A COMPARISON OF THE INFLUENCE OF DIFFERENT COMPONENTS OF THE GOOD BEHAVIOR GAME IN A PRESCHOOL CLASSROOM

By

Elizabeth A. Foley

Submitted to the graduate degree program in Applied Behavioral Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

Chairperson, Claudia L. Dozier, Ph.D.

Florence D. DiGennaro Reed, Ph.D.

Vincent T. Francisco, Ph.D.

Date Defended: March 16, 2017

The Thesis Committee for Elizabeth A. Foley
certifies that this is the approved version of the following thesis:

A COMPARISON OF THE INFLUENCE OF DIFFERENT COMPONENTS OF THE GOOD BEHAVIOR GAME IN A PRESCHOOL CLASSROOM

Chairperson, Claudia L. Dozier, Ph.D.

Date approved: March 16, 2017

Abstract

The Good Behavior Game (GBG) is a multicomponent treatment package that has been demonstrated to decrease disruptive behavior in various populations (kindergarten through high school-age students) and environments (Barrish et al., 1969; Embry, 2002; Tingstrom et al., 2006). However, there is limited research evaluating the GBG with preschool-age children (Swiezy, Matson, & Box, 1992). Furthermore, few studies have evaluated the effects of various components of the GBG, and of those that have, most have done so only after exposure to the GBG package (Fishbein & Wasik, 1981; Harris & Sherman, 1973; Medland & Stachnik, 1972). Finally, few studies have reported data at the individual level (Medland & Stachnik, 1972). Therefore, the purpose of our study was to (a) evaluate the effects of the GBG on disruptive behavior of preschool children during group instruction, (b) evaluate the effects of the major components of the GBG before and after implementation of the GBG package, and (c) examine effects at both the group and individual level. Results suggest that the GBG package was necessary for decreasing disruptive behavior. However, after exposure to the GBG, a response-independent contingency was effective for maintaining low levels of disruptive behavior.

Acknowledgements

I would first like to thank my advisor, Dr. Claudia L. Dozier, for all of her support and guidance over the course of the study. I am forever grateful for the time and effort you have invested in this project and me. Second, I would like to thank my lab mates who helped me conduct sessions and provided valuable input over the course of the study. Finally, I would like to thank my family, who have supported me in countless ways. I could not have accomplished this project without the endless support you all have provided.

Table of Contents

Reports suggest that between 8% and 22% of preschoolers exhibit moderate to significant problem behavior (as cited in Bulotsky-Shearer, Fantuzzo, McDermott, 2010), which includes aggression, noncompliance, emotional outbursts, and inattentive behavior (Del'Homme, Sinclair, & Kasari, 1994; Dominguez, Vitiello, Fuccillo, Greenfield, & Bulotsky-Shearer, 2011). In addition, prevalence of problem behavior increases to 34% for preschoolers from low-income families (Dominguez et al., 2011). Problem behavior in preschool classrooms has been associated with various negative effects such as teacher burnout (Hastings & Bham, 2003), student expulsion (Gilliam & Golan, 2006; Tangorra, 2015), and lower student academic gains both during preschool and over time (Bulotsky-Shearer & Fantuzzo, 2011). Zaghlawan and Ostrosky (2011) noted that problem behavior in preschools often occurs during structured learning times and involves multiple children. Thus, implementing an intervention on an individual level during group instruction may be difficult or impractical for a preschool teacher. Therefore, it is important to identify and evaluate interventions that can be implemented for a group of preschool children during structured learning times.

One intervention that may be beneficial for preschool teachers' management of problem behavior in a group setting is the Good Behavior Game (GBG), which is a multi-component behavioral intervention incorporating an interdependent group contingency. In general, the GBG includes (a) dividing a group or class into teams, (b) establishing rules in the form of clear behavioral expectations, (c) providing feedback for rule violations, (d) establishing a criterion for winning the game, and (e) delivering consequences (e.g., rewards) to winning team(s) at the end of some activity or specified period using an interdependent group contingency. In an interdependent group contingency, a behavior-change agent (e.g., teacher) delivers a common consequence (e.g., rewards) to all members of the group based on the behavior of all of the individuals in the group (Litow & Pomroy, 1975).

In the first study on the GBG, Barrish, Saunders, and Wolf (1969) evaluated its effects on disruptive behavior (i.e., out-of-seat and talking-out behavior) of children in a fourth-grade classroom during math periods. Prior to implementing the GBG, the teacher divided the class into two teams based on current seating assignments. During the GBG phases, the teacher implemented the multi-component intervention in which she first reviewed the rules for winning the game (e.g., staying in seat unless raised hand to request permission) and provided specific examples that described the behavioral expectations. Next, she explained the interdependent contingency in which she told the class that if any individual on a team broke a rule, then the entire team would receive a hatch mark on the chalkboard. In addition, she told the class the criterion for winning the game (i.e., earning the reward). The criteria for a team to win the game was either (a) receiving five or fewer hatch marks (both teams could win if both met this criterion) or (b) receiving the fewest hatch marks (if none of the teams met the criterion of five or fewer hatch marks). Finally, at the end of the math period, the teacher tallied the hatch marks that each team received to determine which team(s) won the game. Individuals on the team(s) who won the game received victory tags and special privileges (e.g., extra recess and time for special projects). The experimenters demonstrated that the GBG decreased problem behavior of the class every time it was implemented.

The GBG has since been repeatedly demonstrated to decrease problem behavior in classrooms with children in kindergarten (e.g., Donaldson, Vollmer, Krous, Downs, &

Berard, 2011; Donaldson, Wiskow, & Soto, 2015), grade school (e.g., Bostow & Geiger, 1976; Harris & Sherman, 1973), and high school (e.g., Flower, McKenna, Muething, Bryant, & Bryant, 2013; Kleinman & Saigh, 2011). In addition, researchers have demonstrated the GBG to increase appropriate behavior such on-task behavior (e.g., Fishbein & Wasik, 1981; Johnson, Turner, & Konarski, 1978). Furthermore, it has been demonstrated to be effective in a variety of settings including the library (Fishbein & Wasik, 1981), the cafeteria (McCurdy, Lannie, & Barnabas, 2009), and in inclusive classrooms (Johnson et al., 1978; Lastapes, 2013). Finally, it has demonstrated generality across cultures with replications conducted in Sudan (Saigh & Umar, 1983) and Belgium (Leflot, van Lier, Onghena, & Colpin, 2013).

Although researchers have demonstrated the GBG to be a robust technology, several limitations to the technology warrant further research. One limitation is that there have been few studies conducted with *young* children in classroom environments, particularly preschool-age children. One exception is an evaluation of a modified version of the GBG conducted by Swiezy, Matson, and Box (1992). In this study, the experimenters showed that the GBG was effective for increasing compliance and cooperation for four preschool children (i.e., two dyads of preschool children) during free-play periods. The experimenters conducted 15-min sessions with each dyad in which the experimenters used a puppet to provide 10 standard instructions (e.g., "Jack and Jill, go put the books on the shelf" or "Jack and Jill, shake hands") to the dyad. If one child in the dyad complied with the instruction, the experimenter delivered praise to that child while the experimenter ignored the behavior of the other child. If both children complied with the instruction, the experimenter delivered praise and a token. The experimenters established a winning criterion for each dyad by determining the mean level of compliance during baseline, and then systematically increasing this number over the course of the experiment. Additionally, each dyad could win an extra reward if they complied as a pair to all 10 instructions. Results showed a large increase in compliance for each dyad with the implementation of the GBG.

Although, Swiezy et al. (1992) provided an example of the utility of the GBG in a preschool setting, there are several limitations of the study. First, although results demonstrated that the GBG was effective when implemented with two children at a time, it is unclear whether it is effective as a class-wide intervention. Second, the researcher evaluated the GBG during free play, thus it is unclear whether the GBG is effective during an instructional period (e.g., circle time) in a preschool setting. Third, there are several potential confounds that limit conclusions that can be made regarding the effects of the GBG. For example, additional procedures such as the use of a puppet to deliver instructions may have increased responding regardless of the implementation of the GBG.

Although, only one study to date has evaluated the influence of the GBG with preschoolers, several studies (Donaldson et al., 2011; Donaldson et al., 2015; McGoey, Schneider, Rezzetano, Prodan, & Tankersley, 2010; Tanol, Johnson, McComas, & Cote, 2010) have been conducted with children in kindergarten classrooms. Donaldson et al. (2011) evaluated the effects of the GBG on the disruptive behavior (i.e., getting out of seat, talking out of turn, and touching other children) of kindergarten students in five kindergarten classrooms during structured, group-learning times (i.e., reading and math). Results showed that the GBG effectively decreased disruptive behavior across all of the kindergarten classrooms. Therefore, these data suggest that the GBG game is effective with young learners during structured learning times and may be a viable procedure to address disruptive behavior during structured learning times in a preschool setting.

In another study, McGoey et al. (2010) evaluated the effects of the Good Behavior Game plus Merit (GBG-PM) in three kindergarten classrooms. In this modification of the GBG, children lost points for engaging in disruptive behavior and earned points for engaging in appropriate behavior. Teachers implemented the GBG throughout the day, across several different activities (e.g., free play, meals, and circle time); however, results showed that levels of problem behavior decreased only slightly. Reasons for the limited effects may be the age of the children, the difference in behavioral expectations across the classroom activities, or both. For example, the expectations for changes in behavior during circle time may be very different than those during a free-choice activity. Thus, it may be best to assess the GBG under one context or activity at a time, similar to the Donaldson et al. (2011) study.

In addition to limited evidence on the efficacy of the GBG with preschoolers, there are additional limitations of the existing research on the GBG that should be addressed in future research. One limitation is that most experimenters have evaluated the effects of the entire multi-component GBG package without systematically determining the necessary and sufficient components of the GBG. Identifying the necessary and sufficient components could potentially save time and resources in the implementation of the technology. Additionally, these more fine-grained analyses may provide insight in identifying the mechanism(s) responsible for behavior change and lead to a more conceptually systematic understanding of the efficacy of the GBG. The few studies that have evaluated the components of the GBG have done so after participants experienced the entire intervention package (Fishbein & Wasik, 1981; Harris & Sherman, 1973; Medland & Stachnik 1972), so it is unclear whether these components would be effective alone, in the absence of a history of the entire package, or both.

Harris and Sherman (1973) evaluated the effects of the GBG and several of its components on disruptive behavior in fifth- and sixth-grade classrooms during math and English periods. After showing that the GBG was effective for decreasing disruptive behavior in both classrooms, the experimenters evaluated the influence of the consequence component of the GBG. That is, the teams, rules, feedback, and winning criterion of the GBG were still in place; however, there was no programmed consequence for winning. Results showed little difference in levels of disruptive behavior between this condition and baseline conditions suggesting the importance of a programmed consequence for winning the game. Next, the experimenters evaluated the influence of the winning criterion. In this phase, the teams, rules, feedback, a winning criterion, and the programmed consequence for winning were in place; however, the criterion for winning was increased and decreased. That is, sometimes the criterion was set at four hatch marks or less, whereas at other times it was set at eight hatch marks or less. Results showed that the occurrence of disruptive behavior was lower than in baseline with both criteria; however, lower levels of problem behavior occurred with the more stringent criterion (four or less hatch marks) than with the less stringent criterion (eight or less hatch marks). Thus, disruptive behavior appeared to change as a function of the criterion, suggesting the importance of this component in the GBG. Next, the experimenters evaluated the influence of the feedback component. In this phase, the teams, rules, a winning criterion, and a programmed consequence for winning were in place; however, the teacher no longer placed hatch marks on the board. Instead, she tallied rule violations on a piece of paper hidden behind a podium. Results showed that levels of disruptive behavior were similar to those observed with the GBG package, which suggests that feedback (at least overt and public feedback in the form of hatch marks) may not be an important variable in the efficacy of the GBG. However, it is possible that teams noticed when the teacher was marking down rule violations on the paper behind the podium, which may have served as a more general form of feedback. Finally, the experimenters evaluated the influence of teams. In this final evaluation, the experimenters eliminated the teams and implemented the GBG with the entire class as a single team. Results showed similar reductions in the level of disruptive behavior in this manipulation as compared to when the class was divided into two teams during the math period; however, similar reductions in disruptive behavior were not observed in the English period.

In another study, Medland and Stachnik (1972) evaluated the effects of two GBG components in a fifth-grade reading class. First, the experimenters implemented the GBG package in which feedback was delivered using green and red lights. During the game, the experimenters placed a light that could be illuminated green or red in view of each team. If everyone on a team was following the rules, the light was illuminated green. If a child on a team violated a rule, the light was illuminated red for at least 30 s; the red light remained on until there were no occurrences of disruptive behavior displayed by that team for 30 s. The experimenters demonstrated that the GBG package with green and red feedback lights effectively decreased disruptive behavior. Next, the experimenters evaluated the effects of two components: rules alone and rules plus feedback (i.e., the green and red lights). Results showed that rules alone did not show similar effects as the GBG. However, rules plus feedback decreased disruptive behavior to levels similar to those with the GBG. After a history with the GBG, the rules and feedback may have acquired stimulus control over appropriate and problem behavior. For example, when the green light was illuminated, it could have served as a conditioned reinforcer due to its history of being paired with winning the game and accessing the reward (i.e., receiving special privileges). Additionally, when the red light was illuminated, it could have served as a conditioned punisher due to its history of being paired with the loss of rewards.

Although a few studies have included evaluation of some components of the GBG, all of these studies involved evaluation of these components after participants had a history with the GBG package. Thus, it is unknown if these components would have the same effect prior to a history with all components of the GBG. If a component (or a combination of components) is effective without a history of the GBG package, this may decrease the amount of time and resources needed to implement the GBG. Additionally, by comparing the effects of components before and after a history with the GBG package, we will be able to determine the potential influence of the GBG on subsequent effects of components of the GBG. If a particular component (or combination of components) is effective after a history with the GBG, this may decrease the amount of time and resources needed to maintain effects of the GBG. For example, a teacher may first implement the entire game to obtain initial effects and then move to rules plus feedback to maintain those effects.

Another limitation of the GBG literature is that previous research has primarily included data collection and analysis of group data (e.g., mean levels of behavior of a group of children or entire class of children). Although a few studies have included data collection and analysis of individual participant data (McGoey et al., 2010; Medland & Stachnik, 1972; Swiezy et al., 1992; Tanol et al., 2010), most either do not report those data or only report the data for the most disruptive children. Therefore, the effect of the GBG on the behavior of individual children is unknown. Analyzing data at the individual level allows one to determine the influence (and the degree of influence) of behavior change for each individual, which may be masked by mean rates of groups of individuals. One reason these data may be beneficial is for determining individuals for whom the GBG is ineffective, thus allowing caregivers to modify the GBG or implement additional interventions to affect behavior changes (Harris & Sherman, 1973; Medland & Stachnik, 1972). Another reason these data may be beneficial is for determining individual differences across all participants within a group. For example, if one is determining the influence of different components of the GBG, some individuals may show behavior change with some components, whereas other individuals may show behavior change with other components.

The purpose of the current study was to replicate and extend research on the use of the Good Behavior Game (GBG) on the disruptive behavior of young children. First, I replicated previous research by evaluating the effects of the GBG on the disruptive behavior of preschool children during structured group instruction (i.e., circle-time activities). Second, I evaluated the influence of the different major components of the GBG before and after implementation of the entire GBG package. Third, throughout all evaluations, I examined the effects on both individual and group disruptive behavior displayed by participants.

Method

Participants

Experimenters conducted the study during structured circle times in a university-based preschool classroom serving children ages 3 1/2 to 5 years. Over the course of the study, participants included 22 typically developing children who were assigned to three different teams (blue, red, and green teams) prior to the beginning of the study. Although a few children with autism spectrum disorders sometimes attended this classroom for short amounts of time, because their attendance was not consistent and their ability to follow simple instructions was limited, they were not included as participants in the study.

The lead experimenter assigned children to the three teams using a matched-allocation procedure (McBurney & White, 2007), which was based on the frequency of disruptive behavior during circle time as reported by the classroom supervisor. Specifically, two graduate-student supervisors for the classroom rank ordered all of the children according to their frequency of disruptive behavior (i.e., out of seat, inappropriate vocalizations, and touching others) during circle time. The child reported to engage in the most frequent disruptive behavior ranked first and the child reported to engage in the least frequent disruptive behavior ranked last. The experimenter then took this rank-ordered list and allocated three children at a time (from most disruptive to least disruptive) to one of the three teams. That is, the lead experimenter quasi-randomly allocated children ranked 1st, 2nd, and 3rd to the three different teams, then she allocated children ranked 4th, 5th, and 6th to the three different teams, and so on until she assigned all children to a team.

Initially, 20 children participated in the study. That is, the Blue Team had seven children (ranked 2nd, 4th, 7th, 11th, 14th, 18th and 20th); the Red Team had seven children (ranked 1st, 5th, 9th, 10th, 15th, 17th, and 19th); and the Green Team had six children (ranked 3rd, 6th, 8th, 12th, 13th, and 16th). However, due to children moving or changing preschools during various points of the study, the Blue Team went from seven to six team members and the Green Team went from six to four team members. Experimenters did not replace the member of the Blue Team that left because the team still had six members; however, they did replace two team-members on the Green Team who left with children from an adjacent classroom. Supervisors from the adjacent classroom completed disruptive behavior rankings for all children in their class. Experimenters selected two children whose ranking was similar to the two children for whom they replaced.

Three undergraduate-level practicum teachers completing a specialty area course sequence in Early Childhood Education and Intervention staffed the preschool on the morning (AM) and afternoon (PM) shifts. The practicum teachers were supervised by graduate students (one on AM shift and one on PM shift) who ran the day-to-day operations of each shift and provided ongoing supervision to the practicum teachers on that shift. Circle time occurred once on the AM shift and once on the PM shift and was approximately 15 min in duration. Experimenters conducted sessions during both AM and PM circle times. During circle time, children sat on mats in a semi-circle facing the lead teacher and engaged in various teacher-led activities (e.g., conducting science experiments, reading books, singing songs, and sorting and sequencing activities).

During circle time, children sat in assigned positions on colored mats that corresponded to their team. That is, the Blue Team sat on blue mats on the left side of the semi-circle, the Red Team sat on red mats in the middle section of the semi-circle, and the Green Team sat on green mats on the right side of the semi-circle (see Appendix A).

Materials

Prior to some sessions, the experimenter presented a poster board (55.9 cm x 71.1 cm) to the children that contained three rules that the children should follow during circle time (see Appendix B). The rules on the poster were (a) sit on your mat, (b) raise your hand, and (c) hands to yourself. Next to each written rule was a picture depicting children following that particular rule. For example, next to the "sit on your mat" rule was a picture depicting children sitting on a mat and facing forward.

During some sessions, the experimenter used a dry-erase board and a black marker to record hatch marks when children on a team violated one of the rules. Under each team-name heading were boxes in which the experimenter recorded each hatch mark for a particular team (see Appendix C). During sessions that included a criterion for winning the game, the experimenters modified the dry-erase board in an attempt to denote this criterion to the children. To do so, experimenters added a line using yellow duct tape that signaled to the children that they could not get any hatch marks below that line to meet criterion. In the current study, the winning criterion was no more than 10 hatch marks per team, thus the experimenter used the yellow strip of duct tape to make a line between the first 10 boxes and the remaining boxes on the board (see Appendix D).

Rewards used in various phases included a variety of small tangible items (e.g., stickers, temporary tattoos, and small toys) and individually wrapped edible items (e.g., ring pops, fun-size candy bars, and tootsie rolls). Children chose rewards from a treasure chest when applicable.

Response Measurement and Interobserver Agreement

During all sessions, trained observers collected frequency data on disruptive behavior displayed by each child using paper-pencil data collection methods. In addition, observers used a timer to determine the duration of each circle-time session. During each session, a separate data collector collected data on disruptive behavior displayed by the children on a particular team. Thus, during all sessions, three data collectors (one assigned to each team) collected data. Disruptive behavior included *out of seat, inappropriate verbal behavior,* and *touching other children.* Data collectors scored *out of seat* if a child's bottom raised off their mat in the absence of a teacher's request to stand up, lean forward, or lay down. Additionally, if the child laid down on his or her back (or stomach) or turned away from circle in the absence of a teacher's request to lay down or turn away, data collectors scored out of seat. Data collectors scored *inappropriate verbal behavior (IVB)* if the child vocalized in the absence of a teacher request for an individual or choral response and included yelling questions out of turn, answering a question asked to another individual, whispering or talking to a peer, humming or singing, cursing, crying, and other inappropriate noises (e.g., tongue clicking or lip smacking). A new occurrence of IVB was scored if at least 2 s elapsed between instances of behavior. Data collectors did not score IVB if children engaged in choral laughter or expressive reactions to teacher activities. Data collectors scored *touching other children* if a child's hands, feet, elbow, or other body part made contact with another child in the absence of a teacher's instruction to do so. Thus, data collectors did not score this behavior if an activity involved appropriately touching others such as holding hands or giving a high five.

Frequency data were converted to a rate measure for individual, team, and class-wide disruptive behavior. For individual data analysis, the experimenter divided the frequency of disruptive behavior for a particular child by the session duration to determine the rate measure. For team data analysis, the experimenter divided the frequency of disruptive behavior displayed by all children on a team by the session duration to determine the team rate measure. For class-wide data analysis, the experimenter divided the frequency of disruptive behavior across all children in the classroom by the session duration to determine the class-wide rate measure. If there were less than four members on a team present during a session, the experimenter collected data on the behavior of the individuals who were present; however, the experimenter only analyzed the data at the individual level for that team.

A second observer simultaneously but independently collected data during at least 33% of sessions across the three teams. We calculated interobserver agreement by dividing the session into 1-min intervals and calculating proportional agreement for each type of disruptive behavior. That is, for each disruptive behavior, the experimenters divided the smaller frequency by the larger frequency in each 1-min interval, summed the quotients across intervals, divided the total by the number of intervals, and multiplied by 100. Mean IOA for the Green Team was 93% (range, 51% - 100%) for IVB, 97% (range, 78% - 100%) for out of seat, and 98% (range, 83% - 100%) for touching others. Mean IOA for the Blue Team was 93% (range, 73% - 100%) for IVB, 97% (range, 83% - 100%) for out of seat, and 98% (range, 82% - 100%) for touching others. Mean IOA for the Red Team was 93% (range, 69%-100%) for IVB, 97% (range, 81% - 100%) for out of seat, and 98% (range, 85% - 100%) for touching others. During a few sessions, IOA for a particular disruptive behavior was below 80%. When this happened, the experimenter retrained data collectors on operational definitions and required the data collectors to collect data until they achieved 90% IOA for two consecutive sessions before they could independently collect data for future sessions.

General Procedure

Prior to all sessions, the research assistants laid out the color mats in a semi-circle facing the lead teacher for the circle time activity (Appendix A). Each child sat in the same assigned seat across all sessions. During all sessions, the lead teacher implemented a daily lesson plan centered on a particular theme and various acquisition goals. For example, the theme might be animals and habitats; the goals for the week might include simple instruction following, restating facts, and identifying places on a map. Throughout the session, the lead teacher implemented various activities and presented various individual and group questions surrounding these themes and goals. Furthermore, teachers implemented several general behavior management strategies as part of the already established classroom-behavior-management system. That is, teachers had previously been trained to implement rule reminders (e.g., “remember to sit quietly during circle”) and positive reinforcement (i.e., praise and physical attention) for appropriate circle-time behavior, as well as planned ignoring, response blocking, and three-step prompting (least-to-most prompting) for inappropriate circle-time behavior.

Experimenters used a reversal design to demonstrate the effects of the GBG and the different components of the GBG before and after implementation of the GBG package.

Phases included baseline, rules, rules + feedback, rules + feedback + criterion, rules + feedback + criterion + noncontingent rewards (end of session rewards regardless of behavior), and the GBG package (rules + feedback + criterion + rewards delivered according to an interdependent group contingency).

Baseline. During baseline, the children sat in their assigned spots on their team mats; however, no other programmed antecedents or consequences were implemented.

Data from the baseline phase were used to determine the winning criterion for instances of disruptive behavior during subsequent phases. To determine this criterion, the lead experimenter first determined the mean frequency of problem behavior across teams in baseline. Next, the experimenter calculated an 80% reduction from this mean baseline level. Specifically, the mean frequency of disruptive behavior across the three teams during baseline was 50.2. An 80% reduction from this frequency was 10 occurrences of disruptive behavior. Therefore, our winning criterion was 10 or fewer occurrences of disruptive behavior.

Rules. Rules sessions were identical to baseline with the addition of rules provided to the children prior to beginning circle. Prior to each rules session, the experimenter showed the children the rules poster board (Appendix B) and reviewed the three rules. That is, the experimenter told the children, "These are the rules for circle time. The first rule is that we will sit on our mats (while pointing to the corresponding picture)—we will sit with our bottom on the mat and face forward, unless a teacher tells you that you can get off your mat. The second rule is that we raise our hands (while pointing to the corresponding picture)—we will raise our hands if we need help, have a question, or just want to say something. The third rule is that we keep our hands to ourselves (while pointing to the corresponding picture)—we will not touch our friends during circle time. Next, the experimenter asked the children to repeat each of the rules while she pointed to each of the pictures.

Prior to session 120, we modified the rules presentation in order to clarify the behavioral expectations for circle time (based on some occurrences of disruptive behavior that recently emerged). The modification included providing additional exemplars of acceptable and unacceptable behavior. For example, the experimenter told the children that they should not make noises like clicking their tongue or smacking their lips during circle.

Rules + Feedback (Hatch Marks). Rules + feedback (RF) sessions were identical to rules sessions; however, during the session the experimenter held up the dry-erase board (Appendix C) and provided hatch marks for rule violations. After reviewing the rules, as described above, the experimenter informed the children that their team would now receive a mark every time someone on their team broke a rule during circle. For example, during a session, if a child on the Red Team talked out of turn during circle, the experimenter placed a hatch mark on the dry-erase board under the Red Team's section of the board. As the experimenter placed the hatch mark, she stated the team and the delivery of the hatch mark, "Point red." The experimenter held the dry-erase board in a position such that all the children could see the board when she recorded hatch marks for each team.

Rules + Feedback (Hatch Marks) + Criterion. Rules + feedback + criterion (RFC) sessions were similar to RF sessions; however, the yellow tape was added to the dry erase board to denote the winning criterion (Appendix D). After the experimenter reviewed the rules and the delivery of hatch marks, she introduced the winning criterion (10 or fewer instances of disruptive behavior) to the students. That is, she told the children that they were going to play a game, and in the game, they should try to avoid getting marks on the dry erase board. In addition, she pointed to the yellow line on the dry-erase board and told the children that the object of the game was to make sure the marks for their team did not go past the yellow line. The experimenter told the children that if their marks did not go past the yellow line, then their team would be the winners of the game. After each session, the experimenter counted the number of hatch marks delivered to each of the teams and stated in a neutral tone whether each team won or did not win the game.

Rules + Feedback (Hatch Marks) + Criterion+ Noncontingent Rewards. Rules + feedback + criterion + noncontingent rewards (RFC+NCR) sessions were similar to RFC sessions except the experimenter delivered the rewards noncontingently (regardless of behavior) after each circle-time session. The experimenter reviewed the rules, hatch marks, and winning criterion as previously described. However, at the end of each circle-time session, the experimenter told the children, "You all get a treat today for being at circle time. You can pick a treat or a toy from the treasure chest. I will bring the treasure chest to everyone who was at circle today so you can pick a treat or toy." The experimenter then immediately walked around to each child and allowed the child to select one treat or toy from the treasure chest. The purpose of this condition was to determine the influence of the contingency when the experimenter delivered rewards on a response-independent schedule as compared to a response-dependent schedule.

Rules + Feedback (Hatch Marks) + Criterion+ Contingent Rewards (GBG).

During GBG sessions, all of the components of the GBG were in effect. That is, sessions were similar to RFC+NCR sessions; however, an interdependent group contingency was implemented (Litow & Pumroy, 1975) in which the behavior of everyone on a team must meet the winning criterion for the entire team to earn the reward after the circle-time session.

Prior to GBG sessions, the experimenter reviewed the rules as previously described then told the children, “Today we are going to play the good behavior game. When we play the game, if you don’t follow the rules for circle time, then your team will receive a mark on the board (and pointed to the dry-erase board). You don’t want to get marks, because if you get too many marks and your marks go past the yellow line then you might not get a toy or treat at the end of circle today. If your team’s marks do not go past the yellow line then your team will get a toy or a treat from the treasure box. But if all of the teams go past the yellow line then the team with the smallest number of marks will win.”

At the end of the GBG session, the experimenter reviewed the marks on the dry-erase board with the children. The experimenter rotated through each team, showed each team the dry-erase board, and counted the number of hatch marks aloud to the team. If a team did not receive any marks, the experimenter told the team that they did not receive any marks. After the experimenter counted the number of hatch marks that a team received, the experimenter asked the children on the team to identify if the marks went past the winning criterion line (e.g., “Red Team, did your marks go past the yellow line?”). Children were always able to identify if their team’s marks went past the winning criterion line. Any team who did not receive marks past the winning criterion won the game. If all of the teams received more marks than the winning criterion, then the experimenter declared that the team with the smallest number of hatch marks was the winner. The experimenter then went to each individual child on the winning team(s) and allowed them to select a reward from the treasure box.

Treatment Integrity

Experimenters calculated treatment integrity of hatch-mark delivery (experimenter feedback) by comparing the number of hatch marks delivered to the number of disruptive behaviors scored for that session. That is, during 20% of sessions across phases in which hatch marks were delivered, data collectors scored the number of hatch marks the experimenter delivered across teams and the number of disruptive behaviors scored across teams, and divided the smaller of these numbers by the larger of these numbers to get a percentage agreement measure. The evaluation of treatment integrity focused on implementation of this component, instead of adherence to the entire GBG procedure (e.g., creating teams, reviewing rules, and providing contingent rewards) as it provides a more conservative measure of integrity. That is, correctly and consistently providing hatch marks for disruptive behavior may be the most difficult component of the procedure. Furthermore, it is the aspect of the intervention that determines reward delivery. Thus, if it is implemented incorrectly, then it may impede the efficacy of the intervention. Mean treatment integrity for rules + feedback sessions was 60% (range, 60% - 60%). Mean treatment integrity for rules + feedback + criterion sessions was 70% (range, 68% - 71%). Mean treatment integrity for the rules + feedback + criterion + noncontingent rewards sessions was 79% (range, 53%-92%). Mean treatment integrity for GBG sessions was 77% (range, 50% - 95%).

Low treatment-integrity scores appear to be a result of two scenarios: (a) extremely high occurrences of disruptive behavior or (b) extremely low occurrences of disruptive behavior. For example, during session 164, the data collectors scored 260 occurrences of disruptive behavior and the experimenter delivered 157 hatch marks across all teams. Given this high rate of disruptive behavior, the experimenter missed some instances of disruptive behavior that the data collectors scored. Alternatively, during sessions with only a few occurrences of disruptive behavior, if the experimenter missed a few occurrences, treatment integrity was low. However, poor treatment integrity did not appear to have an adverse effect the efficacy of the GBG. For example, during session 176 (the fifth GBG phase) treatment integrity was 67% for the session. Specifically, data collectors scored 18 instances of disruptive behavior across the three teams, whereas the experimenter provided only 12 hatch marks across the three teams. However, results show that even when integrity for the delivery of hatch marks was low (did not completely correspond with recorded instances of disruptive behavior), the GBG resulted in more than an 80% decrease in disruptive behavior as compared to baseline levels.

Results

Figure 1 depicts the class-wide data and Figures 2-4 depict the different team data. The top panel of each figure depicts the rate of all disruptive behavior across sessions. The bottom panel depicts the mean rate of disruptive behavior in the last five sessions for each phase. However, if a phase had fewer than five sessions, then all sessions in that phase were used for the calculation (these phases are denoted by an asterisk in the graph).

The top panel of Figure 1 shows that disruptive behavior occurred at moderate to high rates in the first baseline (BL) phase and in all of the additive component phases conducted before exposure to GBG. That is, moderate to high rates of disruptive behavior continued to occur in Rules (R), Rules + Feedback (RF), Rules + Feedback + Criterion (RFC), and Rules + Feedback + Criterion + Noncontingent Rewards (RFC+NCR) phases. However, after the first nine sessions of the first GBG phase, disruptive behavior decreased and maintained at low levels (i.e., it occurred at levels lower than an 80% reduction from the initial baseline as depicted by the dotted line on the graph). Next, in the reversal to BL, after the first three sessions, disruptive behavior increased to levels similar to those in the initial baseline and additive component conditions conducted prior to the GBG. In the second GBG phase, disruptive behavior immediately decreased; however, rates did not consistently remain below an 80% reduction from baseline. Therefore, at session 120, the experimenter modified the rules provided to the children prior to circle time in an attempt to clarify the behavioral expectation, particularly with respect to some different topographies of disruptive behavior that had emerged. After this modification, an immediate decrease to levels below an 80% reduction of disruptive behavior occurred and maintained for the remainder of the phase. Furthermore, these effects were replicated in the three subsequent GBG phases.

After implementation of the GBG, various results occurred across the different additive component phases. Results show that low levels of disruptive behavior initially maintained in the R and RF phases. However, after three sessions in the R phase and seven sessions in the RF phase, levels of disruptive behavior increase to baseline or higher levels. In the RFC phase, disruptive behavior immediately increased to high levels. However, in the first RFC+NCR phase, disruptive behavior maintained at low levels across more than 10 sessions. After a reversal to baseline, disruptive behavior again decreased to low levels in the second RFC+NCR phase. Experimenters were only able to conduct three sessions in this last replication phase because it was the end of the school year.

The bottom panel of Figure 1 shows that the mean rate of class-wide disruptive behavior (during the last five sessions of each phase) was never below an 80% reduction from baseline prior to the implementation of GBG. However, there is somewhat of a reduction in mean rate of disruptive behavior in all of the phases implemented after the initial baseline and before the GBG. The mean rate of disruptive behavior is at or below an 80% reduction from baseline every time the GBG was implemented. Furthermore, in looking at the phases conducted after an immediate history with the GBG, only during the RFC+NCR phase is the reduction in disruptive behavior at or below an 80% reduction from baseline. However, lower mean rates of disruptive behavior occurred in the R phase and the BL phases conducted after the GBG as compared to those phases conducted prior to the GBG. In contrast, higher mean rates of disruptive behavior occurred in RF and RFC phases conducted after the GBG as compared to those phases conducted prior to the GBG.

Figure 2 depicts data for the Blue Team. The top panel of Figure 2 shows similar results to the class-wide data. That is, prior to the GBG, the Blue Team engaged in moderate to high levels of disruptive behavior in baseline phases as well as each of the component phases. Furthermore, after the first nine sessions of the first GBG phase, disruptive behavior decreased and maintained at low levels. In the reversal to BL, the Blue Team initially maintained relatively low levels of disruptive behavior; however, these levels increased over time. Also similar to the class-wide data, subsequent GBG phases resulted in a decrease in disruptive behavior. After a history with the GBG, initial low levels of disruptive behavior were observed in the R and RF phases; however, disruptive behavior increased after a few sessions. After implementation of the GBG, low levels of disruptive behavior are observed, which maintained over time in the RFC+NCR phases.

The bottom panel of Figure 2 shows that mean rate of disruptive behavior (during the last five sessions of each phase) for the Blue Team was never below an 80% reduction from baseline prior to the implementation of the GBG. However, there is somewhat of a decrease in mean rate of disruptive behavior in the R, RFC, and RFC+NCR phases, as compared to baseline levels. Somewhat higher levels of disruptive behavior occurred in the RF phase. Reductions in the mean rate of disruptive behavior occurred in all five GBG phases, and this reduction is at or below an 80% reduction from baseline levels for four out of the five GBG phases. Furthermore, similar results to those in the class-wide data occurred in the additive component phases after an immediate history with the GBG. That is, only during the RFC+NCR phase is the reduction in disruptive behavior at or below an 80% reduction from baseline. In addition, lower mean rates of disruptive behavior occurred in the R and BL phases following the GBG as compared to those prior to the initial GBG, and higher mean rates of disruptive behavior occurred in RF and RFC phases following the GBG as compared to those prior.

Figure 3 depicts data for the Red Team. The top panel of Figure 3 shows that disruptive behavior occurred at moderate to high levels in the initial BL phase and in the R phase. Somewhat lower levels of disruptive behavior occurred in the RF phase; however, behavior increased again in the RFC and RFC+NCR phases. Similar to the class-wide and Blue Team data, after the first nine sessions of the first GBG phase, disruptive behavior decreased and maintained at low levels. In the reversal to BL, similar results to those seen in class-wide data and Blue Team data were observed. That is, low levels of disruptive behavior maintained for the first three sessions, then increased over time. Also, similar to the class-wide data and Blue Team data, the GBG resulted in a decrease in disruptive behavior that maintained over time. After a history with the GBG, moderate to high levels of disruptive behavior occurred in the R and RFC phases. Unlike the class-wide data and Blue Team data, however, lower levels of disruptive behavior occurred in the RF phase when it was implemented after a history with the GBG package. Finally, reductions similar to those observed in the class-wide data and Blue Team data occurred with implementation of RFC+NCR after a history with the GBG.

The bottom panel of Figure 3 shows high mean levels of disruptive behavior in the initial BL phase. Also, as compared to BL, somewhat lower levels of disruptive behavior occurred in the R, RF, RFC, and RFC+NCR phases conducted prior to implementing the GBG. However, levels of disruptive behavior were not below an 80% reduction from initial baseline levels. Similar to the class-wide data and Blue Team data, an 80% reduction in the mean rate of disruptive behavior occurred in all GBG phases. In the phases conducted after a history with the GBG, lower levels of disruptive behavior occurred in the BL phases as compared to the initial BL phase; however, levels are still above an 80% reduction from baseline levels. Similar to the class-wide data and Blue Team data, an 80% reduction in disruptive behavior was observed in the RFC+NCR phases conducted after the GBG was implemented. However, the mean rate of disruptive behavior was also below an 80% reduction from baseline levels in the RF phase conducted after GBG was implemented.

Figure 4 depicts data for the Green Team. The top panel shows similar results to the class-wide data and other teams' data. That is, in the initial BL and all of the additive component phases conducted prior to the GBG, the Green Team engaged in moderate to high levels of disruptive behavior. Additionally, after the first nine sessions in the first GBG phase, disruptive behavior decreased and maintained at low levels. Also, similar to class-wide data and other teams' data, after initial maintained low levels of disruptive behavior in the reversal to BL, disruptive behavior increased. Subsequently, every time the GBG was implemented, levels of disruptive behavior decreased. In the analysis of the effects of the additive components after a history with the GBG, decreased levels of disruptive behavior occurred in the R and RFC+NCR phases. In addition, initial low levels of disruptive behavior occurred in the RF phase; however, this decrease did not maintain over time.

The bottom panel of Figure 4 shows moderate levels of disruptive behavior in the initial BL and additive component phases prior to the GBG; however, somewhat lower levels of disruptive behavior occurred in all of the additive component phases as compared to the initial BL phase. In addition, mean rates of disruptive behavior are below an 80% reduction from BL levels every time the GBG was implemented. In addition, after implementation of the GBG, lower mean rates of disruptive behavior occurred in the R and final BL phases as compared to levels in those phases conducted prior to a history with the GBG. However, levels are still above an 80% reduction from baseline levels. Finally, the mean rate of disruptive behavior in the RFC+NCR phases conducted after a history with the GBG was below an 80% reduction from baseline levels.

In addition to the group analyses, experimenters analyzed 17 individual data sets. Individual data for five children are not included because they were not present both before and after the implementation of the GBG. Three of these children were only present for the phases prior to the GBG; two of these children were only present for the phases after the GBG. Experimenters analyzed individual data in various ways. First, experimenters determined the percentage change from initial BL levels of disruptive behavior in each of the GBG phases for each participant. Specifically, for each child, they calculated the mean rate of disruptive behavior for the last five sessions of BL and the mean rate of disruptive behavior for the last five sessions in each of the GBG phases, then they calculated the percentage change across these means. As with the group data, if a phase had fewer than five sessions, then experimenters calculated the mean rate using all of the sessions in that phase. To calculate the percentage change, the experimenters subtracted the mean rate of disruptive behavior in a GBG phase from the mean rate of disruptive behavior in the initial BL phase, then divided this quotient by the initial BL mean rate. If the percentage change was an increase over BL levels, then the percentage is a positive number; if the percentage change was a decrease below BL levels then the percentage is a negative number. As depicted in Table 1, results showed that most children displayed a decrease in disruptive behavior across all administrations of the

GBG. Only two children did not show a decrease in disruptive behavior in one or more of the GBG phases (Charlie during the first phase and Aiden during the first, second, and fourth phases). Results across phases showed that of the 17 children, 10 in the first phase, 13 in the second phase, 15 in the third phase, 12 in the fourth phase, and 12 in the fifth phase showed at least an 80% reduction in disruptive behavior over baseline levels. Thus, similar to the group results, the GBG was effective for decreasing disruptive behavior for most participants across most phases.

Second, experimenters used a similar calculation to determine the percentage change from initial BL levels of disruptive behavior for each component phase (R, R+F, RFC, and RFC+NCR) before and after the implementation of the GBG for each participant. This was calculated by subtracting the mean rate of disruptive behavior in the last five sessions of a component phase from the mean rate of disruptive behavior in the last five sessions of the initial BL phase, then dividing this number by the initial BL mean rate. If the percentage change was an increase over BL levels, then the percentage is a positive number; if the percentage change was a decrease below BL levels then the percentage is a negative number. Table 2 depicts the results of this analysis.

As shown in Table 2, in the R phase conducted prior to the GBG, 14 out of 17 children showed a decrease in disruptive behavior as compared to initial BL levels. Of these 14 children, only one (Tony) showed a decrease of 80% or more from BL levels. In addition, three children (Charlie, Etta, and Daria) showed an increase in disruptive behavior over BL levels. In the R phase conducted after the GBG, 13 out of 17 children showed a decrease in disruptive behavior as compared to initial BL levels. Of these 13 children, six (Lester, Tony, Xander, Andrew, Noah, and Megan) showed a decrease of

80% or more from BL levels. In addition, four children (David, Charlie, Quilla, and Aiden) showed an increase over BL levels. In fact, Aiden showed a 273% increase, which is in large contrast to his 18.2% decrease in the R condition conducted prior to the GBG.

In the RF phase conducted prior to the GBG, 11 out of 17 children showed a decrease in disruptive behavior as compared to initial BL levels. Of these 11 children, five (David, Tony, Etta, Andrew, and Daria) showed a decrease of 80% or more from BL levels. In addition, six children (Brenden, Lester, Lucas, Charlie, Harrison, and Aiden) showed an increase in disruptive behavior. In the RF phase conducted after the GBG, 13 out of 17 children showed a decrease in disruptive behavior as compared to BL levels. Of these children, 10 showed a decrease of 80% or more from BL levels. In addition, four children (Lester, Charlie, Noah, and Aiden) showed an increase over BL levels. In fact, three of these children (all but Noah) showed an increase by more than 100%.

In the RFC phase conducted prior to the GBG, 10 out of 17 children showed a decrease in disruptive behavior as compared to BL. Of these children, four (Tony, Etta, Andrew, and Daria) showed a decrease of at least 80% from BL. In addition, seven kids showed an increase over BL levels, with one child (David) showing a 192.6% increase over BL levels. In the RFC phase conducted after the GBG, 12 out of 17 children showed a decrease in disruptive behavior as compared to BL. Of these children, five (Tony, Xander, Matthew, Andrew, and Megan) showed at least an 80% reduction. Six children showed an increase over BL levels, with one child (David) showing a 1,503% increase over BL.

In the RFC+NCR phase conducted prior to the GBG, 12 out of 17 children showed a decrease in disruptive behavior as compared to BL levels. Of these 12 children, four (Etta, Andrew, Daria, and Megan) showed at least an 80% decrease. Five children showed an increase over BL levels, with two children (David and Aiden) displaying more than a 100% increase. In the first administration of RFC+NCR conducted after implementation of the GBG, 16 out of 17 children (all but Aiden) showed a decrease in disruptive behavior as compared to initial BL levels, and all 16 of these children showed at least an 80% decrease over BL levels. With the second administration, all 17 children showed a decrease; however, only 12 showed that the decrease was at least 80% from BL levels. Four participants percentage decrease was just below 80%; however, for one child (Aiden), his mean percentage decrease was only 8.3%.

General Discussion

The first purpose of the present study was to evaluate the efficacy of the GBG on decreasing the disruptive behavior of a group of preschoolers during a structured learning activity. I aimed to determine the efficacy of the intervention by evaluating group and individual data. Results of the group analysis showed that the GBG produced a significant decrease in the rate of problem behavior at both the class-wide and team level. In the analysis of the group data, a decrease in both variability and level of disruptive behavior occurred every time the GBG was implemented. Thus, the data from the current study suggest that the GBG was an effective intervention for the reduction of disruptive behavior in a group of preschool children during a structured learning activity. This replicates the results of previous research evaluating the GBG with young children

(Donaldson et al., 2011; Donaldson et al., 2015; McGoey et al., 2010; Tanol et al., 2010) while extending the generality of the intervention to a new population and setting (i.e., preschoolers in a structured learning activity). Given that previous research suggests that problem behavior originating in the preschool years persists over time (Breitenstein, Hill, & Gross, 2009, Del'Homme et al., 1994) and is associated with poor academic outcomes in subsequent school years (Bulotsky- Shearer & Fantuzzo, 2011), preschools should consider using the GBG to address problem behavior that presents in these early years. Unlike common interventions such as time-out, the GBG allows the child to remain in the learning environment during the intervention such that learning time is not lost. In addition, the GBG does not require additional effort from the teacher that may be required to implement individual interventions like reprimands, rationales, and time out to all children.

Results of the individual data analysis of the GBG showed that it was effective for decreasing disruptive behavior across children; however, it was not effective for all children every time it was implemented. That is, an 80% or more reduction from mean BL levels of disruptive behavior only occurred for 7 out of 17 participants in all five of the GBG phases. However, for an additional five participants, an 80% or more reduction from BL levels occurred in 4 of the 5 GBG phases. Thus, for 12 out of 17 children, an 80% or more reduction from BL levels occurred in at least four of the five GBG phases. It is important to note; however, that for the majority of participants, even when their reductions were not 80% or above, most were close to that percentage. For example, for the children ranked as the two most disruptive children during circle time (Lester on the Blue Team and Quilla on the Red Team), although they did not display an 80% or more reduction in disruptive behavior in all phases of the GBG, when they did not meet this criterion, their reductions were in the 70%-79% range. However, for a few individuals (e.g., Charlie and Aiden) this was sometimes not the case.

As a practical matter in the classroom, the decrease demonstrated by the majority of children may benefit teachers and children. Although we did not systematically evaluate the social validity of the GBG in our classroom, teachers commented that circle time ran smoother and that they enjoyed conducting circle time more when the GBG was in place. There are several reasons why teachers may prefer the GBG. First, given that disruptive behavior in preschools is associated with teacher burnout (Hastings & Bham, 2003), an intervention that decreases overall levels of disruptive behavior may help combat the fatigue and stress associated with the occurrence of disruptive behavior in the classroom. Second, with a large decrease in disruptive behavior, the teacher may be able to focus on providing effective teaching strategies, rather than the management of problem behavior. Thus, future research should involve a systematic evaluation of the social validity of the GBG by determining the components that the teachers prefer and the ease with which it can be implemented within the context of preschool activities. Furthermore, future research could involve evaluating some of the reasons why the GBG might be preferred, such as determining the degree to which teachers have more time to teach skills and the associated skill acquisition of the children when the GBG is implemented.

The GBG may fit well within the response-to-intervention (RTI) model, which is an assessment and intervention model used by public schools and more recently early childhood education settings such as preschools (Ball & Trammell, 2009). In this model, ongoing assessment is used to allow teachers and administrators to make databased decisions on teaching and intervention, which are implemented using a three-tiered approach. Tier I involves strategies that are implemented to be effective at the class or large-group level. Tier II involves a slightly more intense intervention, which may include individual or group tutoring, additional review and practice sessions, or more systematic and structured intervention. Tier III involves the most intensive and individualized interventions, which are often implemented by a special-education teacher or some other specialist (Bayat, Mindes, & Covitt, 2010). Given that the GBG was effective at the group level, preschools could implement it as a Tier I intervention, and teachers could then use data from individual child performance to identify which children require additional support under a Tier II or III intervention.

Another purpose of the present study was to evaluate the influence of the major components that make up the GBG package in an additive fashion to determine the efficacy of these components when they are implemented before and after exposure to the GBG. Results of this analysis at the class and team level showed that none of the additive components decreased disruptive behavior to at least 80% below initial BL levels prior to exposure to the GBG. However, after implementation of the GBG, consistent decreases to at least 80% below initial BL levels occurred both times the RFC+NCR phase was implemented.

Another outcome of the additive component analysis was that the Red Team demonstrated at least an 80% decrease above BL levels in disruptive behavior in the RF phase that was conducted after exposure to the GBG. Thus, after a history with the GBG, it is possible that the hatch marks became conditioned punishers for disruptive behavior

(Hake & Azrin, 1965). That is, when a rule violation occurred on a specific team, the experimenter delivered a hatch mark to that team, which may have acquired aversive properties due to the hatch marks (or a relatively large number of them) being associated with a loss of access to reinforcement. However, given that this effect was not observed with the other teams, or with the class, it does not appear that rules and feedback could be used in isolation for an entire class. In fact, the highest rates of disruptive behavior for the Blue Team were observed during this same phase.

Results of the additive component analysis at the individual level yielded several interesting results. That is, six participants (Tony, Xander, Matthew, Andrew, James, and Megan) showed a decrease from BL levels in all phases, both before and after the GBG. Furthermore, two other children (Etta and Daria) showed a decrease over BL levels in all phases except the R phase conducted prior to the GBG. In addition, one child (Charlie) did not show a decrease over BL levels until RFC+NCR was implemented after the GBG. Another child (Aiden) only showed a decrease over BL levels in the R phase prior to the GBG phase and the second RFC+NCR phase after the GBG. Furthermore, overall results showed that for 11 out of 17 children, there was not at least an 80% decrease in disruptive behavior across any phase prior to the implementation of the GBG; however, for all but one participant (Aiden), there was at least an 80% reduction in at least one phase after the implementation of the GBG. Finally, although the individual data suggest that some of the additive components either before or after the GBG were effective for reducing disruptive behavior, sometimes to at least an 80% reduction of BL levels, the most consistently effective condition for maintaining low levels of disruptive behavior within and across participants was the RFC+NCR phase conducted after the GBG.

Consistently at all levels of evaluation, RFC+NCR after the GBG was effective for maintaining low levels of disruptive behavior. These results replicate research on the maintenance of behavior when response-independent schedules follow response-dependent schedules (e.g., Dozier et al., 2001; Ringdahl, Vollmer, Borrero, & Connell, 2001). However, without further analysis, it is unclear why this procedure resulted in maintained low levels of disruptive behavior. One possibility is that maintained effects occurred due to adventitious reinforcement (e.g., Ecott & Critchfield, 2004). That is, it is possible that appropriate behavior (i.e., the absence of rule violations or the occurrence of appropriate circle-time behavior) may have occurred in close temporal proximity to the delivery of response-independent reinforcement, resulting in maintained effects. Another possibility is that maintained effects were due to stimulus control (Gamzu & Schwartz, 1973). That is, the various GBG stimuli (e.g., rules, dry-erase board, hatch marks, criterion line, and treasure box) may have acquired stimulus control over responding such that when those stimuli that were previously associated with delivery of a response-dependent reward (a reward delivered for winning the game) became to exert control over levels of disruptive behavior. Thus, in the presence of these stimuli, behavior maintained at low levels. It is also possible that maintained effects were due to a combination of these mechanisms. Regardless of the mechanism of action, there are important practical implications of these results. That is, after initial implementation of the GBG to decrease disruptive behavior, teachers could implement RFC+NCR to maintain effects, which may decrease the effort of implementing the intervention. Future research should be conducted to provide a more systematic analysis of RFC+NCR and determine whether it is easier to implement for teachers.

Although results suggest the utility of the GBG during circle time in preschool classrooms and the potential use of a response-independent GBG procedure for maintaining initial effects, there are several limitations of the current study. One limitation is that a graduate student conducted all of the GBG sessions, which does not allow us to know whether a preschool teacher could implement the game while conducting circle time. In addition, challenges associated with the implementation of the GBG may decrease the acceptability of the intervention. Future research should involve evaluation of the efficacy of GBG with teacher implementation, and as mentioned before, evaluation of the social validity of the GBG by preschool teachers. Another limitation was the lack of control over teacher training and lesson planning. That is, we did not control for the quality of instruction or number of demands presented during each circle-time lesson. Given that the preschool was staffed by undergraduate students completing a practicum, the quality of lesson planning and implementation varied across students. However, this potential confound was present across during the entire study and lawful effects were still observed across phases.

Another procedural challenge was participant attrition, which influenced the makeup of our teams throughout the course of the study, which may have influenced some of the class-wide and team outcomes. Furthermore, the two children who replaced the children on the Green Team were from another classroom. This present two challenges. First, the two children who replaced the original participants were selected because they received similar rankings on their reported frequency of disruptive behavior. However, rankings were relative to the levels of disruptive behavior exhibited by the children in each classroom. That is, although they received similar subjective rankings, their baseline levels of disruptive behavior may not have been similar. After conclusion of the study, the mean rates of disruptive behaviors for these four children were calculated and compared in order to determine the similarity or difference in BL levels of responding between the two sets of children. The differences were .687 and .017 respectively. Second, the children from the adjacent classroom did not experience all of the potential side effects of the GBG that may have occurred throughout the day. For example, children from the classroom in which the GBG was implemented often spoke about the game and sometimes even played the game during free-play periods. In addition, they were observed providing each other rule reminders and reprimands regarding how to play the GBG. Thus, given the nature of the group contingency, it is possible that these experiences influenced the efficacy of the GBG for some individuals. Future research should involve evaluation of the influence of positive and negative comments from peers in order to determine their role in the efficacy of GBG.

Another limitation of the current study was that a decrease in disruptive behavior with the implementation of the first GBG phase did not occur until after nine sessions, which was longer than some of the additive component phases. This may be a threat to internal validity because effects may have been observed in some of these component phases had they been conducted for a longer time (i.e., longer than nine sessions). However, given that the components were implemented in an additive fashion, a total of 30 sessions were conducted in which decreases in disruptive behavior did not occur across component phases. Thus, when the effects of the initial GBG phase are compared to these effects observed across these 30 sessions, there is a clear, albeit delayed, difference in effects. Furthermore, the final RFC+NCR phase was only conducted for three sessions because it was the end of the school year. Thus, it is unknown whether the maintained low effects would have continued to occur over a longer period of time.

Another limitation was the use of a frequency measure rather than a duration measure for the occurrence of some topographies of disruptive behavior. For example, given that some behaviors such as out-of-seat behavior could have occurred for a long duration but only be counted as one instance of disruptive behavior, an interval or duration measure may have been more sensitive to capture the occurrence of behavior.

Another limitation was that modification made to the rules during the second GBG phase, which limits our ability to compare the first rules phase with the rules phase conducted after exposure to the GBG, as well as all of the other phases that included rules prior to this change. However, based on the children's behavior during the second GBG phase, additional disruptive behaviors emerged that were not captured in the rules that were currently provided. Thus, we modified the rules in an attempt to clarify behavioral expectations for the children.

Finally, another limitation was the low treatment integrity scores of experimenter delivery of hatch marks in various phases of the study. Given that high treatment integrity is vital to determining the efficacy of an intervention, low integrity weakens the assertion that the treatment variable(s) is responsible for behavior change. Although treatment integrity scores for the delivery of hatch marks was as low as 50% in the current study, large decreases in disruptive behavior were observed. These results are similar to previous studies (e.g., Donaldson et al., 2011) who also reported low treatment integrity scores (e.g., 60%) while demonstrating a large effect. One reason treatment integrity was low in the present study was that the experimenter conducting the session was observing all 20 children and delivering hatch marks accordingly, whereas data collectors were only responsible for collecting data on the behavior of children on a particular team (approximately six children). Given the results of this study and others, future research should involve evaluating the efficacy of the GBG at varying degrees of treatment integrity with respect to the various components of the GBG.

Overall, there are several conclusions based on the results of this study. First, the GBG is an effective intervention for preschool children during structured learning activities. Second, the entire GBG package was required before a large decrease in disruptive behavior occurred for the class and the different groups. Third, individual data collection allowed for determination of the GBG and various additive components at the level of the individual. These individual results suggest that the additive components had varying levels of efficacy across individuals. Additionally, the results suggest that exposure to the GBG can alter the effects of the additive components. Finally, response-independent schedules may have the potential to maintain low levels of responding after exposure to GBG and may help teachers manage problem behavior.

References

Ball, C.R. & Trammell, B.A. (2011). Response-To-Intervention in high-risk preschools: Critical issues for implementation. *Psychology in Schools, 48,* 502-512. doi: 10.1002/pits.20572

Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good Behavior Game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavioral Analysis, 2,* 119-124. doi: 10.1901/jaba.1969.2-119

Bayat, M., Mindes, G., & Covitt, S. (2010). What does RTI (response to intervention) look like in preschool? *Early Childhood Education Journal, 37,* 493-500. doi: 10.1007/s10643-010-0372-6

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. doi: 10.1007/978-3-319-44124-5_20

Breitenstein, S. M., Hill, C., & Gross, D. (2009). Understanding disruptive behavior problems in preschool children. *Journal of Pediatric Nursing, 24*, 3-12. doi: 10.1016/j.pedn.2007.10.007

Bulotsky-Shearer, R. J., Fantuzzo, J. W., & McDermott, P. A. (2010). Typology of emotional and behavioral adjustment for low-income children: A child-centered approach. *Journal of Applied Developmental Psychology, 31*, 180-191. doi: 10.1016/j.appdev.2009.12.002

Bulotsky-Shearer, R. J., & Fantuzzo, J. W. (2011). Preschool behavior problems in classroom learning situations and literacy outcomes in kindergarten and first grade. *Early Childhood Research Quarterly, 26*, 61-73. doi: 10.1016/j.ecresq.2010.04.004

Del'Homme, M. A., Sinclair, E., & Kasari, C. (1994). Preschool children with behavioral problems: Observation in instructional and free play contexts. *Behavioral Disorders, 19*, 221-232. Retrieved from http://www.ccbd.net/publications/behavioraldisorders

Dominguez, X., Vitiello, V. E., Fuccillo, J. M., Greenfield, D. B., & Bulotsky-Shearer, R. J. (2011). The role of context in preschool learning: A multilevel examination of the contribution of context-specific problem behaviors and classroom process quality to low-income children's approaches to learning. *Journal of School Psychology, 49*, 175-195. doi: 10.1016/j.jsp.2010.11.002

Donaldson, J. M., Vollmer, T. R., Krous, T., Downs, S., & Berard, K. P. (2011). An evaluation of Good Behavior Game in kindergarten classrooms. *Journal of Applied Behavioral Analysis, 44,* 605-609. doi: 10.1901/jaba.2011.44-605

Donaldson, J. M., Wiskow, K. M., & Soto, P. L. (2015). Immediate and distal effects of the Good Behavior Game. *Journal of Applied Behavioral Analysis, 48*, 685-689. doi: 10.1002/jaba.229

Dozier, C. L., Carr, J. E., Enloe, K., Landaburu, H., Eastridge, D., & Kellum, K. K. (2001). Using fixed-time schedules to maintain behavior: A preliminary investigation. *Journal of Applied Behavior Analysis, 34*, 337-340. doi: 10.1901/jaba.2001.34-337

Ecott, C. L. & Critchfield, T. S. (2004). Noncontingent reinforcement, alternative reinforcement, and the matching law: A laboratory demonstration. *Journal of Applied Behavior Analysis, 37*, 249-265. doi: 10.1901/jaba.2004.37-249

Fishbein, J. E., & Wasik, B. H. (1981). Effect of the Good Behavior Game of disruptive library behavior. *Journal of Applied Behavioral Analysis, 14,* 89-93. doi: 10.1901/jaba.1981.14-89

Flower, A., McKenna, J., Muething, C. S., Bryant, D. P., & Bryant, B. R. (2013). Effects of the Good Behavior Game on classwide off-task behavior in a high school basic algebra resource classroom. *Behavior Modification, 38*, 45-68. doi: 10.1177/0145445513507574

Review of Educational Research, 84, 546-571. doi: 10.3102/0034654314536781

Gamzu, E. & Schwartz, B. (1973). The maintenance of key pecking by stimulus-contingent and response-dependent food presentation. *Journal of the Experimental Analysis of Behavior, 19,* 65-72. doi: 10.1901/jeab.1973.19-65

Gilliam, W. S., & Golan, S. (2006). Preschool and child care expulsion and suspension: Rates and predictors in one state. *Infants & Young Children, 19*, 228-245. Retrieved from http://journals.lww.com/iycjournal/pages/default.aspx

Hake, D.F., & Azrin, N.H. (1965). Conditioned punishment. *Journal of the Experimental Analysis of Behavior, 8,* 279-293. doi: 10.1901/jeab.1965.8-279

Harris, V. W., & Sherman, J. A. (1973). Use and analysis of the “Good Behavior Game” to reduce disruptive classroom behavior. *Journal of Applied Behavioral Analysis, 6*, 405-417. doi: 10.1901/jaba.1973.6-405

Hastings, R. P., & Bham, M. S. (2003). The relationship between student behaviour patterns and teacher burnout. *School Psychology International*, *24,* 115-127. doi: 10.1177/0143034303024001905

Johnson, M. R., Turner, P. F., & Konarski, E. A. (1978). The “Good Behavior Game”: A systematic replication in two unruly transitional classrooms. *Education and Treatment of Children, 1*, 25-33. doi: 10.1023/A:1020977107086

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. doi: 10.1177/0145445510392213

Lastrapes, R. E. (2013). Using the Good Behavior Game in an inclusive classroom. *Intervention in School and Clinic, 49*, 225-229. doi: 10.1177/1053451213509491

Leflot, G., van Lier, P. A. C., Onghena, P., & Colpin, H. (2013). The role of children’s on-task behavior in the prevention of aggressive behavior development and peer rejection: A randomized controlled study of the Good Behavior Game in Belgian elementary classrooms. *Journal of School Psychology*, *51*, 187-199. doi: 10.1016/j.jsp.2012.12.006

Litow, L. & Pumroy, D. K. (1975). A brief review of classroom group-oriented contingencies. *Journal of Applied Behavior Analysis, 8*, 341-347. doi: 10.1901/jaba.1975.8-341

McBurney, D.H. & White, T. L. (2007) Control. In *Research Methods* (pp. 202-205). Belmont, CA: Wadsworth.

McCurdy, B. L., Lannie, A. L., & Barnabas, E. (2009). Reducing disruptive behavior in an urban school cafeteria: An extension of the Good Behavior Game. *Journal of School Psychology, 47*, 39-54. doi: 10.1016/j.jsp.2008.09.003

McGoey, K. E., Schnieder, D. L., Rezzetano, K. M., Prodan, T., & Tankersley, M. (2010). Classwide intervention to manage disruptive behavior in the kindergarten classroom. *Journal of Applied School Psychology, 26*, 247-261. doi: 10.1080/15377903.2010.495916

Medland, M. B., & Stachnik, T. J. (1972). Good-Behavior Game: A replication and systematic analysis. *Journal of Applied Behavior Analysis, 5*, 45-51. doi: 10.1901/jaba.1972.5-45

Ringdahl, J. E., Vollmer, T. R., Borrero, J. C., & Connell, J. E. (2001). Fixed-time schedule effects as a function of baseline reinforcement rate. *Journal of Applied Behavior Analysis, 34*, 1-15. doi: 10.1901/jaba.2001.34-1

Saigh, P. A., & Umnar, A. M. (1983). The effects of Good Behavior Game on the disruptive behavior of Sudanese elementary school students. *Journal of Applied Behavioral Analysis, 16*, 339-344. doi: 10.1901/jaba.1983.16-339

Swiezy, N. B., Matson, J. L., & Box, P. (1992). The Good Behavior Game: A token reinforcement system for preschoolers. *Child and Family Behavior Therapy, 13*, 21-32. http://dx.doi.org/10.1300/J019v14n03_02

Tangorra, C. Jr. (2015). *Expulsion and suspension of children, ages 3-5, from regular early childhood and special education programs.* Retrieved from New York State Department of Education website:

http://www.regents.nysed.gov/common/regents/files/meetings/May%202015/515p12d3.pdf

Tanol, G., Johnson, L., McComas, J., & Cote, E. (2010). Responding to rule violations or rule following: A comparison of two versions of the Good Behavior Game with kindergarten students. *Journal of School Psychology, 48*, 337-355. doi: 10.1016/j.jsp.2010.06.001

Zaghlawan, H. Y., & Ostrosky, M. M. (2011). Circle time: An exploratory study of activities and challenging behavior in head start classrooms. *Early Childhood Education Journal, 38*, 439-448. doi: 10.1007/s10643-010-0431-z

Figures

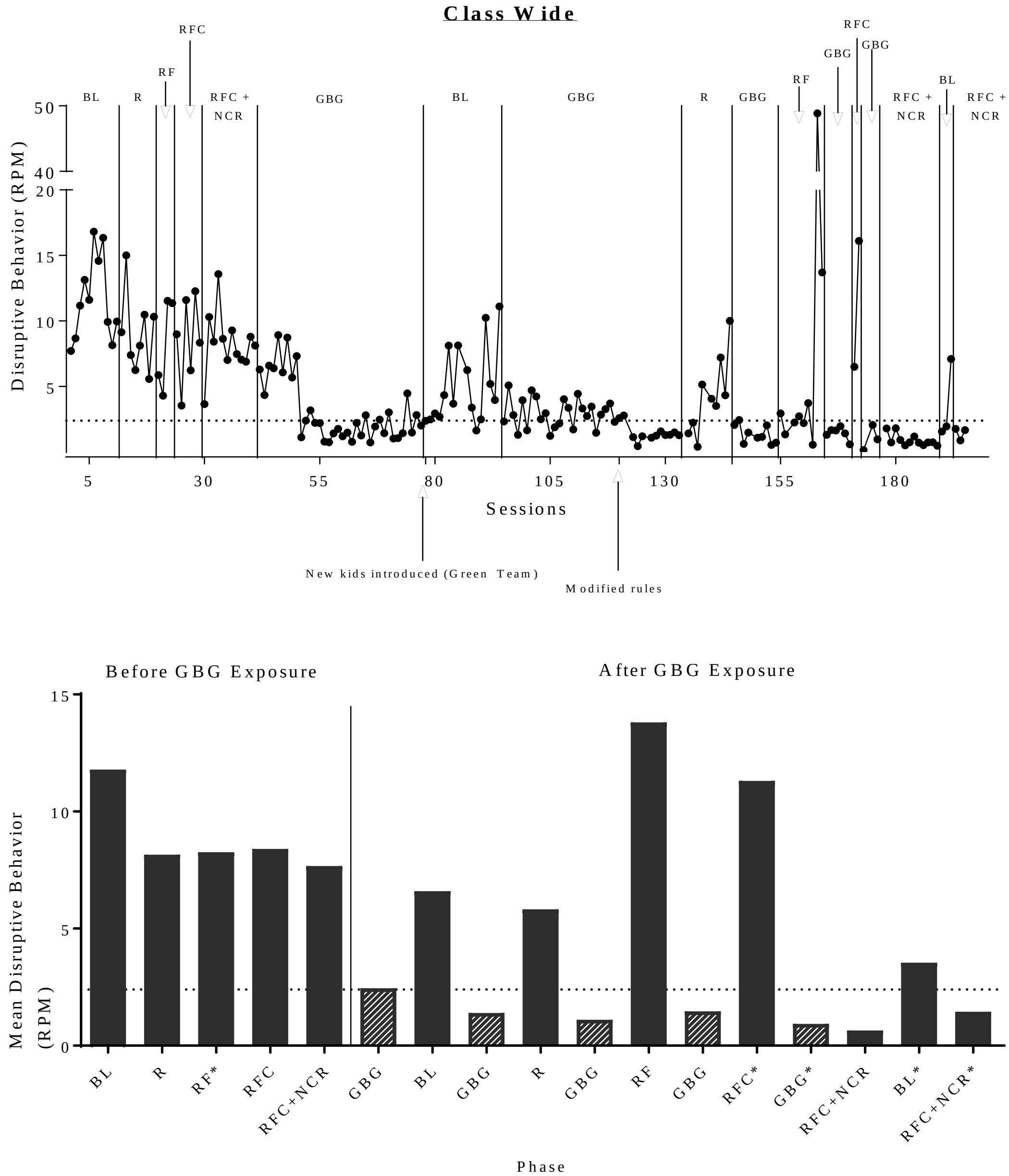


Figure 1. The top panel depicts the class-wide rate of disruptive behavior across all sessions. The bottom panel depicts the mean rate of disruptive behavior for the last five sessions in each phase. An asterisk denotes phases in which there was less than five sessions. The dotted line on both graphs represents an 80% reduction in disruptive behavior from baseline levels. BL=Baseline, R=Rules, RF=Rules + Feedback, RFC=Rules + Feedback + Criterion, RFC+NCR= Rules + Feedback + Criterion + Noncontingent Reinforcement, GBG=Good Behavior Game.

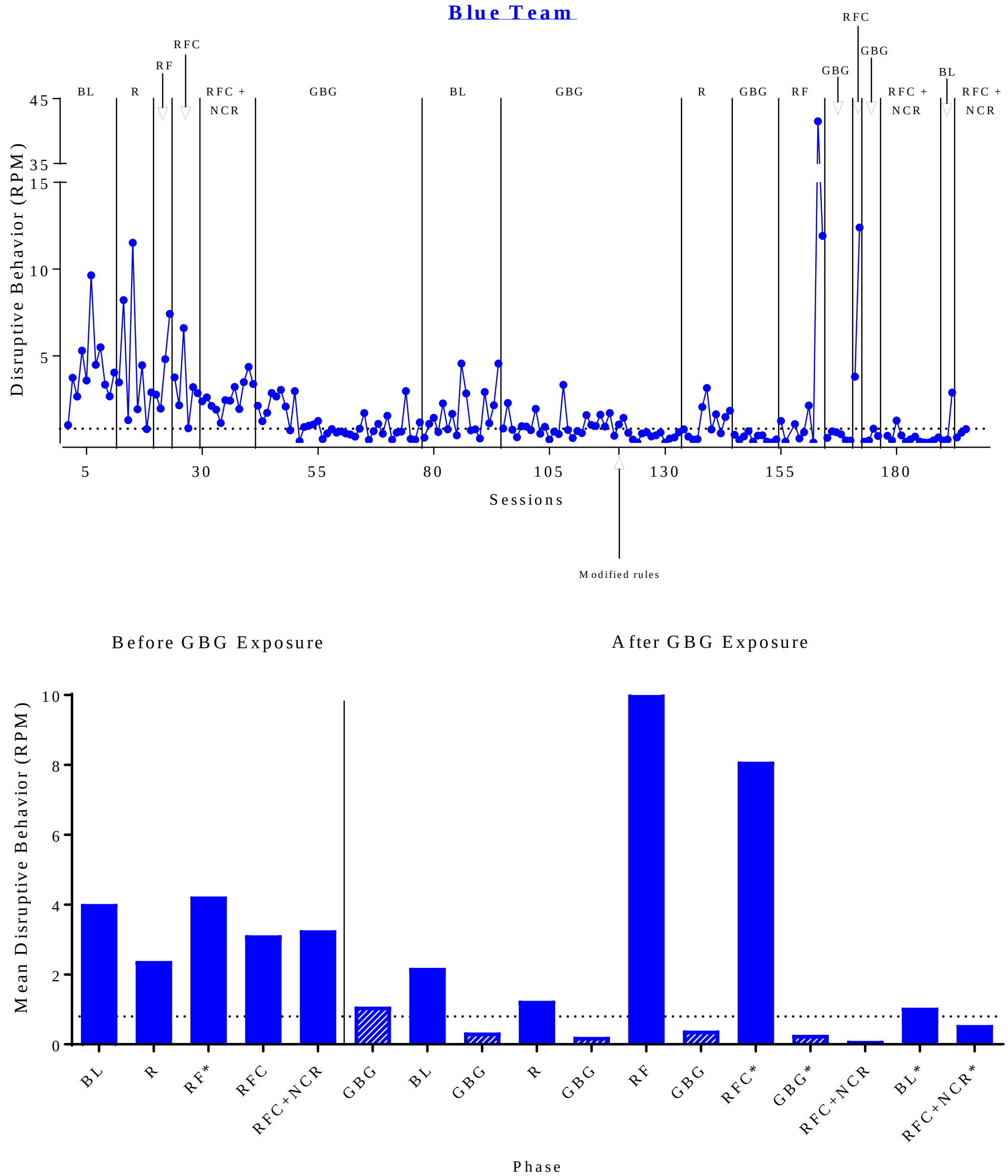


Figure 2. The top panel depicts the Blue Team rate of disruptive behavior across all sessions. The bottom panel depicts the mean rate of disruptive behavior for the last five sessions in each phase. An asterisk denotes phases in which there was less than five sessions. The dotted line on both graphs represents an 80% reduction in disruptive behavior from baseline levels. BL=Baseline, R=Rules, RF=Rules + Feedback, RFC=Rules + Feedback + Criterion, RFC+NCR= Rules + Feedback + Criterion + Noncontingent Reinforcement, GBG=Good Behavior Game.

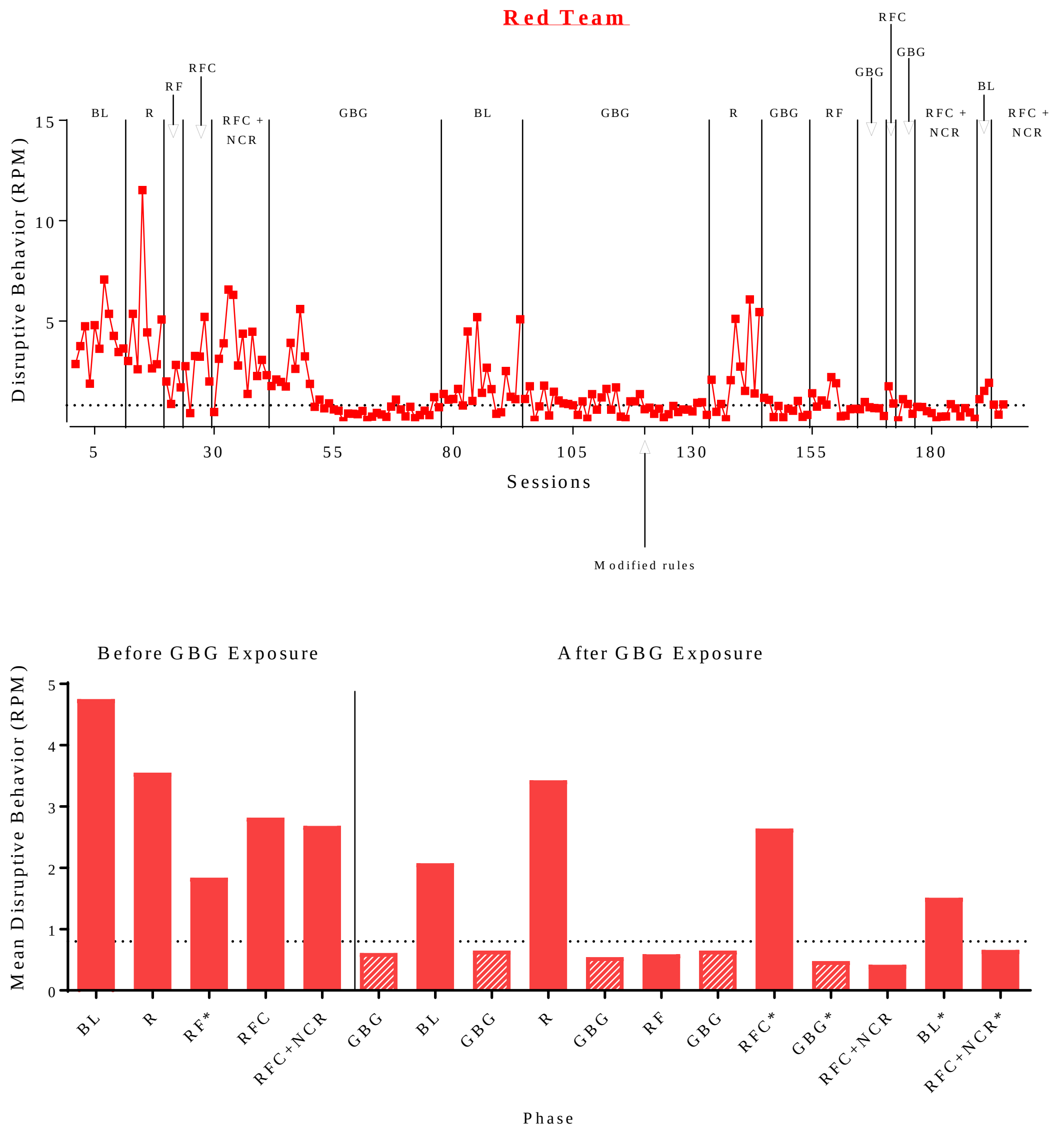


Figure 3. The top panel depicts the Red Team rate of disruptive behavior across all sessions. The bottom panel depicts the mean rate of disruptive behavior for the last five sessions in each phase. An asterisk denotes phases in which there was less than five sessions. The dotted line on both graphs represents an 80% reduction in disruptive behavior from baseline levels. BL=Baseline, R=Rules, RF=Rules + Feedback, RFC=Rules + Feedback + Criterion, RFC+NCR= Rules + Feedback + Criterion + Noncontingent Reinforcement, GBG=Good Behavior Game.

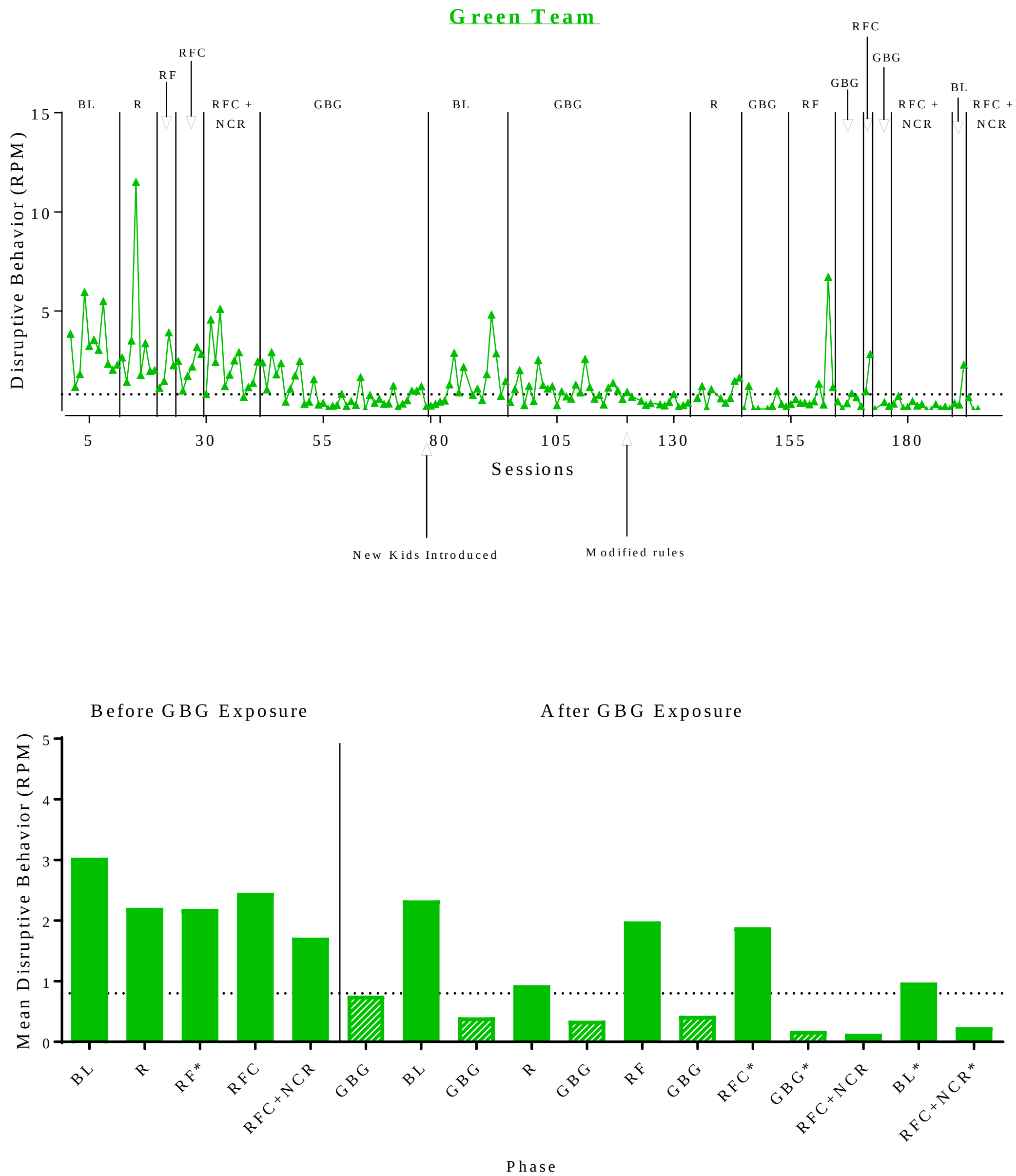


Figure 4. The top panel depicts the Green Team rate of disruptive behavior across all sessions. The bottom panel depicts the mean rate of disruptive behavior for the last five sessions in each phase. An asterisk denotes phases in which there was less than five sessions. The dotted line on both graphs represents an 80% reduction in disruptive behavior from baseline levels. BL=Baseline, R=Rules, RF=Rules + Feedback, RFC=Rules + Feedback + Criterion, RFC+NCR= Rules + Feedback + Criterion + Noncontingent Reinforcement, GBG=Good Behavior Game.

Tables

Percentage Change from BL in Mean Rate of Disruptive Behavior during GBG Phases

Teams	Phase				
Individuals	GBG 1	GBG 2	GBG 3	GBG 4	GBG 5
Blue Team					
Brenden	-100%	-90.9%	-100%	-100%	-100%
David	-92.6%	-85.2%	-100%	-100%	-55.6%
Lester	-74.1%	-96.6%	-94.0%	-92.7%	-96.1%
Tony	-84.4%	-93.3%	-100%	-95.6%	-100%
Lucas	-52.8%	-77.8%	-88.9%	-58.3%	-91.7%
Charlie	27.3%	-27.3%	-81.8%	-68.2%	-40.9%
Red Team					
Xander	-98.2%	-98.8%	-95.8%	-88.6%	-89.2%
Matthew	-93.5%	-95.7%	-82.6%	-97.8%	-97.8%
Etta	-94.1%	-82.4%	-76.5%	-85.3%	-100%
Andrew	-92.9%	-94.0%	-100%	-94.0%	-97.6%
James	-92.9%	-82.1%	-96.4%	-98.2%	-100%
Quilla	-70%	-65.4%	-80%	-76.9%	-77.7%
Daria	-88.5%	-84.6%	-94.2%	-78.8%	-63.5%
Green Team					
Noah	-89.9%	-85.3%	-84.4%	-88.1%	-88.1%
Megan	-34.9%	-90.5%	-87.3%	-82.5%	-93.7%
Harrison	-46.2%	-89.7%	-97.4%	-82.1%	-100%
Aiden	36.4%	27.3%	-45.5%	0%	-63.6%

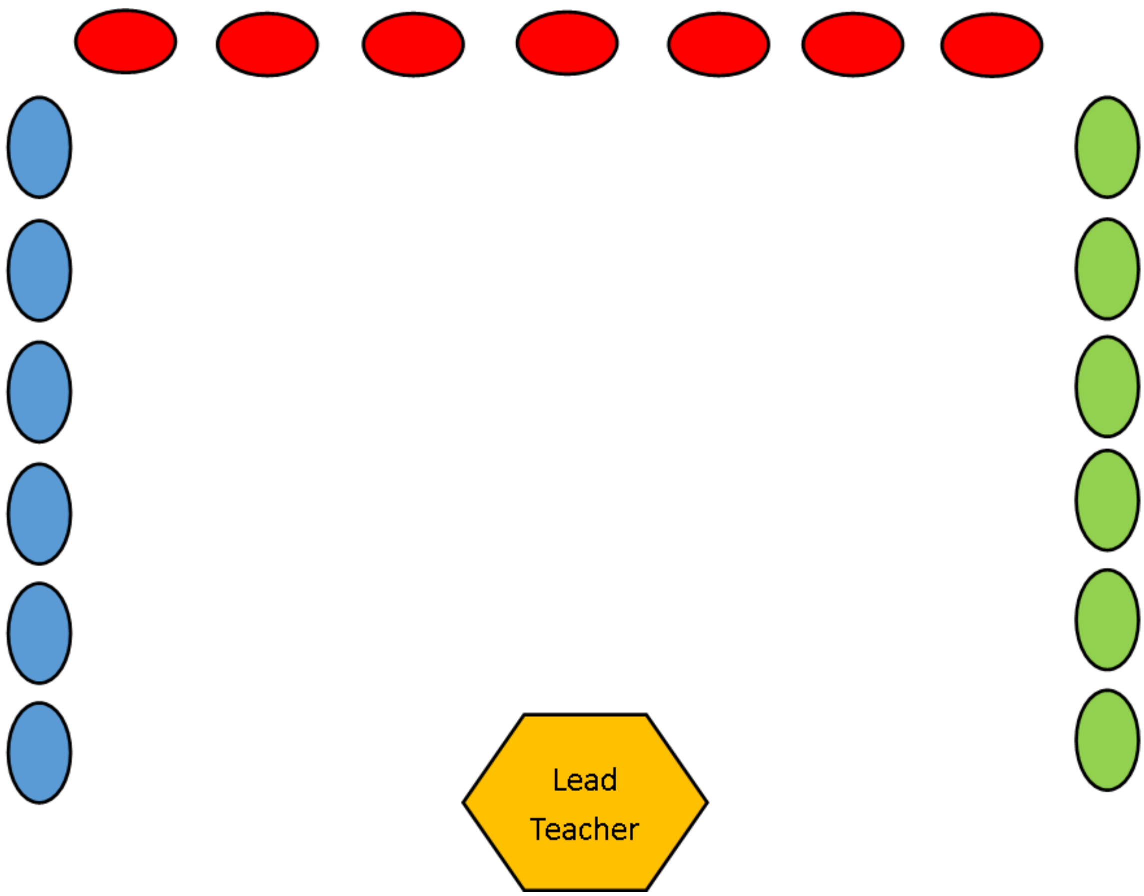
Table 1. Table 1 depicts the percentage change in the mean rate of disruptive behavior as compared to the mean rate in the initial baseline phase in each of the components before and after a history with the GBG. A positive number indicates an increase in the mean rate; a negative number indicates a decrease in the mean rate.

Percentage Change from BL in Mean Rate of Disruptive Behavior Before and After GBG

Team	Component Phases								
	R		RF		RFC		RFC+NCR		
Individual	Before	After	Before	After	Before	After	Before	After (1)	After (2)
Blue Team									
Brenden	-45.5%	-72.7%	45.5%	-100%	-9.1%	-72.7%	-45.5%	-100%	-100%
David	-25.9%	74.1%	-85.2%	-25.9%	192.6%	1,503%	263%	-96.3%	--88.9%
Lester	-51.3%	-85.3%	18.5%	345.7%	-46.6%	49.1%	-49.6%	-98.3%	-86.2%
Tony	-86.7%	-93.3%	-80%	-100%	-86.7%	-93.3%	-60%	-100%	-93.3%
Lucas	-8.3%	-41.7%	44.4%	-100%	2.8%	-77.8%	38.9%	-86.1%	-72.2%
Charlie	40.9%	13.6%	9.1%	290.9%	31.8%	59.1%	63.6%	-100%	-68.2%
Red Team									
Xander	-50.6%	-81.3%	-69.3%	-84.9%	-51.2%	-97%	-53.6%	-88.6%	-81.3%
Matthew	-41.3%	-47.8%	-65.2%	-95.7%	-65.2%	-89.1%	-67.4%	-100%	-100%
Etta	70.6%	-58.8%	-85.3%	-88.2%	-91.2%	-55.9%	-82.4%	-94.1%	-94.1%
Andrew	-76.2%	-84.5%	-82.1%	-82.1%	-95.2%	-96.4%	-97.6%	-96.4%	-100%
James	-30.4%	-69.6%	-75%	-92.9%	-71.4%	-67.9%	-67.9%	-94.6%	-100%
Quilla	-36.2%	79.2%	-33.1%	-90%	51.5%	-46.9%	13.8%	-92.3%	-74.6%
Daria	21.2%	-48.1%	-100%	-75%	-92.3%	-65.4%	-94.2%	-86.5%	-78.8%
Green Team									
Noah	-22.9%	-81.7%	-3.7%	39.4%	9.2%	11%	-32.1%	-95.4%	-100%
Megan	-23.8%	-95.2%	-71.4%	-81%	-73%	-84.1%	-84.1%	-96.8%	-95.2%
Harrison	-48.7%	-48.7%	20.5%	-64.1%	48.7%	-35.9%	-53.8%	-100%	-100%
Aiden	-18.2%	273%	54.5%	118.2%	18.2%	172.7%	134.7%	53.9%	-8.3%

Table 2. Table 2 depicts the percentage change in the mean rate of disruptive behavior as compared to the mean rate in the initial baseline phase in each of the components before and after a history with the GBG. A positive number indicates an increase in the mean rate; a negative number indicates a decrease in the mean rate.

Appendix A



Appendix B

Circle Time Rules
1 Sit on your mat
2 Raise your hand
3 Hands to yourself

Appendix C

Appendix D