THE GOOD BEHAVIOR GAME: MAINTENANCE AND SIDE-EFFECTS IN PRESCHOOLERS

By

Elizabeth A. Foley

Submitted to the graduate degree program in The Department of Applied Behavioral Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairperson Claudia L. Dozier, Ph.D.

Pamela L. Neidert, Ph.D.

Vincent T. Francisco, Ph.D.

Edward K. Morris, Ph.D.

Jose Rafael Martinez, Ph.D.

Date Defended: 25 July 2019

The Dissertation Committee for Elizabeth A. Foley

certifies that this is the approved version of the following dissertation:

THE GOOD BEHAVIOR GAME: MAINTENANCE AND SIDE-EFFECTS IN PRESCHOOLERS

Chairperson Claudia L. Dozier, Ph.D.

Date approved: 25 July 2019

Abstract

The Good Behavior Game (GBG) is an effective intervention which has been used by teachers to change a variety of behaviors, across a range of populations, and in various settings (see Tingstrom, Sterling-Turner, & Wilczynski, 2006, for a review). However, there is limited research on the intervention with preschoolers (Foley et al., 2019; Ling & Barnett, 2013; Wiskow et al., 2019, Swiezy et al., 1992), uninterrupted treatment effects when the GBG is faded or removed (Dadakhodjaeve et al, 2019; Lynch & Keenan, 2018; Ruiz-Olivares et al., 2010), and the potential side-effects associated with the GBG (Groves & Austin, 2019). Therefore, the purpose of the current study was to replicate Foley et al. (2019) by implementing the GBG with preschoolers and further evaluating whether effortful components of the GBG can be faded while maintaining treatment effects. Additionally, we conducted a systematic analysis of positive and negative peer interactions as a potential side effect of the GBG. Results of the study suggest that the GBG is an effective intervention for reducing the disruptive behavior of preschoolers. Additionally, we were able to fade several components of the GBG while maintaining treatment effects. Results of the side-effect analysis suggest that the GBG is associated with an increase in peer interactions and specific variables (i.e., the occurrence of target behavior and delivery of reward) were associated with specific types of peer interactions.

Acknowledgements

To my lab mates, thank you for your collaboration on this project. I could not have conducted this project without your hard work, suggestions, and support. To my husband, Ed Dugan, I will never forget the look on your face when I said I wanted to move to Kansas to pursue my graduate degree. You knew it was my dream, and you supported me through all of the ups and downs. Thank you does not begin to express my gratitude for all you do. To my children, Emmett and Savanna, thank you for the perspective and joy you provided to me during graduate school. To my parents, Michael and Margaret Foley, thank you for always supporting me. Words cannot express how grateful I am for all the sacrifices you made to support me on this journey. Finally, to my advisor, Dr. Claudia Dozier, I am beyond honored to have worked with you. I am especially thankful for all you have invested in me. You have supported me as a scientist, practitioner, teacher, and human. You have served as the best model in some of the most difficult situations. I hope I always make you proud and I will always be proud to be part of the Dozier Lab.

Table of Contents

The Good Behavior Game: Maintenance and Side-effects in Preschoolers

Teachers rank classroom management as one of their greatest professional development needs (Coalition for Psychology in the Schools and Education, 2006). Teachers may use a variety of individual strategies to increase appropriate behavior or decrease inappropriate behavior in their classrooms. For example, a teacher may provide praise (e.g., Burnett & Mandel, 2010; Fullerton, Conroy, & Correa, 2009) or implement guided compliance (e.g., Cote, Thompson, & McKerchar, 2005; Wilder & Atwell, 2006) with an individual student to increase compliance for a single student. However, manipulating these individual contingencies would be inefficient and cumbersome if a teacher needed to modify compliance for several students. Interventions using group contingencies offer a more efficient and practical method for changing the behavior of a group of individuals (Bushell, Wrobel, & Michaelis, 1968; Maggin, Johnson, Chafouleas, Ruberto, & Berggren, 2012; Stage & Quiroz, 1997).

Group Contingencies

Group contingencies involve applying a uniform performance criterion, a common consequence, or both to a group of individuals (Bushell et al., 1968; Litow & Pumroy, 1975). Three common group contingencies include dependent, independent, and interdependent group contingencies (Litow & Pumroy, 1975). In a dependent group contingency, a common consequence is delivered to a group of individuals based on the behavior of one member or a small subset of the group. Thus, a performance criterion is applied to a single individual and a common consequence is delivered to the group. In the independent group contingency, a common consequence is delivered to each member of the group based on their individual performance. That is, a uniform performance criterion is applied to each individual in a group and delivery of a common consequence is contingent on each individual meeting that performance criterion. Finally, the interdependent group contingency involves providing a common consequence to all members of a group based on the entire group meeting a performance criterion. That is, the entire group must meet a specified response criterion in order to access a common consequence. Thus, each individual's performance contributes to the group's performance.

Each type of group contingency offers benefits and has limitations. For example, a dependent group contingency may put undue pressure on the individual in the group who must meet the performance criterion in order for the group to receive a reward; however, it may be easier for a teacher to implement as they only need to monitor the behavior of a single individual (Gresham & Gresham, 1982; Kelshaw-Levering, Sterling-Turner, Henry, & Skinner, 2000; Little, Akin-Little, & O'Neill, 2015). Independent group contingencies allow a teacher to target the behavior of all members of a group; however, they do not promote cooperation and other group skills like other group contingencies (Little et al., 2015; Slavin, 1977). The interdependent group contingency is useful because it (a) does not put undue pressure on a single individual (or subgroup), (b) allows a teacher to target the behavior of all members of a group, and (c) may promote cooperation and collaboration (Little et al., 2015). However, interdependent group contingencies also have limitations. First, behavior change across individuals in the group can vary due to unequal contributions across group members. For example, a single star player on a basketball team may be responsible for scoring all the points in a championship game; however, all players would receive a trophy for winning. Second, interdependent contingencies may evoke negative peer perceptions and interactions between students. For example, a player on that same team who does not contribute to the group's success may be reprimanded for the group's failure. However, more research is needed to support this assumption.

The Good Behavior Game

One group-contingency intervention is the Good Behavior Game (GBG; Barrish, Saunders, & Wolf, 1969; Flower, McKenna, Bunuan, Muething, & Vega, 2014; Harris & Sherman, 1973), which is a multi-component intervention that uses an interdependent group contingency and has been demonstrated to be effective for decreasing disruptive behavior and increasing appropriate behavior. Although there are several variations of the GBG (e.g., Galbraith & Normand, 2017; Maloney & Hopkins, 1973; Radley, Dart, & O'Handley, 2016; Robertshaw, & Hiebert, 1973), components primarily consist of (a) dividing a group or class into teams, (b) establishing rules describing clear behavioral expectations, (c) providing feedback for rule violations, (d) establishing a criterion for winning the game, (e) delivering consequences (e.g., rewards) to winning team(s) at the end of some activity or specified period, and (f) an interdependent group contingency.

Barrish et al. (1969) first evaluated the effects of the GBG on the disruptive behavior of students in a fourth-grade classroom during math instruction. First, the teacher divided the class into two teams. Second, the teacher reviewed the rules for the math instruction period in the form of clear behavioral expectations. Third, the teacher explained the hatch marks, winning criterion, interdependent contingency, and reward delivery (i.e., the rules for playing and winning the GBG). Specifically, the teacher explained that every time any child on a team engaged in a rule violation, the teacher would place a hatch mark by the team's name on the chalkboard. The teacher explained that if the team's hatch marks exceeded the winning criterion (i.e., no more than 5 hatch marks), they would not win the game and receive a reward (i.e., victory tags and special privileges). However, if all teams exceeded the criterion, then the team with the fewest number of hatch marks would win. The GBG effectively decreased disruptive behavior every time it was implemented.

Since this first evaluation of the GBG, numerous replications have demonstrated its efficacy (see Embry, 2002 and Tingstrom, Sterling-Turner, & Wilczynski, 2006, for reviews). That is, the GBG has been shown to decrease disruptive behavior (e.g., Barrish et al., 1969; Donaldson, Matter, & Wiskow, 2018; Harris & Sherman, 1973; Kosiec, Czernicki, & McLaughlin, 1986; Rubow, Vollmer, & Joslyn, 2018) and increase appropriate behavior (e.g., Fishbein & Wasik, 1981; Johnson, Turner, & Konarski, 1978; Ling & Barnett; Maloney & Hopkins, 1973; Scott et al., 2017) across a variety of settings (e.g., Fishbein & Wasik, 1981; Galbraith & Normand, 2017; McCurdy, Lannie, & Barnabas, 2009), populations (e.g., Bostow & Geiger, 1976; Cheatham, Ozga, St. Peter, Mesches, & Owsiany, 2017; Lastapes, 2013), and cultures (e.g., Leflot, van Lier, Onghena & Kolpin, 2010; Ruiz-Olivares, Pino, & Herruzo, 2010; Saigh & Umar, 1983). Furthermore, numerous studies have demonstrated both the short-term (e.g., Bostow & Geiger, 1976; Gresham & Gresham, 1982; Groves & Austin, 2019; Harris & Sherman, 1976; Hegerle, Kesecker, & Couch, 1979; Joslyn, Vollmer, & Hernández, 2014; Medland & Stachnik, 1972) and long-term (e.g., Dolan et al., 1993; Kellam, Ling, Merisca, Brown, & Ialongo, 1998; Kellam et al., 2011; Kellam et al. 2014) effects of the multicomponent intervention. Short-term effects include decreasing inappropriate behavior (e.g., talking out of turn, getting out of seat, aggression, and swearing) and increasing appropriate behavior (e.g., on-task behavior, worksheet completion, and accuracy of task completion). Long-term effects include lower levels of aggression (as reported by teachers) and a decreased likelihood of engaging in risky sexual behavior, becoming dependent on drugs, becoming incarcerated, engaging in suicide ideation, or being diagnosed with an antisocial personality disorder. These long-term outcomes associated with the GBG have led the Centers for Disease Control and

Prevention (CDC) to identify the GBG as a universal violence prevention program (CDC, 2017). Additionally, Embry (2002) has suggested the GBG is a behavioral vaccine (Embry, 2002). Specifically, Embry argues that the GBG is a relatively inexpensive intervention that can be applied to a variety of populations while decreasing the likelihood of a variety of adverse outcomes (e.g., substance abuse and juvenile crime). Although the long-term outcomes of the GBG are promising, it is unclear why these results are observed. Given these studies only provide correlations between long-term outcomes and experience with the GBG in childhood, the exact mechanisms for behavior change are unclear.

The Good Behavior Game with Young Children

The GBG has been shown to be effective for behavior change in a variety of populations including very young students (e.g., Donaldson,et al., 2011; Foley et al., 2019; Ling & Barnett, 2013), grade-school students (e.g., Barrish et al., 1969; Medland & Stachnik, 1972), high-school students (e.g., Flower, McKenna, Muething, Bryant, & Bryant, 2013; Kleinman & Saigh, 2013), college students (Cheatham et al., 2017), and state hospital residents (Lutzker & White-Blackburn, 1979). However, given the long-term effects of the GBG (e.g., Dolan et al., 1993), it may be important to evaluate the efficacy of the GBG with children as they first enter school.

Donaldson and colleagues (Donaldson et al., 2011; Donaldson et al., 2015; Donaldson et al., 2018) evaluated the effects of the GBG on the disruptive behavior of kindergarteners during structured learning times (e.g., math instruction) across several studies. Donaldson et al. (2011) evaluated the GBG across five kindergarten classrooms during math and reading instruction. The GBG effectively reduced inappropriate behavior across all five classrooms. Additionally, the researchers were able to transfer implementation of the GBG from the researchers to the teachers in all the classrooms. In a later study, Donaldson et al. (2015) evaluated the effects of the GBG during periods immediately before, during, and after implementation of the GBG across five kindergarten classrooms in order to determine the effects of the GBG on non-targeted contexts. That is, it is possible that the GBG could have positive effects on non-targeted contexts if the appropriate behaviors occurring in the intervention context generalize to other activities (e.g., appropriate behavior reinforced during math instruction generalized to reading instruction), or the GBG could have negative effects on non-targeted contexts due to the lack of availability of reinforcement. Results of the study suggest the GBG did not worsen behavior in non-targeted contexts. That is, levels of inappropriate behavior in the non-targeted contexts remained at baseline levels when the GBG was implemented in the target context. Thus, it was necessary for teachers to implement the GBG in order to decrease inappropriate behavior in the previously non-targeted contexts. Most recently, Donaldson et al. (2018) evaluated preference for teacher- or student-led implementation of the GBG during whole-group instruction in three kindergarten classrooms and one first-grade classroom. Both the teacher-led and student-led versions were effective for decreasing disruptive behavior; however, the student-led version required teacher prompts to tally rule-violations. Preference varied across classrooms. Two of the classrooms preferred the student-led version, one class preferred the teacher-led version, and one class preferred not to play the GBG. Overall, results of the evaluations with kindergarten students suggest the GBG is effective with this young population.

Although several studies have shown the utility of the GBG with kindergarten students, it is important to also evaluate its effects with even younger students (i.e., preschoolers). Evaluations of the GBG with preschoolers may be particularly important. First, the number of children enrolled in preschool is steadily increasing (National Center for Education Statistics, 2018). Thus, preschool is often the first school setting young children experience. Second, research suggests that problem behavior that originates in early childhood persists if not treated (Breitenstein, Hill, & Gross, 2009; Del'Homme, Sinclair, & Kasari, 1994). Thus, the GBG may provide preschool teachers with an intervention to address problem behavior early in a child's school experience.

Swiezy, Matson, and Box (1992) offer one of the first evaluations of the GBG with preschool students. The researchers used a modified version of the GBG that involved a puppet to deliver instructions to a dyad of preschoolers during free play. If only one child complied with the instruction, that child received praise from the puppet. If both students cooperated to complete the instruction, then the dyad received a token. Results of the study showed the GBG increased compliance and cooperation, as compared to baseline, across both dyads. Although compliance and cooperation are important behaviors to increase, studies suggest that problem behavior in preschool often occurs during structured learning times (Zaghlawan & Ostrosky, 2011). Thus, it is beneficial to evaluate the GBG during structured learning times in preschool settings.

Only a few studies (Foley, Dozier, & Lessor, 2019; Ling & Barnett, 2013; Wiskow, Matter, & Donaldson, 2019) have evaluated the GBG with preschoolers during structured group instruction (i.e., circle time). These studies demonstrated the GBG was effective for reducing disruptive behavior during structured group instruction. For example, Ling and Barnett (2013) evaluated the effects of the GBG in two preschool classrooms. The researchers collected data on disruptive behavior and engagement with the lesson; however, engagement was not specifically targeted. Implementation of the GBG resulted in a decrease in disruptive behavior and an increase in engagement every time researchers implemented the intervention.

Recently, Wiskow et al. (2019) conducted an evaluation of the GBG in two preschool classrooms during whole-group instruction. The researchers collected data on disruptive behavior (e.g., talking out, out of seat, and manipulating objects unrelated to the lesson). The researchers compared the GBG with three different types of feedback (i.e., visual feedback only, vocal feedback only, and visual + vocal feedback, described in detail in the component analyses section) and found the GBG with vocal feedback only and with both vocal and visual feedback were similarly effective at decreasing the disruptive behavior of preschool students during whole-group instruction.

Foley et al. (2019) also evaluated the GBG during circle time with preschoolers. The researchers collected data on disruptive behavior (i.e., inappropriate vocalizations, out of seat, and touching other students) at the individual level (17 students) and aggregated the data for team (three teams) and class-level analyses. Overall, results demonstrated that the GBG resulted in an 80% decrease in disruptive behavior as compared to baseline levels at the class and team levels. That is, an 80% decrease was observed for the class and all three teams. At the individual level, an 80% decrease from baseline levels of disruptive behavior was observed for 11 to 15 students across the implementations of the GBG. Although evaluations of the GBG with preschoolers show promising results, more replications are needed to strengthen the support for the use of the GBG with this population. Additionally, the results of Foley et al. and Wiskow et al. suggest that components can be manipulated (e.g. pictorial rules) and removed (e.g. visual feedback) while maintaining the efficacy of the GBG. Given the GBG is a multicomponent treatment package, an important consideration involves identifying which components are necessary and sufficient for behavior change. Additionally, it is important to consider how modifications to specific components influence the efficacy of the GBG.

The Good Behavior Game: Component Analyses

Traditionally, the GBG involves (a) dividing a class or group into teams, (b) establishing rules which describe behavioral expectations, (c) providing feedback (e.g. tallying hatch marks on a chalk board) for rule violations, (d) stabling a uniform criterion for winning, and (e) an interdependent group contingency to determine reward delivery. In the first study to determine the influence of various components of the GBG, Medland and Stachnik (1972) evaluated the rules and feedback components of the GBG with two fifth grade reading groups. The researchers used lights to deliver feedback to students. A green light was illuminated when the team was engaged in appropriate behavior; however, if an individual on the team engaged in a rule violation the green light was turned off and a red light was illuminated. Results of the component analysis suggested rules alone were not sufficient for behavior change, but rules and feedback produced similar results to the GBG.

In another study, Harris and Sherman (1973) evaluated the effects of several components of the GBG on the disruptive behavior (i.e., talking out of turn, out of seat, and throwing objects) of students from a fifth and sixth grade classroom. After a history with the entire GBG package, the researchers manipulated (a) the programmed rewards, (b) the winning criterion, (c) delivery of feedback, and (d) the number of teams. In the programmed rewards manipulation, the class was divided into two teams, rules for appropriate behavior were described, feedback for rule violations was delivered on the chalkboard, and a criterion for the number of rule violations was in place; however, there were no programmed consequences (i.e., rewards) for winning the game. During this condition, levels of disruptive behavior were similar to those observed in baseline. Next the researchers manipulated the winning criterion. For a portion of this condition, the winning criterion was set at either four or eight hatch marks in order to win the game. Results of this manipulation showed that disruptive behavior fluctuated as a function of the winning criterion. That is, when the winning criterion was set at four hatch marks or fewer, lower levels of disruptive behavior were observed as compared to when the winning criterion was set at eight hatch marks or fewer. Next, the researchers manipulated the feedback component. In this condition, the teacher recorded hatch marks for rule violations on a small piece of paper hidden behind a podium instead of recording on a chalkboard where the teams could see the hatch marks. Results of this manipulation showed that levels of disruptive behavior were similar to those observed when the entire GBG package was implemented. Finally, the researchers manipulated the number of teams competing by treating the class as a single team. Results of this manipulation were mixed. That is, during math class period, reductions in disruptive behavior were similar to those observed when the entire GBG package was implemented; however, similar reductions in disruptive behavior were not observed during the English class period.

Overall, results of Medland and Stachnik (1972) and Harris and Sherman (1973) suggest that some components of the GBG may be more necessary than other components. Harris and Sherman found that rewards and winning criterion were important components; however, results were mixed in the evaluation on the number of teams, and the researchers did not observe a decrease in efficacy when feedback was delivered covertly. Medland and Stachnik found rules alone did not produce a large decrease in disruptive behavior, but rules plus feedback did. The results of Harris and Sherman appear to conflict with those of Medland and Stachnik. That is, they seem to suggest that feedback was not an important component because when feedback was delivered covertly (the teacher tallied rule violations on a small piece of paper behind a podium) reductions in disruptive behavior were similar to those observed under the GBG. However, in both studies the different modalities of feedback may have exerted stimulus control (Gamzu & Schwartz, 1973). In Medland and Stachnik's evaluation of the GBG, the lights may have acquired stimulus control after a history with the GBG package. That is, the green light may have signaled the availability of reinforcement because of a previous history in which appropriate classroom behavior was reinforced in the presence of the light. Furthermore, the red light may have been conditioned as a punisher (Azrin & Holz, 1966; Hake & Azrin, 1965) after a history in which the presence of the red light signaled loss of access to rewards. A similar process may have occurred in Harris and Sherman; however, the teacher writing on paper behind the podium functioned as the discriminative stimulus. These results suggest that it may be important to evaluate components prior to and after they have been paired with the entire GBG package. That is, the efficacy of various components may be altered by repeated exposure to the contingency used in the GBG.

Foley et al. (2019) evaluated components of the GBG before and after exposure to the GBG package on the disruptive behavior (i.e., inappropriate verbal behavior, out-of-seat, and touch other students) of a group of preschool students during circle time. Researchers collected and analyzed data at the class, team, and individual level. The study included three teams and the researchers evaluated results for 17 individuals. The researchers used a reversal design to conduct an additive component analysis (Ward-Horner & Sturmey, 2010) to evaluate the efficacy of components for behavior change. That is, researchers did not evaluate the components in isolation; instead they were added based on ease of implementation. Specifically, the researchers evaluated (a) rules alone; (b) rules plus feedback; (c) rules, feedback, plus a winning criterion; and (d) rules, feedback, a winning criterion, plus noncontingent delivery of rewards. During the rules-only phase, the researchers presented a rules poster board with written rules and pictures of children following the rules, reviewed the rules with the children, and had the children repeat the rules back to the experimenter. During the rules plus feedback phase, the researchers followed the same procedures as the rules-only phase; however, they used a dry-erase board to deliver tallies to teams when any member of the team engaged in a rule violation. In the rules, feedback, plus winning criterion phase, the researchers followed the same procedure as described in the rules plus feedback phase; however, they added a yellow line using colored tape to the dry erase board in order to depict a winning criterion. The researchers did this to increase the saliency of the criterion due to the participants limited numeracy skills. At the end of sessions in this phase, researchers announced if each team had won (i.e., their hatch marks did not go past the yellow line); however, no rewards were delivered for winning. In the rules, feedback, winning criterion, plus noncontingent delivery of rewards phase, researchers followed the procedures described for the previous phase; however, rewards were delivered to all children after circle regardless of behavior during circle. This phase served as a control for the interdependent group contingency. That is, the noncontingent delivery of rewards eliminates the contingency while stimuli associated with the contingency (i.e., rewards and discriminative stimuli) continue to be presented (Thompson et al., 2003). Results showed that prior to a history with the GBG package, some components or some combination of components were effective for some individuals. However, an 80% reduction from baseline levels was never observed in any component phase prior to implementation of the GBG at the team or class level. After all components were evaluated prior to exposure to the GBG, the researchers implemented the GBG. Implementation of the GBG was associated with large decreases in disruptive behavior at the class, team, and individual level. After the students had a history with the GBG, the component phases were evaluated again, in the same additive fashion. Rules and rules plus feedback resulted in an 80% reduction from baseline levels for some teams and individuals; however, the rules, feedback, winning criterion, and noncontingent rewards (RFC+NCR) phase demonstrated similar effects to the GBG package for the class, all teams, and most individuals. That is, the response-independent schedule used in the RFC+NCR phase maintained the treatment effects observed under the response-dependent schedule (i.e., the interdependent group contingency) used in the GBG. Overall, results of the additive analysis suggest that a history with the GBG package influenced the efficacy of the various components. The continuation of treatment effects observed in the RFC+NCR phase replicates the results of previous 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). Continued treatment effects under a response-independent schedule may be due to adventitious reinforcement (e.g., Ecott & Critchfield, 2004) or stimulus control (Gamzu & Schwartz, 1973; Terrace, 1966). That is, appropriate behavior may have continued to occur due to adventitious reinforcement if the delivery of rewards occurred in close temporal proximity to these behaviors even though there was not a dependent relation between the reward and the behavior. Alternatively, the stimuli associated with the GBG were present during the RFC+NCR phase. That is, these stimuli may have acquired stimulus control because behavior was differentially reinforced in the presence of the GBG stimuli (e.g., team mats, experimenters, feedback, and treasure box) which may have exerted control over responding. Regardless of the mechanism responsible for maintenance during the RFC+NCR phase, these results suggest that it may be possible to fade components of the GBG without interrupting treatment effects. However, the researchers were only able to conduct a brief replication of this phase. Thus, additional replications are needed to determine how long behavior would maintain under RFC+NCR.

Additionally, the results of Foley et al. (2019) should be replicated with more populations in order to determine if results were specific to this population. That is, even though researchers did not observe an effect in a particular component phase with a preschool population, that same component (or group of components) could be effective with a different population. For example, rules, feedback, and a winning criterion may be effective with an older population that, presumably, has a longer history playing games or sports. For example, just being declared the winning team may be more effective with an older population as compared to this younger population.

Results of Foley et al. (2019), Medland and Stachnik (1972), and Harris and Sherman (1973) suggest exposure to the GBG package may be used to increase the efficacy of certain components. This has important implications because some components of the GBG may be easier to implement than others. For example, presenting rules, briefly noting some rule violations, and delivering rewards to a group may require less effort than scanning the class for rule violations, providing vocal and visual feedback, and correctly tallying rule violations. Thus, it may be possible to implement the GBG for a period of time, then fade some of the more effortful components making implementation of the GBG easier for teachers who are attempting to deliver effective instruction to a class while implementing the intervention. Additionally, feedback appears to be an important component in that all three of the component analysis studies discussed above observed an effect with the feedback component. It is particularly interesting that three different modalities of feedback were effective. That is, lights, covert feedback, and overt feedback (tallies on a dry-erase board) were all effective. These results suggest that it may be important to compare how different modalities of feedback influence the efficacy of the GBG.

Recently, Wiskow et al. (2019) compared different modalities of feedback within the context of an evaluation of the GBG in two preschool classrooms. Specifically, the researchers compared (a) no feedback, (b) vocal feedback only, (c) visual feedback only, and (d) vocal and visual feedback combined. The researchers collected data on the students' disruptive behaviors (i.e., talking out, out of seat, and manipulating items unrelated to lesson). First, they evaluated the GBG with no feedback. In this condition, the teacher discretely recorded rule violations on a whiteboard out of view from the students. There was no difference in the level of disruptive behavior in this condition as compared to baseline. Next, the researchers used a multielement design to evaluate the GBG with vocal feedback only, visual feedback only, and vocal and visual feedback combined. In the vocal feedback only condition, the teacher provided vocal feedback for rule violations (e.g., "Team 1, sit criss-cross") but discretely recorded hatch marks on a whiteboard. In the visual feedback only condition, the teacher did not provide any vocal feedback for rule violations but recorded a hatch mark on a whiteboard that was visible to the class. In the vocal and visual feedback condition, the teacher provided vocal feedback for rule violations and recorded a hatch mark on a whiteboard that was visible to the class. Overall, results suggested feedback was a necessary component. That is, there was no difference in the level of disruptive behavior observed in baseline as compared to GBG with no feedback. Additionally, GBG with visual feedback alone was not effective. However, GBG with vocal feedback and GBG with vocal and visual feedback were both effective. Considering the results of Foley et al. (2019), Harris and Sherman (1973) and Medland and Stachnik (1972), these results may be due to exposure to the GBG package. The researchers used a multi-element design, in which the conditions were rapidly alternated. Therefore, the vocal delivery of feedback may have acquired control over responding after repeated pairings with the vocal plus visual feedback. That is, excessive hatch marks likely became discriminative for loss of access to rewards. After repeatedly pairing the vocal feedback with the visual feedback, the vocal feedback may have acquired control over responding. These results suggest that occasional exposure, like those in a multielement design, may be sufficient for transferring stimulus control to less effortful versions of components (i.e., vocal feedback is easier to implement than visual feedback and requires less materials).

The various types of feedback (e.g., covert, overt, vocal, and visual) may be an important component of the GBG for several reasons. First, feedback provided contingent upon disruptive behavior may function as a conditioned punisher (Hake & Azrin, 1965; Honig, 1966) because the feedback (e.g., hatch marks) is paired with loss of rewards. Second, feedback may serve as an antecedent for rule-governed behavior (Haas & Hayes, 2006; Johnson, Houmanfar, & Smith, 2010). That is, participants receiving feedback may generate rules about behavior after receiving the feedback. For example, if a child receives feedback for shouting out an answer, they may generate a rule such as “If I shout out the answers, then I get a point for my team.” Finally, the feedback may also serve as a prompt for correct future performance (Aljadeff-Abergel, Peterson, Wiskirchen, Hagen, & Cole, 2017). Additional research on modifications of feedback should be conducted in order to identify the most efficient and effective forms of feedback within the context of the GBG.

Side-effects of the Good Behavior Game: Students

Another important consideration when implementing the GBG is the potential side-effects that have been reported to occur. That is, researchers have provided anecdotal reports of changes in student behavior (e.g., Barrish et al., 1969; Fishbein & Waskik, 1981; Groves & Austin, 2017; Harris & Sherman, 1973) even when those behaviors are not specifically targeted by the GBG. Specifically, several studies have included anecdotal reports of positive and negative peer interactions. Positive interactions that have been reported include students encouraging each other to follow the rules (e.g., Groves & Austin, 2017), congratulating peers for improved performance (e.g., Packard, 1970), and commenting on reward delivery (e.g., Barrish et al., 1969). Additionally, negative peer interaction that have been reported include students harassing the losing team (e.g., Harris & Sherman, 1973), complaining about team members (e.g., Mitchell et al., 2015), and complaining about loss of rewards (e.g., Barrish et al., 1969). Systematic analyses of peer interactions are needed to provide insight on how the GBG contributes to their occurrence and how they may influence the efficacy of the GBG. Additionally, some teachers report that they are hesitant to use the GBG because of the potential for negative peer interactions due to the interdependent group contingency (e.g., Fishbein & Wasik, 1981; Groves & Austin, 2017; Mitchell et al., 2015). Thus, evaluation on side-effects would allow researchers to determine if the GBG is associated with positive, negative, or neutral side-effects.

To date, only three studies have collected and reported data on side-effects associated with group contingencies or the GBG on student behavior (Groves & Austin, 2019; Payne, Dozier, Briggs, & Newquist, 2017; Speltz et al., 1982). Speltz et al. (1982) compared the effects of different group contingencies (i.e., independent, interdependent, dependent with known or unknown target student) on the accurate completion of math problems. That is, the number of correct math problems determined reward delivery during the group contingency comparison. Additionally, the researchers collected data on positive (e.g., compliments, assistance, statements of concern), neutral (i.e., general discussion about the day or current activity), and negative (e.g., name calling, verbal aggression, obscene gestures) peer-to-peer interactions. However, these behaviors were not specifically targeted. The different group contingencies were equally effective in increasing the number of correct math problems completed. Results of the side effect analysis showed that the dependent (known target student) group contingency was associated with the highest amount of peer interactions. Specifically, this condition was associated with a high level of positive peer interactions.

Payne et al. (2017) provide a strong methodology for conducting analyses of potential side-effects associated with group contingencies. The researchers evaluated the occurrence of positive and negative vocalizations associated with three group contingencies (i.e., independent, interdependent, and dependent) across two groups of five preschool students. The researchers collected data on two target behaviors: on-task behavior and problem behavior. One or both behaviors were specifically targeted during the various group contingency conditions. Additionally, researchers collected data on positive and negative vocalizations to determine if any specific group contingency was associated with a higher rate of these potential side-effects. Positive vocalizations included praise or encouragement of on-task behavior of another child, statements about earning rewards, and vocal niceties (e.g., "I like you"). Negative vocalizations included reprimands delivered to another child, statements about losing a reinforcer, and vocal admonitions (e.g., "I don't like you"). The researchers conducted several analyses to determine the variables that may influence the occurrence of positive and negative vocalizations. They evaluated the mean rate of positive and negative vocalizations (a) across each of the group contingencies, (b) across sessions with high rates and low rates of problem behavior, and (c) across sessions with and without reinforcer delivery. Although results varied across both groups of preschool students, some general conclusions were found. First, the group contingencies were generally associated with an increase in positive vocalizations as compared to baseline regardless of whether problem behavior or on-task behavior were targeted. Furthermore, when problem behavior was targeted, the highest level of positive vocalizations occurred in the dependent contingency and the highest level of negative vocalizations occurred in the interdependent contingency. However, when on-task behavior was targeted, the highest level of positive vocalizations was observed in the independent contingency and the highest level of negative vocalizations was observed in the dependent and interdependent contingencies. Second, negative vocalizations were higher in sessions with high rates of problem behavior; however positive vocalizations did not differ much across sessions with high and low rates of problem behavior. Finally, positive vocalizations were generally higher when a reinforcer was delivered, and negative vocalizations were higher when no reinforcer was delivered. Overall, these outcomes suggest group contingencies may influence the occurrence of positive and negative vocalizations exhibited by individuals experiencing the group contingency. Additionally, certain variables may have differential effects on positive and negative vocalizations. Specifically, reward delivery (or the absence of reward delivery) and problem behavior were associated with a higher frequency of positive and negative vocalizations. These results suggest there may be several environmental variables that are associated with the side-effects that are reported anecdotally in the GBG and group contingency literature and not just the group contingency.

Most recently, Groves and Austin (2019) conducted an analysis to determine if positive and negative peer interactions increase when the GBG is implemented. The researchers implemented the GBG in two classrooms in an alternative school setting and specifically asked teachers not to target peer interactions. Researchers collected data on positive and negative peer interactions. Positive peer interactions included encouraging peers, congratulating peers, and requesting or offering assistance from peers. Negative peer interactions included name calling, laughing at a peer's mistake, threats or acts of aggression, and interfering with a peer's work. In both classrooms, negative peer interactions decreased when the GBG was implemented. In the first classroom, a slight increase in positive peer interactions was observed; however, a large increase in positive interactions was observed in the second classroom when the GBG was implemented.

Overall, results of these evaluations are mixed. However, they all provide insight into how the GBG or the interdependent group contingency used in the GBG may influence peer interactions. That is, Speltz et al. (1982) did not observe an effect on peer interaction when the interdependent contingency was in place, Payne et al. (2017) observed negative peer interactions when the interdependent contingency was implemented, and Groves and Austin (2019) observed a decrease in negative peer interactions when the GBG was implemented. The differences in these findings may be due to one of two things. First, Speltz et al. and Payne et al. both compared several different group contingencies in session rooms with a single group of students, whereas Groves and Austin evaluated the GBG in a more applied setting (i.e., a classroom in an alternative school setting) with several groups competing against each other. The difference in these experimental arrangements may have contributed to the difference in responding. That is, the competition across groups was not present in Spelt et al. and Payne et al., which may have contributed to the limited positive interactions between peers. The competition between groups may be necessary to evoke these behaviors. Second, the populations across studies were different and may have contributed to the different results observed in each study. Payne et al. included preschool students; Speltz et al. included 7- to 10-year-olds from a class for individuals with learning disabilities; and Groves and Austin included 9- to 16-year-olds in an alternative school setting for students with behavioral issues or learning disabilities. It may be that that the preschool students were still developing some of the skills associated with the positive outcomes associated found by Groves and Austin. For example, these students may still be learning to work in groups, cooperate, and provide reinforcement to peers. Systematic replication across populations of these studies would be needed to determine if the results are in any part due to the populations used in these studies.

Results of Payne et al. validate some of the concerns with group contingencies. That is, negative vocalizations were associated with dependent and interdependent group-contingencies. As Litow and Pumroy (1975) discussed, a dependent group contingency may put undue pressure on a single individual and an interdependent group contingency may increase negative peer interaction for students who contribute to the loss of access to rewards. However, given the results of Payne et al., it may be that negative vocalizations occur more frequently when a group contingency is first implemented as peers may reprimand or admonish behaviors which result in the loss of access to a reward; and positive vocalizations may increase and negative interactions may decrease after the target behavior change has occurred and reinforcers are being earned on a regular basis. That is, Payne et al. found the overall mean frequency of negative vocalizations was highest when there were high rates of problem behavior and no reinforcer was earned. Session-by-session analysis may provide a finer analysis of whether this pattern would be observed with the GBG in an applied setting (e.g., a classroom).

Maintenance of the Good Behavior Game

There is a large body of literature demonstrating the efficacy of the GBG; however, there is limited research on the maintenance of treatment effects of the GBG. Maintenance refers to the durability of treatment effects over time (Baer, Wolf, & Risley, 1968), and little research has systematically planned for and evaluated maintenance effects associated with the GBG

(Dadakhodjaeva, Radley, Tingstrom, Dufrene, & Dart, 2019; Lynch & Keenan, 2018; Ruiz-Olivares, Pino, Herruzo, 2010). Additionally, few studies have conducted follow up probes to assess maintenance of treatment effects in the absence of the intervention (Flower et al., 2014; Johnson et al., 2018).

Ruiz-Olivares et al. (2010) demonstrated maintenance of GBG treatment effects by gradually withdrawing the game while implementing say-do correspondence (SDC) training for appropriate behaviors. Specifically, the researchers implemented SDC training immediately before and immediately after GBG sessions. The SDC procedure prior to GBG sessions involved asking the students if they were going to follow the rules, then praising the students if they said they were going to follow the rules. If the student said they were not going to follow the rules or did not answer, the researchers prompted a correct response. For example, the researcher would ask the student, "Are you going to shout?" and the students would respond, "No we are not going to shout." The SDC procedure following GBG sessions involved asking the students if they had followed the rules. For example, the researcher would ask the student, "Have you shouted?" and the students would respond "No we have not shouted." The researchers were able to remove the GBG and effects maintained at a one-year follow-up.

Lynch and Keenan (2018) also attempted to fade the GBG while maintaining treatment effects. The researchers used a reversal design to evaluate effects of the GBG on disruptive behavior (e.g., talking out, out of seat, and turning around in chair) of 11- to 12-year-olds during History, English, and Geography instruction. The researchers trained teachers to provide praise to participants for appropriate behavior during final baseline and GBG phases. The researchers hoped that the praise delivered by the teachers would function as reinforcement for appropriate behavior and the teachers would become discriminative stimuli for the delivery of reinforcement for appropriate behavior (e.g., sitting in chair and oriented to task materials), which could maintain appropriate behavior in the absence of the intervention. However, results were mixed likely because of the variability of positive comments delivered by the teachers. That is, not all teachers increased positive comments as instructed. Additionally, the authors reported that teachers continued to implement the GBG outside of experimental sessions (i.e., during different classroom contexts), which may have influenced responding during the reversal to baseline.

Recently, Dadakhodjaeva et al. (2019) evaluated maintenance of treatment effects of the GBG by reducing the frequency of implementation. After collecting baseline data, the researchers used a multiple baseline across classrooms design to evaluate the effects of daily implementation of the GBG on disruptive behavior (e.g., talking out, out of seat, and touching others) and academic engagement in three kindergarten classrooms. After a treatment effect was observed with daily implementation of the GBG, the researchers thinned the schedule of GBG implementation by instructing teachers to only implement the intervention once or twice a week. The researchers instructed the teachers to select implementation days randomly such that the students would not know when it was being played. Teachers were instructed to still review the rules for appropriate behavior (clear behavioral expectation) with the students on the days that the GBG was not played. Additionally, teachers were instructed to respond to student requests for the game by stating that it would not be played today, but maybe tomorrow. Data were collected on implementation and non-implementation days. Results showed disruptive behavior remained at low levels and academic engagement remained high during both implementation and non-implementation days. These results suggest it may be possible to reduce the effort of the GBG by reducing the frequency of implementation. However, when the researchers interviewed the teachers who implemented the GBG in their classroom in order to determine acceptability of the intervention, the teachers reported that they would prefer to implement the GBG initially, then only implement portions of the GBG. Thus, although the results of the study are promising, it may be more important to determine how to maintain treatment effects of the GBG while decreasing the number of components implemented.

Both Flower et al. (2014) and Johnson et al. (2018) conducted brief follow-up phases to assess whether treatment effects maintained in the absence of the GBG. Flower et al. evaluated effects of the GBG on off-task (e.g., not attending to instruction) behavior of high school students in a special education resource room. Johnson et al. evaluated effects of the GBG on disruptive behavior (e.g., talking out of turn, out of seat, and cursing) of third- and fourth-grade students. During the follow-up phases in both studies, the teachers had discontinued use of the GBG. Results of both studies showed that treatment effects did not maintain in the absence of the GBG. These results and the results of Ruiz-Olivares et al. (2010), Lynch and Keenan (2018), and Dadakhodjaeva et al. (2019) suggest that researchers need to explicitly plan for maintenance of treatment effects. That is, Ruiz-Olivares et al. implemented SDC training with the participants, Lynch and Keenan implemented teacher training in order to increase the likelihood that appropriate behavior would come under stimulus control of the teacher's presence, and Dadakhodjaeva et al. thinned the schedule of implementation of the GBG. However, given the results of the social validity assessment of Dadakhodjaeva et al., research on maintenance may need to focus on reducing the number of GBG components implemented in order to maintain treatment effects. Results from the studies on component analyses of the GBG (Harris & Sherman, 1973; Foley et al., 2019; Medland & Stachnik, 1972) suggest that after a history with the GBG, various components associated with the GBG may acquire stimulus control over student behavior. Thus, it may be possible to implement the GBG, then implement only a few components while maintaining treatment effects.

Purpose

Although some research has demonstrated the efficacy of the GBG with preschoolers (Foley et al., 2019; Ling & Barnett, 2013; Wiskow et al., 2019, Swiezy et al., 1992), additional replications are needed to increase the credibility of the GBG with this population. Additionally, there is little research on maintaining treatment effects when the GBG is removed (Lynch & Keenan, 2018; Ruiz-Olivares et al., 2010). Although there is some support for maintaining effects with a reduced frequency of GBG implementation (Dadakhodjaeva et al., 2019), it may be more acceptable to determine ways to fade intervention components while maintaining treatment effects (Dadakhodjaeva et al., 2019). Finally, there is limited research on how the GBG influences peer interactions that may be due to the interdependent group contingency used in the GBG (Groves & Austin, 2019). Evaluation of potential side-effects of the GBG, such as peer interactions, may provide insight into the mechanisms responsible for both short-term and long-term effects observed in the GBG literature.

The primary purpose of the current study was to replicate the GBG with preschoolers and evaluate a potential method for fading effortful components of the GBG while maintaining treatment effects. First, we replicated Foley et al. (2019) by replicating the response-independent schedule (i.e., the RFC+NCR phase) to determine if treatment effects would maintain over time. Second, when treatment effects maintained under a response-independent schedule, we faded some of the more effortful components of the good behavior game (e.g., delivery of hatch marks for rule violations) to determine if treatment effects would continue to maintain. As a secondary purpose, we conducted a systematic analysis of positive and negative peer interactions as a potential side effect of the interdependent group contingency used in the GBG.

Method

Participants

The study was conducted in a university-based preschool classroom located at the University of Kansas serving typically developing children ranging in age from 3.5 to 5 years. Over the course of the study, 17 typically developing children participated. The experimenter divided the group of children into three teams. Each team consisted of 4 to 6 students. Prior to beginning the study, researchers determined teams by collecting data on the frequency of the target behaviors (described below) for each individual child during group instruction (i.e., circle time). For each of these observations, the researchers summed the total frequency of the target behaviors and divided by the duration of the circle time to calculate the rate of the target behaviors for each individual child. Next, the researchers calculated the average rate of the target behaviors for each child by summing the rates for each circle-time observation and dividing by the total number of observations. The researchers then created a rank-order list of the children based on the average rate of the target behaviors. The child with the highest average rate of target behaviors was ranked at the top of the list; the child with the lowest average rate of target behaviors was ranked at the bottom of the list. Next, the researchers used a matched allocation procedure (McBurney & White, 2007) to form the teams. Specifically, each of the top three children on the ranked list were quasi-randomly assigned to one of the three groups. Researchers then repeated this process with the next three children on the ranked list, and so on, until all children were assigned to a team.

Initially 17 children were included in the study. At the beginning of the study, the Red Team had six children (ranked 1^{st}, 6^{th}, 8^{th}, 10^{th}, 15^{th}, and 17^{th}); the Green Team had five children (ranked 3^{rd}, 4^{th}, 7^{th}, 12^{th}, 13^{th}, and 16^{th}); and the Blue Team had five children (ranked 2^{nd}, 5^{th}, 9^{th}, 11th, and 14th). Some of the children had previous exposure to the GBG because teachers implemented the GBG as part of the preschool's behavior management procedures. That is, teachers implemented the GBG after less intensive behavior management procedures were ineffective at decreasing disruptive behavior during circle time. Specifically, three children on the Red Team (Michael, Elanor, and Mina), three children on the Green Team (Sasha, Everett, and Murphy), and one child on the Blue Team (Layla) had previous exposure to the GBG. After session 122, a child from the Red Team (George) and a child from the Green Team (Murphy) moved. A child from the Blue Team, Connor, was dropped from the study after session 88 because the researchers did not observe a consistent treatment effect for this individual. The researchers made several manipulations in an attempt to demonstrate a treatment effect with Connor. Fist, prior to session 74, the lead experimenter reminded teachers to be sure to implement general classroom behavior management (e.g., proactive attention, differential reinforcement of alternative behavior (DRA), and descriptive praise) because observations suggested they were not consistently implementing behavior management procedures that were typically provided to the children during circle time. Once the teachers resumed typical behavior management practices and a treatment effect was still not observed for Connor, the researchers provided pre-session training with Connor prior to sessions 83-87. The training was modeled after behavioral skills training (BST; Parsons, Rollyson, & Reid, 2012) and involved individual instruction, rehearsal, and feedback. First, the researcher provided Connor with individualized instruction on how to play the GBG. These were the same rules presented to the group (described in the procedures section below); however, they were presented to Connor individually prior to sessions. Second, the researcher prompted Connor to engage in rule following behavior and rule-violation behavior. For example, the researcher would prompt

Connor to raise his hand quietly (rule-following behavior) and raise his hand while shouting "Excuse me!" (rule-violation behavior). Third, the researcher provided praise for rehearsals of rule-following behavior and delivered feedback (in the form of hatch marks) for rehearsals of rule-violations. For example, if Connor raised his hand quietly the researcher provided praise. When he raised his hand and shouted, "Excuse me," he received a hatch mark for the rule violation (described in the procedures section below). After 5 sessions, the researchers stopped providing the pre-session training because Connor's individual data suggested that it was not effective. Prior to session 88, we moved him to his own team. Researchers continued to collect individual data on his behavior; however, his data were no longer included in the class or Blue Team data.

Setting and Materials

The preschool classroom is staffed by three student teachers and a graduate-student supervisor during a morning and an afternoon shift. All sessions were conducted during circle time. Teachers conducted circle time once per shift for approximately 10 to 15 min. During circle time, a teacher led an activity (e.g., singing songs, telling stories, sorting activities, and gross motor activities) that fit within a specific theme (e.g., animals and habitats) and targeted programmed curricular skills (e.g., identifying colors). Children sat in a semi-circle facing the teacher and engage in individual and choral responses during circle. During circle-time, each child sat in an assigned position on a colored mat that corresponded to their team color. See Appendix A for a sample seating chart. Additionally, a bin of children's books was used during pre- and post-session periods.

A poster board with written rules and a picture of a child following the rules was presented to the children prior to all sessions except during baseline. The written rules included raise your hand to talk, keep your hands to yourself, and sit on your mat. See Appendix B for a depiction of the poster board. During some sessions, a dry-erase board was used to record hatch marks. The dry erase board was designed to increase the saliency of the hatch marks and the winning criterion. Specifically, the researchers made small square boxes on the dry erase board on which researchers recorded hatch marks. Additionally, the winning criterion was denoted by a thick yellow line placed below the maximum number of rule violations allowed to occur while still winning the game. The winning criterion in the current study was 11 hatch marks; therefore, the yellow line was placed below the first 11 boxes and above the remaining boxes. See Appendix C for a depiction of the dry-erase board. Rewards were delivered at the end of some sessions and included tangible items (e.g., small toys) and individually wrapped edible items (i.e., different types of candies) that were kept in a plain plastic bin.

Data Collection, Data Analysis, and Interobserver Agreement

During circle-time sessions and pre- and post-session time periods, trained observers collected data on the target behaviors displayed by each child on a team using paper-pencil data collection methods. During circle-time sessions, observers collected data on target behavior (i.e., out of seat, inappropriate verbal behavior, and touching others) for each individual child on each team using a frequency within 1-min intervals measure (see Appendix D for a sample data sheet for a team). *Out of seat* was scored if the child's bottom raises off their mat in the absence of a teacher's instruction to do so or if the child is seated but turned away from the teacher (e.g., sitting backward or lying flat on their back). *Inappropriate verbal behavior (IVB)* was scored if a child vocalizes in the absence of a teacher's instruction. This will include yelling out, talking to others, humming, tongue clicking, crying, laughing, and other disruptive vocalizations. For less discrete vocalizations, a new occurrence was scored if the child pauses the vocalization for at least 2 s. For more discrete vocalizations, such as lip smacking, each discrete occurrence was counted. IVB was not scored if the children engage in expressive reactions to teacher interactions. For example, if the teacher makes a joke or uses a silly voice and a child laughs, data collectors did not score IVB. *Touching other students* was scored if the child's hands, feet, elbows, or other body part contacted another student without being instructed to do so. Touching others was not scored if the teacher delivered an instruction which involved making contact with another child (e.g., passing another child an object or giving a high-five) or contact occurred as a result of teacher instruction (e.g., the child falls when attempting to follow the instruction "Balance on one foot.").

Data during circle-time was analyzed as a rate measure at three levels (i.e., individual child level, team level, and whole class level). At the individual level, the rate of target behavior for an individual child was calculated by dividing the frequency of target behavior by the session duration. At the team level, the rate of target behavior for the team was calculated by dividing the frequency of target behavior for all children on the team by the session duration. At the class level, the rate of target behavior of the whole class was calculated by dividing the frequency of target behavior for all children in the class by the session duration. At the conclusion of the study, researchers also calculated the mean rate of disruptive behavior for each condition at the class, team, and individual level. That is, the rates of disruptive behavior for all sessions in a particular condition were summed and divided by the total number of sessions for those conditions conducted over the course of the study.

During pre- and post-session periods, observers collected data on the frequency of potential side-effects (i.e., positive and negative interactions) of the GBG (see Appendix E for a sample data sheet). A *positive interaction* was scored if an individual vocalized or made a statement to a peer that involved encouraging another peer (e.g., “You can do it.”), providing praise (e.g., “You did a good job following the rules at circle.”), commenting about rewards or reward delivery (e.g., “Yay, we get a prize today.”), congratulating another peer (e.g., “You won today!”), offering assistance (“I’ll help you at circle today.”), providing a rule reminder in an appropriate tone (e.g., “Remember to raise your hand in circle if you want to say something so we don’t get points.”), providing a supportive physical gesture (e.g., thumbs up, high fives, or pats on the back), or making statements of approval or friendship to their peers (e.g., “I like you.” or “I’m glad you’re on my team.”). A *negative interaction* was scored if an individual vocalized or made a statement to a peer that involved making a threat (e.g., “We won’t play with you if you make us lose today.”), insulting (e.g., “You are stupid”), or demeaning another child (e.g., “Nobody wants you on our team.”); providing a reprimand to another child (e.g., “Stop shouting out the answers.” or “We got points because you didn’t raise your hand”); telling the child how to behave in a demeaning tone (e.g., “Stay on your bottom!”); making a statement of concern about not receiving a reward (e.g., “Oh no, we won’t win a prize if you talk during circle.”); making statements of disapproval (e.g., “I don’t like you.” or “You’re not my friend anymore.”); and swearing, name calling, or engaging in a combative physical gesture (e.g., sticks tongue out at another child, or attempts to push or hit another child).

Data during pre- and post-session periods were analyzed at the class and team level. The researchers graphed the frequency of pre-session positive interactions, pre-session negative interactions, post-session positive interactions, and post-session negative interactions on a session-by-session basis at the class level. Furthermore, in order to compare the overall level of side-effects across phases, we calculated the mean frequency of positive and negative interactions that occurred in the pre- and post-session periods for each phase of the study. For example, we summed all of the positive interactions that occurred during the pre-session of all sessions conducted during the first baseline phase, then divided that number by the total number of sessions conducted in the first baseline phase. We repeated that process for all positive interactions that occurred in the post-session during the initial baseline. We then graphed these as a stacked bar graph in order to depict the mean rate of all positive interactions during the initial baseline. We repeated this calculation for negative interactions during the initial baseline phase. This analysis was conducted for each phase of the study at both the class and team level.

At the conclusion of the study we conducted several retrospective analyses to determine potential associations between the different side-effects (i.e., positive and negative interactions) and other variables (i.e., occurrence of target behavior, reward delivery, and type of reward contingency) at the class and team level. To determine the association between the occurrence of positive and negative interactions and the delivery of rewards, we calculated the mean frequency of positive and negative interactions that occurred in post-session periods associated with sessions in which rewards were (a) delivered during response-independent sessions (NCR sessions), (b) delivered during response contingent sessions (GBG), and (c) were not delivered (but were available for delivery) during GBG sessions at the team level. For example, to determine the mean frequency of positive interactions during post-session periods for all sessions in which rewards were available but not delivered to a team, we divided the frequency of positive interactions during post-session periods for all sessions in which a reward was available but not delivered to a team by the total number of GBG sessions.

To determine the association between the occurrence of positive and negative interactions and the level of disruptive behavior, we calculated the mean frequency of positive and negative interactions that occurred in post-session periods. We did this for sessions that occurred in intervention conditions (i.e., NCR and GBG sessions) in which high rates of disruptive behavior (i.e., less than an 80% reduction from baseline levels) occurred and low rates of disruptive behavior (i.e., an 80% or more reduction from baseline levels) occurred. We did this at both the class and team level. For example, to determine the mean frequency of positive interactions that occurred in the post-session period of intervention sessions with high levels of target behavior, we divided the frequency of positive interactions during post-session periods for all intervention sessions in which target behavior occurred at a less than 80% reduction from mean baseline levels by the total number of intervention sessions that also met that criterion.

A second observer simultaneously and independently collected data during for at least 33% of sessions in each phase (including pre- and post-sessions) across teams. For disruptive behavior, researchers calculated interobserver agreement (IOA) by dividing the session into 1-min intervals and calculating interval-by-interval agreement for each target behavior. That is, for each behavior, researchers 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 to convert the score to a percentage. Mean IOA for the Blue Team was 92% (range, 70% - 100%) for IVB, 94% (range, 78% - 100%) for out of seat, and 98% (range, 88% - 100%) for touching others. Mean IOA for the Red Team was 95% (range, 80% - 100%) for IVB, 96% (range, 84% - 100%) for out of seat, and 99% (range, 96% - 100%) for touching others. Mean IOA for the Green Team was 93% (range, 71%-100%) for IVB, 96% (range, 86% - 100%) for out of seat, and 99% (range, 84% - 100%) for touching others. IOA fell below 80% for two sessions and was associated with a single data collector. When this happened the lead experimenter provided a retraining that included reviewing the definitions and data collection procedures with the data collector.

For positive and negative interactions, researchers calculated IOA using the total method. That is, for each pre- and post-session, researchers summed the total frequency of positive and negative interactions scored by each observer, then divided the smaller frequency by the larger frequency, and multiplied by 100 to convert the score to a percentage. Mean IOA for the Blue Team was 79% (range, 0% - 100%) for positive interactions and 91% (range, 0% - 100%) for negative interactions. Mean IOA for the Red Team was 77% (range, 0% - 100%) for positive interactions and 82% (range, 0% - 100%) for negative interactions. Mean IOA for the Green Team was 76% (range, 0% - 100%) for positive interactions and 85% (range, 0% - 100%) for negative interactions. IOA for side-effects was at 0% for sessions in which only a few instances occurred. For example, if there was only a single occurrence during a session and that was missed by one of the data collectors, this resulted in an agreement of 0%.

General Procedures

All sessions were divided into three parts: pre-session, circle-time session, and post-session. Pre- and post-sessions were 3 min. Prior to all sessions, researchers used a seating chart (Appendix A) to set up the circle-time mats such that the children sat in the same spot across all sessions. That is, they laid out the color-correlated team mats in the designated area of the classroom according to the seating chart. Next, they set up a bin of approximately 25 children's books at the front of the circle. Any other stimuli (e.g., rules poster-board or dry-erase board) that was needed for the session was placed in the front of circle (i.e., next to the spot where the teacher sat to lead circle). Once all children were present, the lead experimenter passed out a single book to each child. The experimenter provided the rule that the children would get one book, it would be a surprise which book they received, and that they could not trade books. After all children had a book, the experimenter set a 3-min timer and cued the start of the pre-session by saying “3, 2, 1, start.” The data collector assigned to a specific team then began collecting data during the pre-session period. After 3 min elapsed, data collection stopped. The experimenter collected all the books and informed the lead teacher that circle could begin. Children’s books were used during pre- and post-sessions in order to mimic typical transition activities used in the classroom. That is, it would not be typical to have the children sit on their mats with nothing to do for 3 min. Typical classroom procedures involve allowing the children to select a book from the library and sit on their mat as they wait for a circle activity to begin.

Prior to beginning the circle-time session, the experimenter cued the beginning of the session by saying “3, 2, 1, start.” Data collectors started timers and began collecting data. During the circle-time session period, a teacher implemented a daily lesson plan that fit within a specific theme and targeted curricular skills. During the lesson, the teacher alternated between individual and choral responses while leading a circle activity. Additionally, the teachers supporting the lead teacher were trained to implement several behavior management strategies. Specifically, teachers were trained to implement antecedent rule reminders, differential reinforcement (e.g., praise) of appropriate behavior (e.g., sitting quietly), planned-ignoring, response blocking, and prompting strategies (e.g., least-to-most prompting). The teacher informed the experimenter when the circle activity ended and the experimenter then cued the end of the circle-time session by stating “3, 2, 1, end.”

Post-sessions were conducted immediately after the conclusion of the circle-time session. Across phases, post-session varied slightly; however, the following procedures were consistent across phases. During all post-sessions, the children were seated on their mats in their assigned position, and the experimenter gave each child a single book. After all children had a book, the experimenter started a 3-min timer and cued data collection to begin. After 3 min elapsed, data collectors discontinued data collection.

Post-sessions varied across baseline, GBG, and NCR phases. During post-sessions in which there was no reward delivery (i.e., baseline), the children sat on their mats for the 3-min period. After 3 min elapsed, data collectors discontinued data collection and the lead teacher transitioned the children to the next activity. During post-sessions in which reward delivery was response dependent (i.e., GBG sessions), after the children had books and the 3-min timer was started, the experimenter announced which teams won the game and would receive a reward, then began reward delivery. Data collectors discontinued data collection after 3-min elapsed; however, if needed, reward delivery continued until all children who earned a reward received a reward. After rewards were delivered, the lead teacher transitioned the children to the next activity. During post-session in which the reward delivery was response-independent (i.e., NCR sessions), after the children received books and the 3-min timer was started, the lead experimenter began reward delivery. Data collectors discontinued data collection after 3 min elapsed; however, if needed, reward delivery continued until all children present for circle had received a reward. After rewards were delivered, the lead teacher transitioned the children to the next activity.

Baseline. During baseline (BL), children sat with their teams, in their assigned seat, and the teacher conducted circle as described in the previous section. During baseline, teachers often provided a variety of antecedent rule reminders. As part of regular of the classroom's regular behavior management strategies, teachers were trained to use rules that stated clear behavioral expectations using "do" statements. That is, the rule would describe a desired behavior (e.g., "raise your hand if you want to talk") in contrast to telling the child a behavior that they should not engage in during circle (e.g., "don't shout out the answer"). For example, if the activity is sorting animals across three different habitats the teacher may have told the children to sit quietly and raise their hand if they want to take a turn to sort an animal. No other programmed manipulations were made. Researchers used baseline data to determine the winning criterion which was an 80% reduction from the mean frequency of disruptive behavior for the class during baseline averaged across the three teams. Specifically, the mean frequency of disruptive behavior for the class was 168.5 occurrences. This was divided by three in order to average the mean frequency across the three teams. An 80% reduction from this frequency was 11 occurrences of disruptive behavior which was used as the winning criterion for each team.

Good Behavior Game (GBG). Prior to all GBG sessions, the experimenter showed the children the poster board and reviewed the rules for circle and the GBG. First, the experimenter told the children,

> These are the rules for circle. The first rule is that we will sit with our bottoms on our mats facing the teacher (and point to the picture of the child sitting on their bottom facing forward). The second rule is that we have a quiet mouth and raise our hand if we need to ask a question or say something (and point to the picture of the child raising their hand)—we will raise our hands and wait for our teacher to call on us unless the teacher asks us to talk. The third rule is that we keep our hands to ourselves and in our laps (and point to the child's hand in their lap in the picture)—we will not touch other people during circle.

After reviewing the rules with the children, the experimenter asked the children to repeat each of the rules. That is, the experimenter pointed to each part of the picture that corresponded to a rule while asking the children to state the rule. The experimenter then set the rules poster near circle such that the children could see the rules poster during the session. During the initial GBG sessions, the experimenter told the children how to play the GBG. That is, while showing the children the modified dry-erase board, the experimenter informed the children that they would get a “point” on the board under their team’s section if they broke a rule during circle. Additionally, the experimenter provided exemplars and non-exemplars of behaviors that would and would not result in rule violations in order to ensure the children understood how and when hatch marks would be delivered. Next, the experimenter pointed to the thick yellow line denoting the winning criterion and said, “You don’t want to get points, because if your points go past this yellow line you might not get a prize at the end of circle. If your team’s marks do not go past the yellow line, then your team will definitely get a toy or treat at the end of circle. But if all of the teams go past the yellow line, then the team with the smallest number of points will win.”

At the end of GBG sessions, the experimenter showed the dry-erase board to each team and asked the children on each team if their hatch marks exceeded the criterion. For example, while showing the board to the Red Team, the experimenter would state “Red Team, did your points go past the yellow line?” If necessary, the experimenter would prompt a correct response. If the team’s hatch marks did not exceed the winning criterion the experimenter would tell the team “You will get a reward from the prize bin.” If the team’s point exceeded the yellow line, the experimenter would tell the team that they could try to win at the next session. The experimenter allowed each child on a winning team to select a reward from the prize bin after the dry-erase board was reviewed with each team.

Rules + feedback (hatch marks) + criterion+ noncontingent rewards. Prior to rules + feedback + criterion + noncontingent rewards (RFC+NCR) sessions, the experimenter reviewed the rules for circle while holding the rules poster as previously described. The experimenter did not review rules for receiving hatch marks or other rules associated with the GBG. If the children had asked about any of the missing components (e.g., discussion of the GBG rules or the contingency), the experimenter would have told the children to just try their best at circle today. However, this did not occur. At the end of each session, the experimenter set the dry-erase board down, then walked up to each child and allowed the child to select one prize from the prize bin. The experimenter did not review the dry-erase board or announce winners. The purpose of this condition was to extend Foley et al. (2019) by evaluating the maintained effects of RFC+NCR.

Rules + vocal feedback + noncontingent rewards. Prior to rules + vocal feedback + noncontingent rewards (RVF+NCR) sessions, the experimenter reviewed the rules for circle while holding the rules poster as previously described; however, during the session, the experimenter delivered vocal feedback alone for rule violations. Children did not receive hatch marks for rule violations on the dry-erase board and the dry-erase board depicting the winning criterion was not present. At the end of RVF+NCR sessions, the experimenter did not review the dry-erase board and did not announce winner; however, they did allow each child to pick a prize from the prize bin as described in the RFC+NCR phase. The purpose of this condition is to determine if it is possible to remove the more effortful components of the GBG (e.g., scoring tallies for rule violations) while maintaining treatment effects.

Experimental design. A reversal design was used to demonstrate experimental control. Phases included baseline, Good Behavior Game (GBG), Rules + Feedback + Criterion + Noncontingent Rewards (RFC+NCR), and Rules +Vocal Feedback + NCR (RVF+NCR).

Treatment integrity. During GBG, RFC+NCR, and RVF+NCR sessions, a data collector used a treatment integrity data sheet to record the degree to which the intervention was implemented correctly (see Appendix F for a sample data sheet). Each phase had a specific set of six components which was used to assess if the experimenter correctly implemented specific components or omitted specific components. For GBG sessions, the six components included: (a) present rules poster board to children, (b) review each rule on the poster board with the children, (c) set up rules poster board next to teacher or in a position clearly visible to all children, (d) use dry-erase board to deliver hatch marks for rule violations, (e) announce which teams won at the end of circle, and (f) deliver rewards to winning teams. For RFC+NCR sessions, the six components included: (a) present rules poster board to children, (b) review each rule on the poster board with the children, (c) set up rules poster board next to teacher or in a position clearly visible to all children, (d) use dry-erase board to deliver hatch marks for rule violations, (e) does not announce which teams won at the end of circle, and (f) deliver rewards to all teams at end of circle regardless of behavior during circle. For RVF+NCR sessions, the six components included: (a) present rules poster board to children, (b) review each rule on the poster board with the children, (c) set up rules poster board next to teacher or in a position clearly visible to all children, (d) deliver vocal feedback for rule violations (does not use dry-erase board to deliver hatch marks for rule violations), (e) does not announce which teams won at the end of circle, and (f) deliver rewards to all teams at end of circle regardless of behavior during circle. Each component was scored as correctly implemented, incorrectly implemented, correctly omitted, or incorrectly omitted. Treatment integrity was collected across 20% of sessions and was 100% across all treatment phases.

Results

Figure 1 depicts the class data with and without the data for Connor. The top panel depicts the class data with Connor's data included through session 87. The bottom panel depicts data for the class with all of Connor's data removed. After session 87 the top and bottom panel depict the same data. Across both graphs, disruptive behavior occurred at moderate to high levels during baseline. During the first implementation of GBG, when Connor's data are included, we observed an initial increase that is followed by a decrease to low levels after the first four sessions. When Connor's data are removed, we observed an immediate decrease in the level of disruptive behavior with the first implementation of the GBG. When we moved to RFC+NCR, we observed low levels of disruptive behavior for 15 sessions across both graphs; however, we observed slightly more variability when Connor's data are included. When we reversed to baseline, we observed an immediate increase in the level of disruptive behavior; however, when Connor's data are excluded the rate is not quite as high as the original baseline phase. We were then able to replicate treatment effects with the GBG. In the top panel, we observed more variability until Connor was removed from the study after session 87. In the bottom panel, we observed low and stable levels for the class when Connor's data are removed. When we attempted to replicate the RFC+NCR phase, we continued to observe low and stable levels of disruptive behavior for 12 sessions. When we faded to RVF+NCR, we continued to observe low levels of disruptive behavior for 16 sessions. We conducted a brief reversal to baseline and then replicated treatment effects with RVF+NCR for another 14 sessions.

Figure 2 depicts the data for the Blue Team with and without Connor's data. The top panel depicts the Blue Team's data with Connor's data included through session 87. The bottom panel depicts data for the Blue Team with all of Connor's data removed. After session 87 the top and bottom panel depict the same data. The open data points in the GBG phases depict sessions in which the Blue Team won the game (i.e., they earned the reward). With Connor's data included, the Blue Team demonstrated moderate to high levels of disruptive behavior; however, without Connor's data, the Blue Team demonstrated lower levels of disruptive behavior in the initial baseline. In the initial GBG Phase, when we include Connor's data, we observed an initial increase during the first three sessions followed by a decrease to low levels. In the bottom panel, when we exclude Connor's data, we observed an immediate decrease in both level and overall variability when the GBG is implemented. Although, the Blue Team experienced winning the GBG for the majority of sessions, they also experienced losing the GBG in some sessions. When we moved to RFC+NCR, we observed lower levels of disruptive behavior in the top panel; however, disruptive behavior is more variable and often slightly above the 80% reduction criterion. When we look at the Blue Team's data without Connor's data, levels of disruptive behavior are low and stable. After we replicate baseline, we attempted to replicate the effects of GBG again. In the top panel, we observed an immediate decrease in the overall level of disruptive behavior; however, the rate of disruptive behavior was often above the 80% reduction criterion and the Blue Team was frequently losing access to rewards. Thus, we implemented BST with Connor in an attempt to ensure he was attending to the rules and could tact appropriate behavior and rule-violation behavior. After five sessions, this was discontinued, and he was moved to his own team. We continued to collect data on his behavior; however, the data were not included in the class or Blue Team data. Connor continued to experience the same contingencies present during the session as the other children; however, teachers were instructed to implement additional individualized procedures. Specifically, they implemented time-out for major disruptive behavior (e.g., aggression and excessive IVB). Furthermore, during GBG sessions, if Connor did not exceed his own individualized winning criterion, he received immediate access to a higher magnitude reinforcer and individual teacher attention. During NCR sessions, if Connor was not engaging in aggression at the end of the session, he was provided with the same reward just described. Connor had a bout of teacher-directed aggression at the end of session 120. At the end of that session, we delayed access to the reward in order to decrease the potential for adventitious reinforcement of teacher-directed aggression. After we moved Connor to his own team, we observed low levels of disruptive behavior for the Blue Team for the remainder of the GBG phase. When we replicated RFC+NCR we continued to observe low and stable levels. Next, we faded to RVF+NCR and continued to observe low and stable levels of disruptive behavior. After a brief replication of baseline, we were able to replicate the effects of RVF+NCR.

Figure 3 depicts data for the Red and Green Teams. Both teams displayed moderate to high levels of disruptive behavior in the initial baseline and an immediate decrease to low, stable levels when we implemented the GBG. Both teams regularly won the GBG (denoted by open data points); however, both teams also experienced losing. When we implemented RFC+NCR, disruptive behavior continued to occur at low and stable levels. When we reversed to baseline, we only observed a slight increase for the Red Team. For the Green Team, we observed higher levels of disruptive behavior when we reversed to baseline; however, they do not return to the levels observed during the first baseline. We were able to replicate treatment effects with both the Red and Green Teams when we replicated the GBG and RFC+NCR phases. When we faded to RVF+NCR, we continued to observe low and stable levels of disruptive behavior for both teams. After a brief reversal to baseline, we were able to replicate effects of RVF+NCR.

Figure 4 depicts the mean rate of disruptive behavior for the class, teams, and individuals for each condition. At the class and team level, there is a large decrease in the mean level of disruptive behavior in all the intervention phases as compared to baseline. Although the mean rate for the GBG sessions is not at or below the 80% reduction criterion for the class and Blue Team, the individual data suggest this is the result of a single individual on the Blue Team. The highest individual data point for the Blue Team in baseline, GBG, and RFC+NCR depict the mean frequency for Connor across these phases; his data were not included in any of the RVF+NCR sessions for the Blue Team, so he does not have an individual data point. Although, his individual data were highly variable on a session-by-session basis, this analysis suggests the interventions were effective at decreasing the mean rate of disruptive behavior. The mean rate for the Red and Green Team was below the 80% reduction criterion for GBG sessions. Although a single individual on the Red Team also had a higher mean rate of disruptive behavior as compared to the other individuals on the team, the GBG was effective at reducing that individual's disruptive behavior to the 80% reduction criterion. With the exception of the Red Team, the mean rate for the class and teams was at or below the 80% reduction criterion for the RFC+NCR and RVF+NCR sessions. The individual data suggest that it was a single individual who engaged in elevated levels of disruptive behavior during some of these sessions.

Figure 5 depicts the side-effects data for the class. The top panel depicts the session-by-session data for positive and negative peer interactions that occurred during the pre- and post-sessions. The bottom panel depicts the mean frequency of these interactions across the various phases. In this graph, the first bar for each phase denotes the mean frequency of positive interactions; the second bar for each phase denotes the mean frequency of negative interactions. These bars are further delineated by whether those interactions occurred during pre- or post-sessions. Both graphs show that most interactions occurred during the post-session periods across all phases. Prior to exposure to the GBG, very few peer interactions occurred during the pre- and post- sessions. After exposure to the GBG, we observed an increase in both positive and negative peer interactions. Both the sessions-by-session data and the mean analysis show that across most phases positive interactions occurred at the highest frequency. However, in the second RFC+NCR phase, negative interactions demonstrated a slightly higher mean frequency as compared to positive interactions. The highest levels of positive interactions occurred in the RVF+NCR phases.

Figure 6 depicts the mean frequency of positive and negative peer interactions for all three teams. Similar to the classwide data, we observed few interactions prior to exposure to the GBG, and the majority of interactions occurred during the post-session periods across teams. Furthermore, during most phases, both positive and negative peer interactions occurred across teams. During the first GBG phase, we observed a small increase in positive interactions and low levels of negative interactions for the Blue Team. The Red Team displayed a large increase in positive interactions and a slight increase in negative interactions. The Green Team displayed a large increase in both positive and negative interactions. During the second implementation of the GBG, we observed both positive and negative interactions with all teams. Slightly higher levels of positive interactions as compared to negative interactions occurred for the Red and Green Teams, whereas higher levels of negative interactions as compared to positive interactions occurred for the Blue Team. In the first implementation of RFC+NCR, we observed both types of interactions. Slightly higher levels of positive interactions as compared to negative interactions occurred for the Red and Green Teams, whereas substantially higher levels of positive interactions occurred for the Blue Team. In the second RFC+NCR phase, similar levels of positive and negative interactions occurred for all three teams. For the first implementation of RVF+NCR, we saw similar levels of positive and negative interactions for the Blue Team. Additionally, this is the only phase where more interactions occurred during the pre-session period. For the Red and Green Teams, we observed more positive interactions as compared to negative interactions during this phase. In the second implementation of this phase, we saw the highest levels of positive interactions for all three teams. Additionally, we saw higher levels of positive interactions as compared to negative interactions for all three teams.

Figure 7 depicts the mean frequency of positive and negative interactions that occurred during the post-sessions periods associated with sessions in which rewards (a) were delivered during response-independent sessions (NCR sessions), (b) were delivered during response-contingent sessions (GBG), and (c) were not delivered (but were available for delivery) during GBG sessions for the three teams. Across all three teams, more positive interactions occurred as compared to negative interactions when a reward was delivered contingent on behavior and noncontingently. Across all three teams, more negative interactions occurred as compared to positive interactions when the teams lost the GBG and did not earn rewards.

Figure 8 depicts the mean rate of positive and negative interactions that occurred in the post-session periods of intervention sessions associated with high rates of disruptive behavior (i.e., less than an 80% reduction from baseline levels) and with low rates of disruptive behavior (i.e., an 80% or more reduction from baseline levels) for the class and all three teams. Results varied across the class and three teams for sessions with high levels of disruptive behavior. For the class, we observed similar levels of positive and negative interactions when high levels of problem behavior occurred, whereas we observed higher levels of positive interactions as compared to negative interactions when low levels of problem behavior occurred. Similar to the class data, for the Blue Team, we observed similar levels of positive and negative interactions when high levels of problem behavior occurred but higher levels of positive interactions as compared to negative interactions when lower levels of problem behavior occurred. For the Red Team, we observed more positive interactions as compared to negative interactions both when high and low levels of problem behavior occurred. For the Green Team, we observed slightly more negative interactions as compared to positive interactions when high levels of problem behavior occurred and higher levels of positive interactions as compared to negative interactions when low levels of problem behavior occurred.

Discussion

Results of the current study replicate previous research (Foley et al., 2019; Ling & Barnett, 2013; Wiskow et al., 2019) by showing that the GBG effectively decreased the rate of the disruptive behavior of preschoolers during structured group instruction. Additionally, we were able to fade the interdependent group contingency, visual feedback, and winning criterion while maintaining treatment effects. These results are important for several reasons. First, it strengthens the support for use of the GBG in preschool classrooms by adding to the literature demonstrating the efficacy this intervention with this population. Second, it extends the literature on maintenance of treatment effects observed with the GBG (Dadakhodjaeva et al., 2019; Lynch & Keenan, 2018; Ruiz-Olivares et al., 2010) by offering another method for maintaining treatment effects in the absence of the intervention. Third, results of the current study suggest that it is possible to fade the effortful components while maintaining treatment effects.

Results of the present study also suggest the GBG may be associated with an increase in both positive and negative interactions. However, the occurrence of these interactions appears to be sensitive to specific variables associated with implementation of the GBG. Specifically, the level of problem behavior and access to reward delivery. These results extend the literature in several ways. First, they replicate and extend the literature on side-effects associated with group contingencies by assessing the occurrence of side-effects with preschoolers in an applied setting. Second, they support the assumption that group contingencies are associated with an increase in peer interactions. However, variables associated with the group contingency (e.g., earning a reward and the occurrence of target behavior) influence which type of peer interaction may occur. For example, group success (i.e., earning the reward) was associated with positive interactions, whereas group failure (i.e., contingent loss of a reward) was associated with negative interactions.

Efficacy of the GBG with preschoolers may be particularly important for several reasons. First, disruptive behavior during structured learning times is associated with teacher burnout (Hastings & Bham, 2003), student expulsion (Gilliam & Golan, 2009; Tangora, 2015), and lower academic gains for students (Bulotsky-Shearer & Fantuzzo, 2011). Thus, the GBG provides preschool teachers with an effective intervention for addressing the disruptive behavior associated with these negative outcomes. Second, the GBG provides teachers with an effective class-level intervention. As more preschools adopt a response-to-intervention (RTI) model in order to make data-based decisions for their students, the GBG offers a Tier 1 approach that teachers can implement with the class or a large group (Ball & Trammell, 2009; Bayat, Mindes, & Covitt, 2010). However, results of the current study suggest the GBG was not effective for all students. Thus, it appears there may be some preschool students for whom the GBG is not effective. Modifications to the GBG or individualized interventions may be necessary for these students. The researchers made several modifications to increase the efficacy of the intervention for Connor. Individual presentation of the rules, rehearsal, feedback, an individual contingency, and manipulating dimensions of reinforcement (i.e., immediacy, quantity, and quality) were all used to increase the efficacy of the GBG. Connor's data were highly variable throughout the study which suggests that the GBG, and the other interventions used during the study, did not influence his behavior. It may have been necessary to conduct a functional analysis (Greer et al., 2013; Iwata, Dorsey, Slifer, Bauman, & Richman, 1982/1994) in order to develop a function-based intervention (Carr & Durand, 1985; Hanley, Piazza, Fisher, Contrucci, & Maglieri, 1997) for Connor. For example, escape from circle may have been the function of Connor's disruptive behavior. A function-based treatment may have involved teaching Connor to ask for a break from circle. Alternatively, we may not have observed a treatment effect with Connor due to a skill deficit. For example, a student may need to have acquired rule-governed behavior (Skinner, 1969), the ability to self-generate rules (Cerutti, 1989), or at minimum the ability to follow if-then statements (e.g., if you break a circle-time rule, then you get a point for your team; Tarbox, Zuckerman, Bishop, Olive, & O'Hora, 2011). Future research should conduct skills assessments prior to implementation of the GBG in order to identify skills associated with treatment effects and deficits associated with a lack of treatment effect.

The precise mechanisms responsible for behavior change during the implementation of the GBG are unclear. The GBG may function as a differential reinforcement of low rates (DRL) procedure in which reinforcement is delivered if the number of responses during an interval or full session is less than or equal to a predetermined limit (Dietz & Repp, 1973). A DRL can be implemented with a single individual (Dietz & Repp, 1973) or with a group (Austin & Bevan, 2013) and may be used with or without stimuli that signal the availability of reinforcement (Becraft, Borrero, Davis, Mendres-Smith, & Castillo, 2018). Thus, the GBG may function as a full-session DRL in which the winning criterion reflects the predetermined limit and the visual feedback functions as signals for reinforcement. Alternatively, the visual and vocal feedback may serve as conditioned punishers (Hake & Azrin, 1965) due to their repeated pairing with the loss of access to rewards. Finally, results of the side-effects analysis from the present study and previous research on side-effects associated with the GBG (Groves & Austin, 2019) suggests that peer-interactions increase after exposure to the GBG. Thus, these side-effects may enhance the efficacy of the GBG. That is, the peers may provide unprogrammed prompts or consequences before, during, or after a GBG session. These interactions may serve as prompts (e.g., rule reminders prior to session), reinforcement (e.g., praise after a session), or punishment (e.g., scowls or negative gestures during session) that contribute to the efficacy of the GBG. It is also likely that several mechanisms are acting on behavior given the robust findings associated with the GBG.

Another important finding of the present study was the demonstration of continued treatment effects under RFC+NCR and RVF+NCR. These results replicate Foley et al. (2019) by replicating the RFC+NCR phase. Additionally, these results replicate and extend the literature demonstrating the importance of the feedback component of the GBG (e.g., Foley et al., 2019; Medland & Stachnik, 1973; Wiskow et al., 2019). That is, we did not observe an interruption of treatment effects under NCR with vocal + visual feedback (RFC condition) and vocal feedback (RVF condition). However, we did not evaluate rules + NCR (R+NCR), so we do not know if we could have faded the feedback component from the intervention. Anecdotally, the students generally engaged in a few instances of disruptive behavior at the beginning of RFC+NCR and RVF+NCR sessions which stopped after a team received feedback. Thus, it may require additional manipulations to fade the feedback without losing treatment effects under R+NCR. For example, it may require fading the schedule of implementation days similar to Dadakhodjaeva et al. (2019) by interspersing R+NCR sessions and RVF+NCR sessions. Additionally, it may be important to remove the signal cueing the beginning of a session such that the contingencies are indiscriminable (Stokes & Baer, 1977) in order to maintain treatment effects.

The results demonstrated under the RVF+NCR condition are particularly important because they suggest that it is possible to fade some of the more effortful components of the GBG. That is, we were able to fade the interdependent group contingency, visual feedback, and criterion while maintaining treatment effects. The remaining components included teams, rules, vocal feedback, and noncontingent rewards. The remaining components are the least effortful components and are often part of typical classroom practices. Affiliation with a team and that team's performance no longer contributed to reward delivery. Thus, the team mats and assigned seats served as a seating chart which allowed children to come to circle and find their seat without the assistance of a teacher. Reviewing rules for behavioral expectations requires minimal effort and are common in classroom environments (Gable, Hester, Rocks, & Hughes, 2009). Vocal feedback does not require any additional materials (e.g., dry-erase board or chalk board) and provides an alternative to the delivery of reprimands (Elswick & Casey, 2011; Lannie & McCurdy, 2007; Ling & Barnett, 2013; Rubow et al., 2018). Additionally, the vocal feedback in the current study was concise and did not draw attention to the problem behavior, which follows guidelines for the effective use of reprimands (Gable et al., 2009). The delivery of tangible rewards can be controversial in school settings; however, results of an evaluation of teachers use and perception of rewards in the classroom showed that all teachers surveyed in the study used rewards in their classroom (Hoffmann, Huff, Patterson, & Nietfeld, 2009). Additionally, teachers felt that rewards were more appropriate for addressing problem behavior as compared to academic achievement (Hoffmann et al., 2009). Thus, RVF+NCR provides a less effortful intervention for teachers and uses components common to the classroom setting.

Literature on maintenance under NCR schedules suggest maintenance is likely due to adventitious reinforcement (Skinner, 1948) or stimulus control (Uhl & Garcia, 1969).

Maintenance may be due to adventitious reinforcement if delivery of rewards is in close temporal proximity to appropriate behavior. Alternatively, the stimuli associated with the GBG (e.g., rules poster board, experimenters, feedback, and treasure box) may have acquired discriminative properties after repeated pairing with the contingent delivery of rewards. Additionally, both mechanisms may influence maintained levels of responding. Regardless of the mechanisms, the efficacy of RVF+NCR may increase the acceptability and implementation of the GBG. That is, Dadakhodjaeve et al. (2019) reported that teachers liked the GBG but would prefer to fade components instead of thinning the schedule of implementation. The RVF+NCR intervention may offer teachers the ability to implement the GBG then fade to a more manageable intervention. However, we did not evaluate teacher implementation of any of the interventions evaluated in the present study; thus, future research should evaluate if similar results would occur under teacher implementation.

Another important area addressed by the present study was the analysis of peer interactions that occurred immediately before and immediately after the GBG. To date, only Groves and Austin (2019) have evaluated peer-interactions in the context of the GBG in an applied setting. Similar to Groves and Austin, we found an association between the GBG and positive peer interactions; however, we also observed negative peer interactions. Further analysis showed negative peer interactions were primarily associated with the contingent loss of rewards. Positive interactions were associated with both contingent and noncontingent reward delivery, and with sessions with low rates of disruptive behavior. However, it is important to note that our analysis of the association between disruptive behavior and peer interactions may have been skewed by our criterion for sessions with high and low rates of disruptive behavior. That is, the winning criterion was set at a frequency of 11 occurrences of disruptive behavior per session; however, session durations differed in length which would allow a team to win even when their rate was not below the 80% reduction criterion. Thus, if a team received 11 hatch marks during a circle that lasted 6 min, as compared to a 19 min circle; the differences in rate would not affect reward delivery but the 6 min session would be considered a session with a high rate of PB and the 19 min session would not.

The most common positive interactions included congratulating each other ("e.g., Yay! Everybody wins today!") and commenting on reward delivery (e.g., "Let's all get yellow slime today!"). Anecdotal notes suggest that the most common types of negative interactions included peer delivered reprimands (e.g., "We lost because you kept making noises!") delivered in an inappropriate tone, complaining about the loss of reward (e.g., "We don't get a treat now because you kept shouting out the answers."), and statements of disapproval (e.g., "We're not playing with you anymore."). We also observed harassing members of the losing team(s) during post-sessions. However, researchers primarily observed only two individuals harass members of losing team(s). These individuals were Connor on the Blue Team and Ace on the Red Team. Additionally, they usually directed these comments to each other. Our anecdotal notes suggest that the types of negative peer interactions may serve as additional punishers for disruptive behavior. For example, after the Green Team lost for the first time during the first implementation of the GBG (session 17), two individuals on the team provided several reprimands to the team member who accrued most of the hatch marks for the team during that session. These reprimands included specific examples of the behaviors that resulted in the team getting hatch marks. For example, one of the comments was "You kept getting us points because you kept making fart noises!" However, after this session, the individual stopped making sounds during circle time and the Green Team won during all the subsequent sessions in that phase.

Thus, the negative interactions may be important for behavior change during initial GBG sessions.

Although the present study demonstrated some interesting and important results, there are several limitations that must be considered. First, we were unable to demonstrate a treatment effect with Connor. Additionally, it is unclear why we were not able to find an effective manipulation. Previous evaluations of the GBG have included children for whom the intervention was not effective; however, simple manipulations (e.g., moving the student to their own team) have generally improved these student's behavior (e.g., Harris & Sherman, 1973; Rubow, Vollmer, & Joslyn, 2018). As previously discussed, a function-based approach may be necessary for an individual like Connor. A systematic preference assessment (DeLeon & Iwata, 1996; Fisher et al., 1992) may have also allowed us to identify effective reinforcers. Second, researchers conducted all sessions, so we were unable to assess the acceptability of the NCR conditions. A primary purpose of the present study was to evaluate a method for fading effortful components to increase adoption of the GBG in classroom settings. Therefore, future research should replicate the current study with teacher implementation such that social validity (Wolf, 1978) of the interventions can be assessed. Third, the use of a frequency measure for all topographies of behavior may not have represented the most sensitive measurement of all the behaviors. For example, a duration measure may have provided a more sensitive measure for out of seat because the duration of out of seat can differ across occurrences. However, typical classroom procedures included guided compliance (e.g., Cote, Thompson, & McKerchar, 2005) for sitting. Thus, most instances of out of seat were brief due to typical behavior management strategies used in the classroom. Fourth, we did not control for previous exposure to the GBG when we created the teams. The Red and Green Teams each had three individuals who had previous exposure to the GBG; whereas, the Blue Team only had one individual with previous exposure. Peers with previous exposure may have contributed to the success of the Red and Green Teams because the stimuli associated with the GBG (e.g., the name of the intervention, instructions, color-correlated teams mats, and the treasure box) may have evoked appropriate behavior. This immediate effect for a large portion of a team may have allowed these teams to win more frequently, allowing the other team member to contact reinforcement. Additionally, the children with previous exposure may have engaged in more peer interactions that may have enhanced the efficacy of the GBG.

Additionally, there were several limitations in the analysis of unprogrammed peer interactions. First, the procedures in the current study did not capture all the relevant peer interactions. That is, researchers only collected data during the 3 min pre- and post-session periods, but peer interactions occurred throughout the day. For example, the classroom teachers reported that children would frequently discuss the GBG during mealtimes and often provided appropriate rule reminders during these times. In addition, the children would sometimes play the GBG during free choice play periods. The children often provided more descriptive feedback (e.g., “Ace you got a point for not sitting on your mat.”) during these play sessions. Future research on peer interactions associated with the GBG might involve conducting several observations across various contexts in order to better capture these behaviors. Additionally, we observed side effects during circle time. However, these generally included gestures because vocalizing in the absence of teacher permission would have resulted in point delivery. Thus, we decided not to collect data on side-effects during circle. Future research could evaluate the association between side-effects during circle and behavior during circle. For example, a contiguity analysis (Dozier et al., 2001) analysis may provide insight on the association between peer-interactions during GBG and behavior change within session.

Second, the present study involved collecting data on the quantity (i.e., overall frequency) of peer interactions; however, it may be more important to collect qualitative data on the types of interactions observed. Although we collected detailed anecdotal notes during sessions, we do not have data on the frequencies of the various types of interactions. Future research should evaluate the frequency of the various topographies of peer interaction (e.g., reprimands, rule reminders, harassment, praise, and gestures) and their association with the GBG. Third, additional replications of the side-effect analysis are needed before we can draw confident conclusion on their association with the GBG. Fourth, low IOA for the side-effect analysis represents a threat to the believability of the side effects data. That is, IOA summarizes the extent to which two observers agree on the occurrence and nonoccurrence of a target behavior (Poling, Methot, & Lesage, 1995). Low IOA for data collection may have been the result of low frequencies of the target behavior, poor training, or subjective and ambiguous definitions. However, data collectors frequently demonstrated high levels of IOA, and low levels of IOA were not associated with a single data collector. Future research should evaluate other methods for data collection on side-effects, particularly when observing multiple (i.e., 5-6) individuals as once.

References

Alijadeff-Abergel, E., Peterson, S. M., Wiskirchen, R. R., Hagen, K. K., & Cole, M. L. (2017). Evaluating the temporal location of feedback: Providing feedback following performance vs. prior to performance. *Journal of Organizational Behavior Management, 35*, 171-195. doi: 10.1080/01608061.2017.1309332

Austin, J. L., & Bevan, D. (2013). Using differential reinforcement of low rates to reduce children's request for teacher attention. *Journal of Applied Behavior Analysis, 44*, 451-461. doi: 10.1901/jaba.2011.44-451

Azrin, N. H. & Holz W. C. (1966). Punishment. In W. K. Honig (Ed.), *Operant Behavior: Areas of Research and Application* (380-447). New York: Appleton-Century-Crofts

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91-97. doi: 10.1901/jaba.1968.1-91

Ball, C.R. & Trammell (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 Behavior 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

Becraft, J. L., Borrero, J. C., Davis, B. J., Mendres-Smith, A. E., & Castillo, M. I. (2018). The role of signals in two variations of differential-reinforcement-of-low-rate procedures. *Journal of Applied Behavior Analysis, 51*, 3-24. doi: 10.1002/jaba.431

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

Bushell Jr., D., Wrobel, P. A., Michaelis, M. L. (1968). Applying "group" contingencies to the classroom study behavior of preschool children. *Journal of Applied Behavior Analysis*, 1, 55-61. doi:10.1901/jaba.1968.1-55

Burnett, P. C., & Mandel, V. (2010). Praise and feedback in the primary classroom: Teachers' and students' perspectives. *Australian Journal of Educational & Developmental Psychology,* 10, 145-154. https://files.eric.ed.gov/fulltext/EJ906941.pdf

Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis, 18*, 111–126. doi: 10.1901/jaba.1985.18-111

Centers for Disease Control and Prevention. (2017). *School-based violence prevention.* Retrieved from https://www.cdc.gov/policy/hst/hi5/violenceprevention/index.html

Cerutti, D. T. (1989). Discrimination theory of rule-governed behavior. *Journal of the Experimental Analysis of Behavior, 51*, 259-276. doi: 10.1901/jeab.1989.51-259

Cheatham, J. M., Ozga, J. E., St. Peter, C. C., Mesches, G. A., & Owsiany, J. M. (2017). Increasing class participation in college classrooms with the Good Behavior Game. *Journal of Behavioral Education, 26*, 277-292. doi: 10.1007/s10864-017-9266-7

Coalition for Psychology in Schools and Education. (2006). *Report on the teacher needs survey.* Washington, D.C.: American Psychological Association, Center for Psychology in Schools and Education.

Cote, C.A., Thompson, R.H., & McKerchar, P.M. (2005). The effects of antecedent interventions and extinction on toddlers' compliance during transitions. *Journal of Applied Behavior Analysis, 38*, 235-238. doi: 10.1901/jaba.2005.143-04

Dadakhodjaeva, K., Radley, K. C., Tingstrom, D. H., Dufrene, B. A., & Dart, E. H. (2019). Effects of daily and reduced frequency implementation of the Good Behavior Game in kindergarten classrooms. *Behavior Modification,* published online ahead of print. doi: 10.1177/0145445519826528

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29,* 519–533. doi: 10.1901/ jaba.1996.29-519

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

Dietz, S. M., & Repp, A. C. (1973). Decreasing misbehavior through the use of DRL schedules of reinforcement. *Journal of Applied Behavior Analysis, 6*, 457-463. doi: 10.1901/jaba.1973.6-457

Dolan, L. J., Kellam, S. G., Brown, C. H., Werthamer-Larsson, L., Rebok, G. W., Mayer, L., S., Laudolff, J., & Turkkan, J. S. (1993). The short-term impact of two classroom-based prevention interventions on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology, 14*, 317-345. doi: 10.1016/0193-3973(93)90013-L

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

Donaldson, J. M., Matter, A. L., & Wiskow, K. M. (2018). Feasibility of and teacher preference for student-led implementation of the Good Behavior Game in early elementary classrooms. *Journal of Applied Behavior Analysis, 51*, 118-129. doi: 10.1002/jaba.432

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

Elswick, S., & Casey, L. B. (2011). The Good Behavior Game is no longer just an effective intervention for students: An examination of the reciprocal effects on teacher behaviors. *Beyond Behavior, 21*, 36-46 http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=79627010&site=ehost-live

Embry, D. D. (2002). The good behavior game: A best practice candidate as a universal behavioral vaccine. Clinical Child and Family Psychology Review, 5, 273-297. doi:10.1023/A:1020977107086

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

Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying rein- forcers for persons with severe and profound disabil- ities. *Journal of Applied Behavior Analysis, 25,* 491–498. doi: 10.1901/jaba.1992.25-491

Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., & Vega, R. (2014). Effects of the Good Behavior Game on challenging behaviors in school settings. *Review of Educational Research, 84*, 546-571. doi: 10.3102/0034654314536781

Flower, A., McKenna, J., Muething, C. S., Bryant, D. P., & Bryant, B. R. (2014). Effects of the Good Behavior Game on class-wide off-task behavior in a high school basic algebra resource classroom. *Behavior Modification, 38*, 45-68. doi: 10.1177/0145445513507574

Foley, E. A., Dozier, C. L., & Lessor, A. L. (2019) Comparison of components of the Good Behavior Game in a preschool classroom. *Journal of Applied Behavior Analysis, 52*, 84-104. doi: 10.1002/jaba.506

Fullerton, E. K., Conroy, M. A., & Correa, V. I. (2009). Early childhood teachers' use of specific praise statements with young children at risk for behavioral disorders. *Behavioral Disorders, 34*, 118-135. https://search.proquest.com/docview/219676467?accountid=14556

Gable, R. A., Hester, P. H., Rock, M. L., & Hughes, K. G. (2009). Back to basics: Rules, praise, ignoring, and reprimands revisited. *Interventions in School and Clinic, 44*, 195-205. doi: 10.1177/1053451208328831

Galibraith, L. A., & Normand, M. P. (2017). Step it up! Using the Good Behavior Game to increase physical activity with elementary school students at recess. *Journal of Applied Behavior Analysis, 50*, 856-860. doi: 10.1002/jaba.402

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 doi: 10.1097/00001163-200607000-00007

Greer, B. D., Neidert, P. L., Dozier, C. L., Payne, S. W., Zonneveld, K. L., & Harper, A. M. (2013). Functional analysis and treatment of problem behavior in early education classrooms. *Journal of Applied Behavior Analysis, 46*, 289-295. doi: 10.1002/jaba.10

Gresham, F. M., & Gresham, G. N. (1982). Interdependent, dependent, and independent group contingencies for controlling disruptive behavior. *Journal of Special Education, 16*, 101-110. doi: 10.1177/002246698201600110

Groves, E. A., & Austin, J. L., (2019). Does the Good Behavior game evoke negative peer pressure? Analysis in primary and secondary classrooms. *Journal of Applied behavior Analysis, 52*, 3-16. doi: 10.1002/jaba.513

Haas, J. R., & Hayes, S. C. (2006). When knowing you are doing well hin- ders performance: Exploring the interaction between rules and feedback. *Journal of Organizational Behavior Management, 26*, 91–111. doi:10.1300/ J075v26n01_04

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

Hanley, G. P., Piazza, C. C., Fisher, W.W., Contrucci, S. A., & Maglieri, K. A. (1997). Evaluation for function-based treatment packages. *Journal of Applied Behavior Analysis, 30*, 459-473. doi: 10.1901/jaba.19997.30-459

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. 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

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

Hoffmann, K. F., Huff, J. D., Patterson, A. S., & Nietfeld, J. L. (2009). Elementary teachers’ use and perception of rewards in the clasroom. *Teaching and Teacher Education, 25*, 843-849. doi:10.1016/j.tate.2008.12.004

Honig, W. K. (1966). The role of discrimination training in the generalization of punishment. *Journal of the Experimental Analysis of Behavior, 9*, 377-384. doi: 10.1901/jeab.1966.9-377

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis*, *27*, 197-209. doi: 10.1901/jaba.1994.27-197

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

Johnson, R. A., Houmanfar, R., & Smith, G. S. (2010). The effect of implicit and explicit rules on customer greeting and productivity in a retail organization. *Journal of Organizational Behavior Managementi, 30*, 38–48. doi:10.1080/ 01608060903529731

Joslyn, P. R., Vollmer, T. R., & Hernández, V. (2014). Implementation of the Good Behavior Game in a classroom for children with delinquent behaviors. *Acta de Investigación Psicológica, 4*, 1673-1681. doi:10.1016/S2007-4719(14)70973-1

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H. & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology, 10*, 165-185. http://dx.doi.org/10.1017/S0954579498001564

Kellam, S. G., Mackenzie, A. C., Brown, C. H., Poduska, J. M., Wang, W., Petras, H., & Wilcox, H. C. (2011). The Good Behavior Game and the future of prevention and treatment. *Addiction Science & Clinical Practice, 6*, 73-84. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3188824/

Kellam, S. G., Wang, W., Mackenzie, A. M., Brown, C. H., Ompad, D. C., Or, F., …, Windam, A. (2014). The impact of the Good Behavior game, a universal classroom-based preventive intervention in first and second grades, on high-risk sexual behaviors and drug abuse and dependence into young adulthood. *Prevention Science, 15*, 6-18. doi:10.1007/s11121-012-0296-z.

Kelshaw-Levering, K., Sterling-Turner, H. E., Henry, J. R., & Skinner, C. H. (2000). Randomized interdependent group contingencies: Group reinforcement with a twist. *Psychology in the Schools*, 37, 523-533. doi:10.1002/1520-6807(200011)37:6<523::AID-PITS5>3.0.CO;2-W

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

Lannie, A. L., & McCurdy, B. L. (2007). Preventing disruptive behavior in the urban classroom: Effects of the Good Behavior Game on student and teacher behavior. *Education and Treatment of Children, 30*, 85-98. doi: 10.1353/etc.2007.0002

Kosiec, L. E., Czernicki, M. R., & McLaughlin, T. F. (1986). The Good Behavior Game: A replication with consumer satisfaction in two regular elementary school classrooms. *Techniques, 2*, 15–23.

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., 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 elementart classrooms. *Journal of School Psychology 51*, 187-199. doi: 10.1016/j.jsp.2012.12.006

Ling, S. M., & Barnett, D. W. (2013). Increasing preschool student engagement during group learning activities using a group contingency. *Topics in Early Childhood Special Education, 33*, 186-196. doi: 10.1177/0271121413484595

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

Little, S.G., Akin-Little, A., & O'Neill, K. (2015). Group contingency interventions with children-1980-2010: A meta-analysis. *Behavior Modification,* 39, 322-341. doi: 10.1 177/0145445514554393

Lutzker, J. R.,&White-Blackburn, G. (1979). The good productivity game: Increasingwork performance in a rehabilitation setting. *Journal of Applied Behavior Analysis, 12*, 488. doi: 10.1901/jaba.1979.12-488

Lynch, D., & Keenan, M. (2018). The good behavior game: Maintenance effects. *International Journal of Education, 87*, 91-99. http://dx.doi.org/10.1016/j.ijer.2016.05.005

Maggin, D. M., Johnson, A. H., Chafouleas, S. M., Ruberto, L. M., & Berggren, M. (2012). A systematic evidence review of school-based group contingency interventions for students with challenging behaviors. *Journal of School Psychology,* 50, 625-654. doi: 10.1016/j.jsp.2012.06.001

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. doi: 10.1901/jaba.1973.6-425

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

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

Mitchell, R. R., Tingstrom, D. H., Dufrene, B. A., Ford, W. B. & Sterling, H. E. (2015) The effects of the Good Behavior Game with general-education high school students. *School Psychology Review, 44*, 191-207. doi: 10.17105/spr-14-0063.1

National Center for Education Statistics. (2018). *Preschool and kindergarten enrollment.* Retreived from https://nces.ed.gov/programs/coe/indicator_cfa.asp

Packard, R. G. (1970). The control of "classroom attention": A group contingency for complex behavior. *Journal of Applied Behavior Analysis, 3*, 13-28. doi: 10.1910/jaba.1970.3-13

Parsons, M.B., Rollysons, J.H., & Reid, D.H. (2012). Evidenced-based staff training: A guide for practitioners. *Behavior Analysis in Practice, 5,* 2-11. doi: 10.1007/BF03391819

Payne, S. W., Dozier, C. L., Briggs, A. M., & Newquist, M. H. (2017). An Analysis of group-oriented contingencies and associated side-effects in preschool children. *Journal of Behavioral Education, 26*, 27-52. doi: 10.1007/s10864-016-9255-2

Polling, A., Methot, L. L., & LeSage, M. G. (1995). *Fundamental of Behavior Analytic Research*. New York: Plenum Press. http://dx.doi.org/10.1007/978-1-4899-1436-1

Radley, K. C., Dart, E. H., & O'Handley, R. D. (2016). The quiet classroom game: a class-wide intervention to increase academic engagement and reduce disruptive behavior. *School Psychology Review,* 45, 93-108. doi: 10.17102/SPR45-1.93-108

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

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. https://www-jstor-org.www2.lib.ku.edu/stable/44738315

Rubow, C.C., Vollmer, T.R., & Joslyn, P.R. (2018). Effects of the Good Behavior Game on student and teacher behavior in an alternative school. *Journal of Applied Behavior Analysis, 51*, 382-392. doi: 10.1002/jaba.455

Ruiz-Olivares, R., Pino, M. J., & Herruzo, J. (2010). Reduction of disruptive behaviors using an intervention based on the Good Behavior Game and the say-do report correspondence. *Psychology in the Schools, 47*, 1046-1058. doi: 10.1002/pits.20523

Saigh, P. A., & Umar, 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

Scott, K. C., Skinner, C. H., Moore, T. C., McCurdy, M., Ciancio, D., & Cihak, D. F. (2017). Evaluating and comparing the effects of group contingencies on mathematics accuracy in a first-grade classroom: Class average criteria versus unknown small-group average criteria. *School Psychology Review, 46*, 262-271. doi: 10.17105/SPR-2017-0037.V46-3

Skinner, B. F. (1948). Superstition in the pigeon. The Journal of Experimental Psychology, 38, 168-172. doi: 10.1037/h0055873

Skinner B.F. (1969) Contingencies of reinforcement: A theoretical analysis. New York: Appleton-Century-Crofts.

Slavin, R. E. (1977). Classroom reward structure: An analytic and practical review. *Review of Educational Research, 47*, 633-650. https://www.jstor.org/stable/1170003 accessed: 23-04-2019. doi: 10.3102/00346543047004633

Speltz, M. L., Shimamura, J. W., & McReynolds, W. T. (1982). Procedural variations in group contingencies: Effects on children's academic and social behaviors, *Journal of Applied Behavior Analysis, 15*, 533-544. doi: 10.1901/jaba.1982.15-533

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

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*, 349-367. doi: 10.1901/jaba.1977.10-349

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

Tangora, 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

Tarbox, J., Zuckermann, C. K., Bishop, M. R., Olive, M. L., & O'Hora, D. P. (2011). Rule-governed behavior: Teaching a preliminary repertoire of rule-following to children with autism. *The Analysis of Verbal Behavior, 27*, 125-139. doi: 10.1007/bf03393096

Terrace, H. S. (1966) Stimulus Control. In W. K. Honig (Ed.), *Operant Behavior: Areas of Research and Application* (271-344). New York: Appleton-Century-Crofts

Thompson, R. H., Iwata, B. A., Hanley, G. P., Dozier, C. L., & Samaha, A. L. (2003). The effects of extinction, noncontingent reinforcement, and differential reinforcement of other behavior as control procedures. *Journal of Applied Behavior Analysis, 36*, 221-238. doi: 10.1901/jaba.2003.36-221

Tingstrom, D. H., Sterling-Turner, H. E., & Wilczynski, S.M. (2006). The good behavior game: 1969-2002. *Behavior Modification, 30,* 225.253. doi: 10.1177/0145445503261165

Uhl, C. N., & Garcia, E. E. (1969). Comparison of omission with extinction in response elimination in rats. *Journal of Comparative and Physiological Psychology, 3*, 554-562. doi: 10.1037/h0028243

Ward-Horner, J. & Sturmey, P. (2010). Component analyses using single subject design: A review. *Journal of Applied Behavior Analysis, 43*, 685-704. doi:10.1901/jaba.43-685

Wilder, D. A., & Atwell, J. (2006). Evaluation of a guided compliance procedure to reduce noncompliance among preschool children. *Behavioral Interventions, 21*, 265– 272. https://doi-org.www2.lib.ku.edu/10.1002/bin.222

Wiskow, K. M., Matter, A. L., Donaldson, J. M. (2019). The Good Behavior Game in preschool classrooms: An evaluation of feedback. *Journal of Applied Behavior Analysis, 52*, 105-115. doi: 10.1002/jaba.500

Wolf, M. M. (1978). Social validity: The case for subjective measurement *or* how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 1978*, 203-214. doi:10.1901/jaba.1978.11-203

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

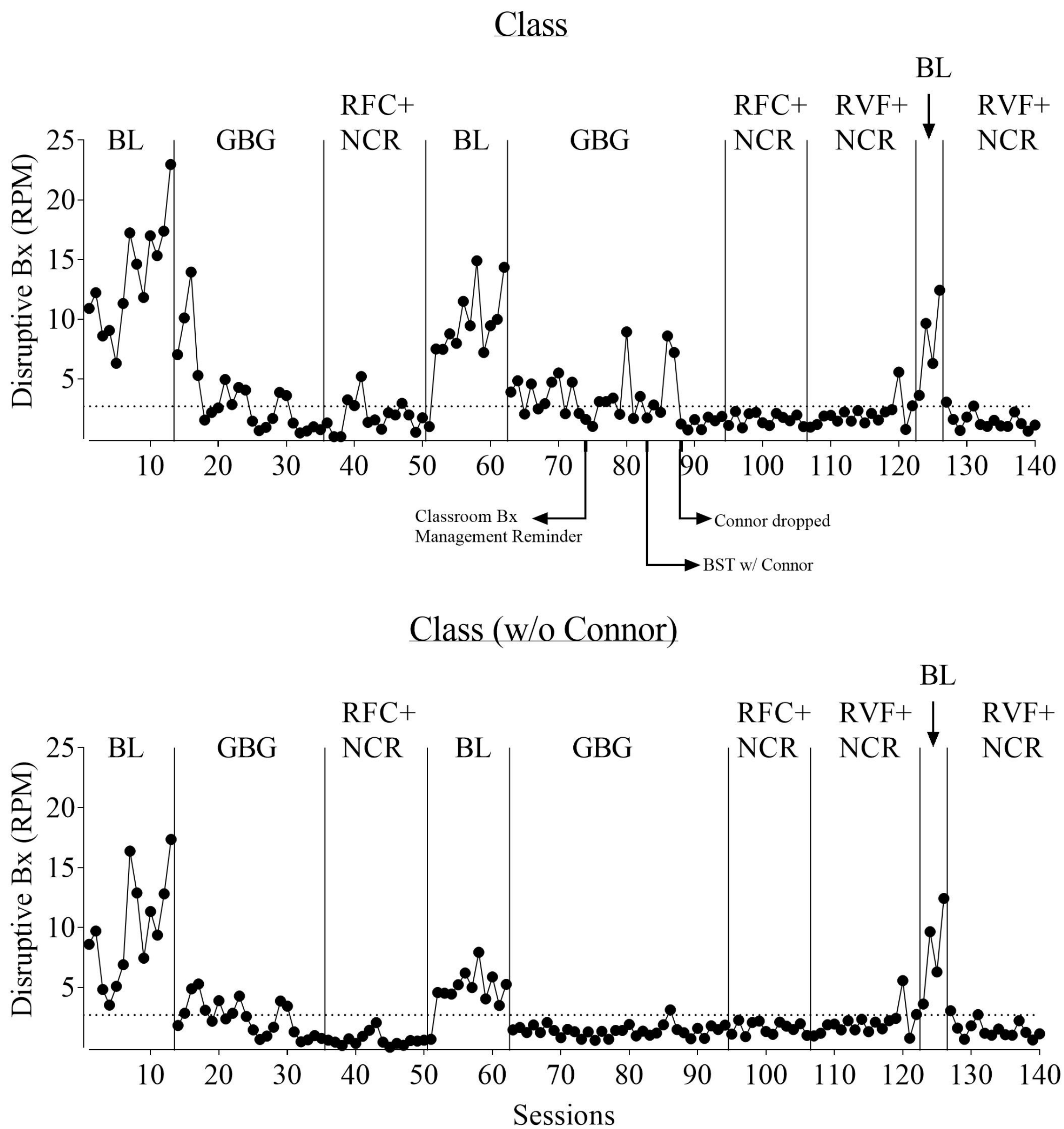


Figure 1. The top panel depicts the class-wide rate of disruptive behavior with Connor's data included through session 87. The bottom panel depicts the class-wide rate of disruptive behavior with Connor's data removed. The dotted line represents an 80% reduction in disruptive behavior from baseline levels. BL=Baseline, GBG=Good Behavior Game, RFC+NCR= Rules + Vocal & Visual Feedback + Criterion + Noncontingent Reinforcement, RVF+NCR= Rules + Vocal Feedback + Noncontingent Reinforcement

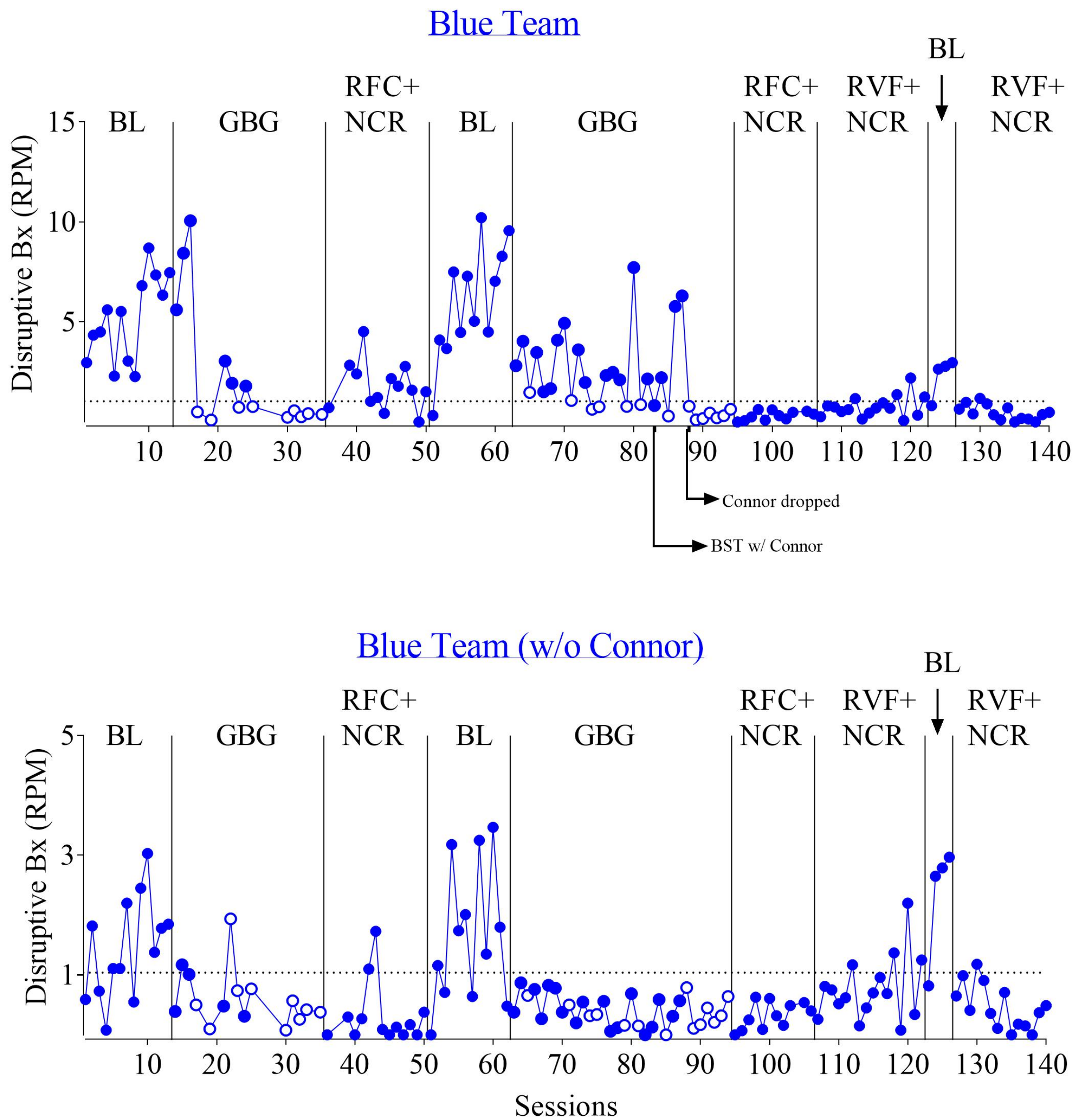


Figure 2. The top panel depicts the rate of disruptive behavior for the Blue Team with Connor's data included through session 87. The bottom panel depicts the rate of disruptive behavior for the Blue Team with Connor's data removed. The dotted line represents an 80% reduction in disruptive behavior from baseline levels. Open data points during GBG sessions denote sessions in which the team won the GBG. BL=Baseline, GBG=Good Behavior Game, RFC+NCR= Rules + Vocal & Visual Feedback + Criterion + Noncontingent Reinforcement, RVF+NCR= Rules + Vocal Feedback + Noncontingent Reinforcement

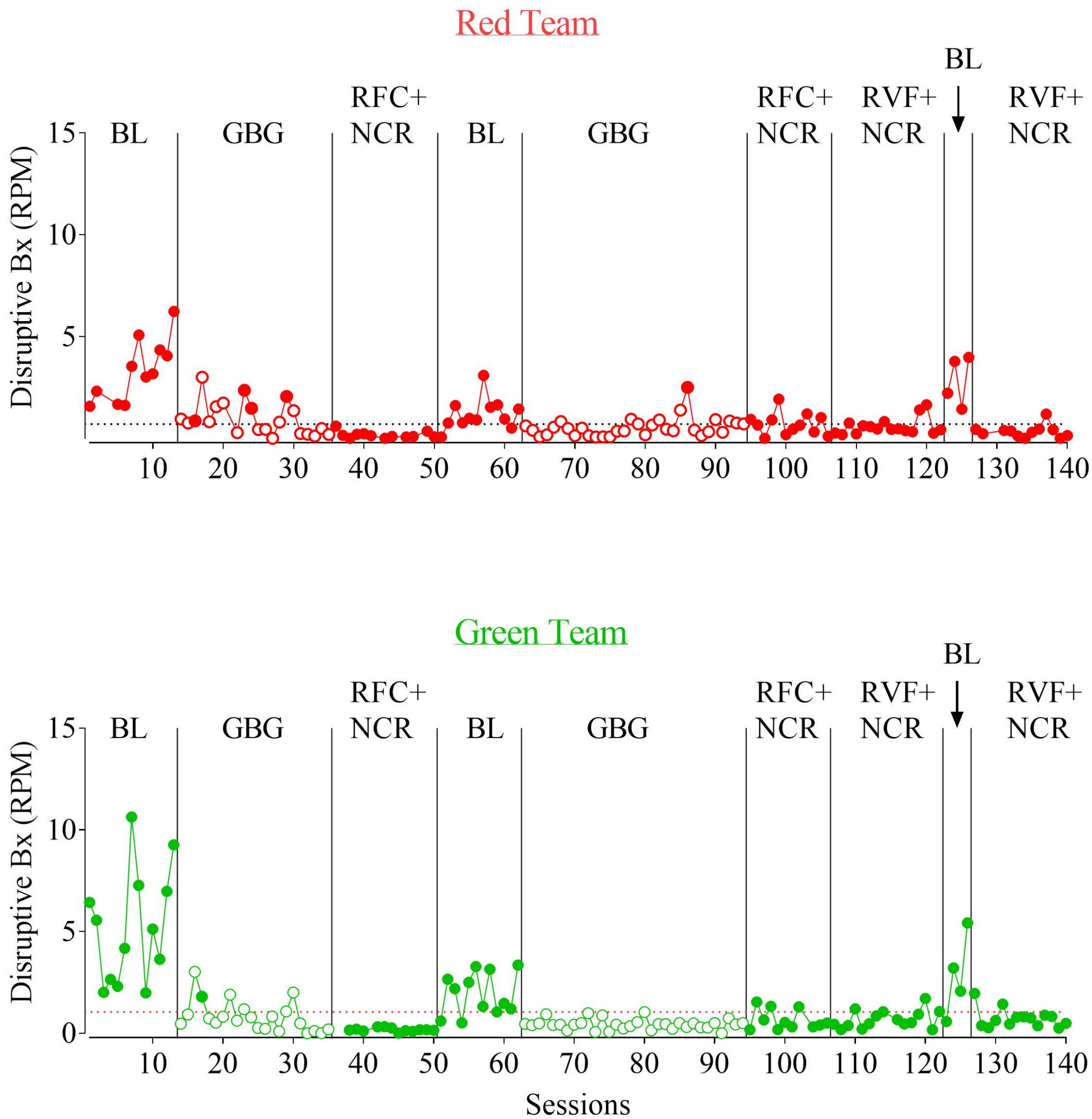


Figure 3. The top panel depicts the rate of disruptive behavior for the Red Team. The bottom panel depicts the rate of disruptive behavior for the Green Team. The dotted line represents an 80% reduction in disruptive behavior from baseline levels. Open data points during GBG sessions denote sessions in which the team won the GBG. BL=Baseline, GBG=Good Behavior Game, RFC+NCR= Rules + Vocal & Visual Feedback + Criterion + Noncontingent Reinforcement, RVF+NCR= Rules + Vocal Feedback + Noncontingent Reinforcement

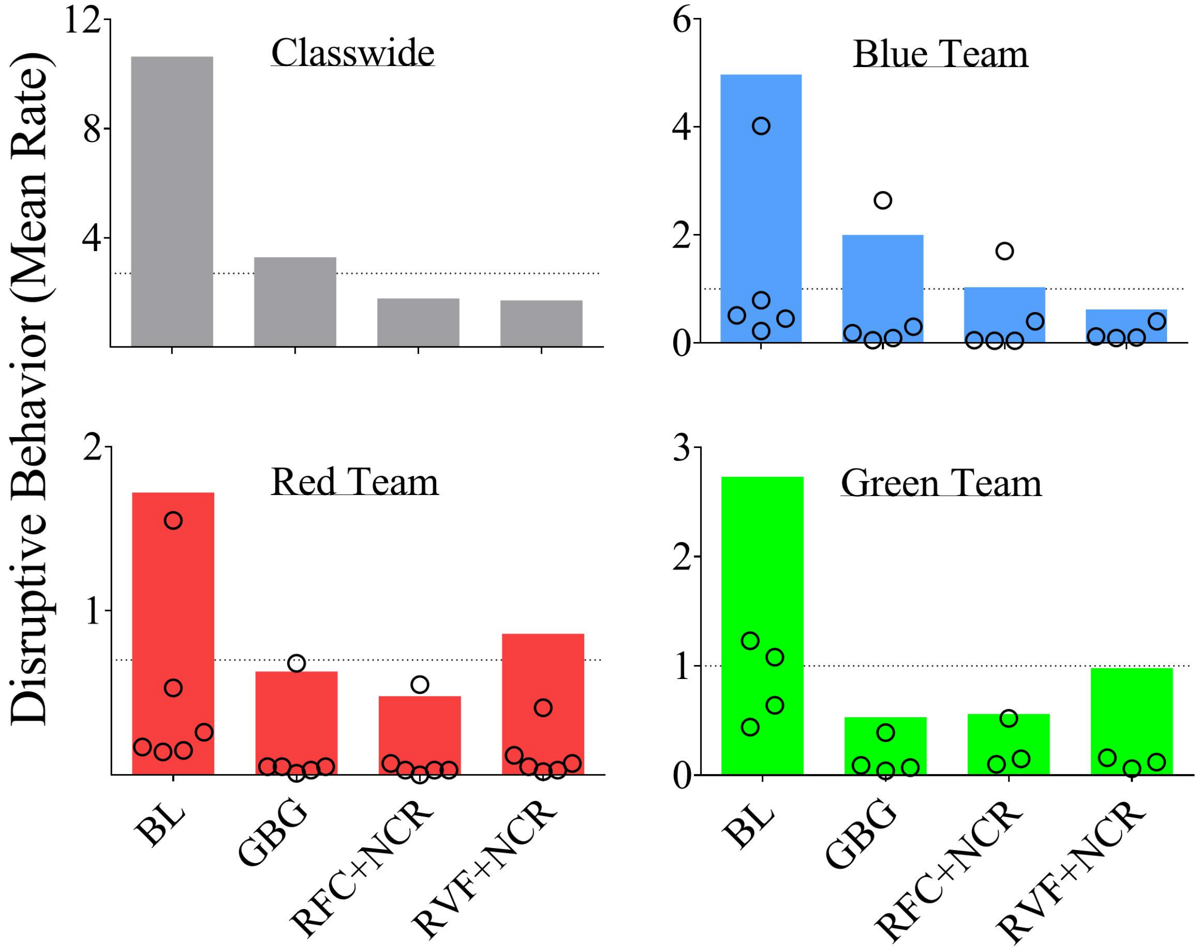


Figure 4. The layout depicts the mean rate of disruptive behavior for the class and all teams. The bars denote the means for the class and each of the three teams. The open circles denote the mean rate for each individual on the team. The dotted line represents an 80% reduction in disruptive behavior from baseline levels. BL=Baseline, GBG=Good Behavior Game, RFC+NCR= Rules + Vocal & Visual Feedback + Criterion + Noncontingent Reinforcement, RVF+NCR= Rules + Vocal Feedback + Noncontingent Reinforcement

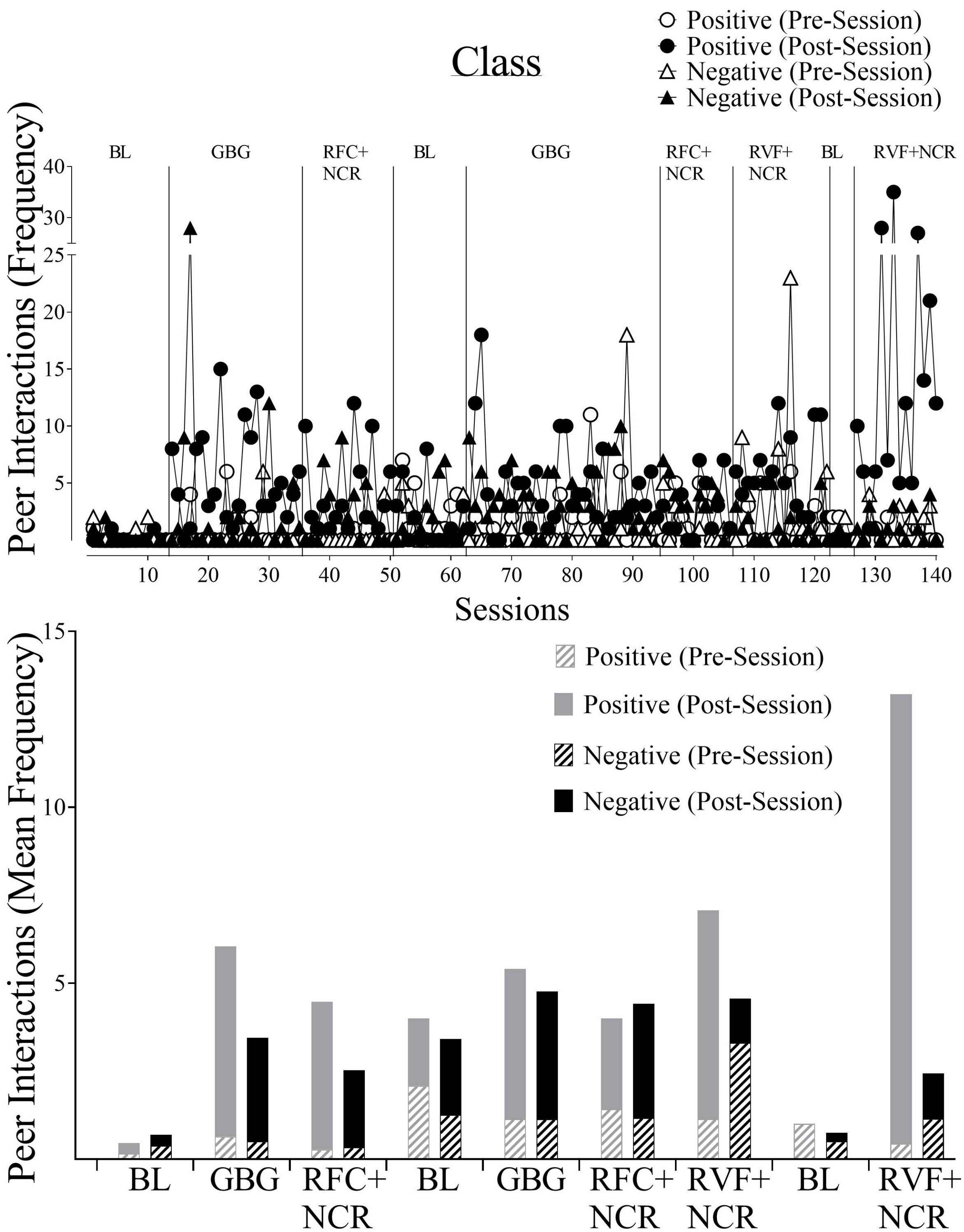


Figure 5. The top panel depicts frequency of positive and negative interactions for the class. The bottom panel depicts the mean frequency of positive and negative interactions for each phase for the class. The grey bars denote positive interactions. The black bars denote negative

interactions. The hatched section of each bar denotes mean interactions during the pre-session. The solid section of each bar denotes mean interactions during the post-session. The black bars denote negative interactions. BL=Baseline, GBG=Good Behavior Game, RFC+NCR= Rules + Vocal & Visual Feedback + Criterion + Noncontingent Reinforcement, RVF+NCR= Rules + Vocal Feedback + Noncontingent Reinforcement

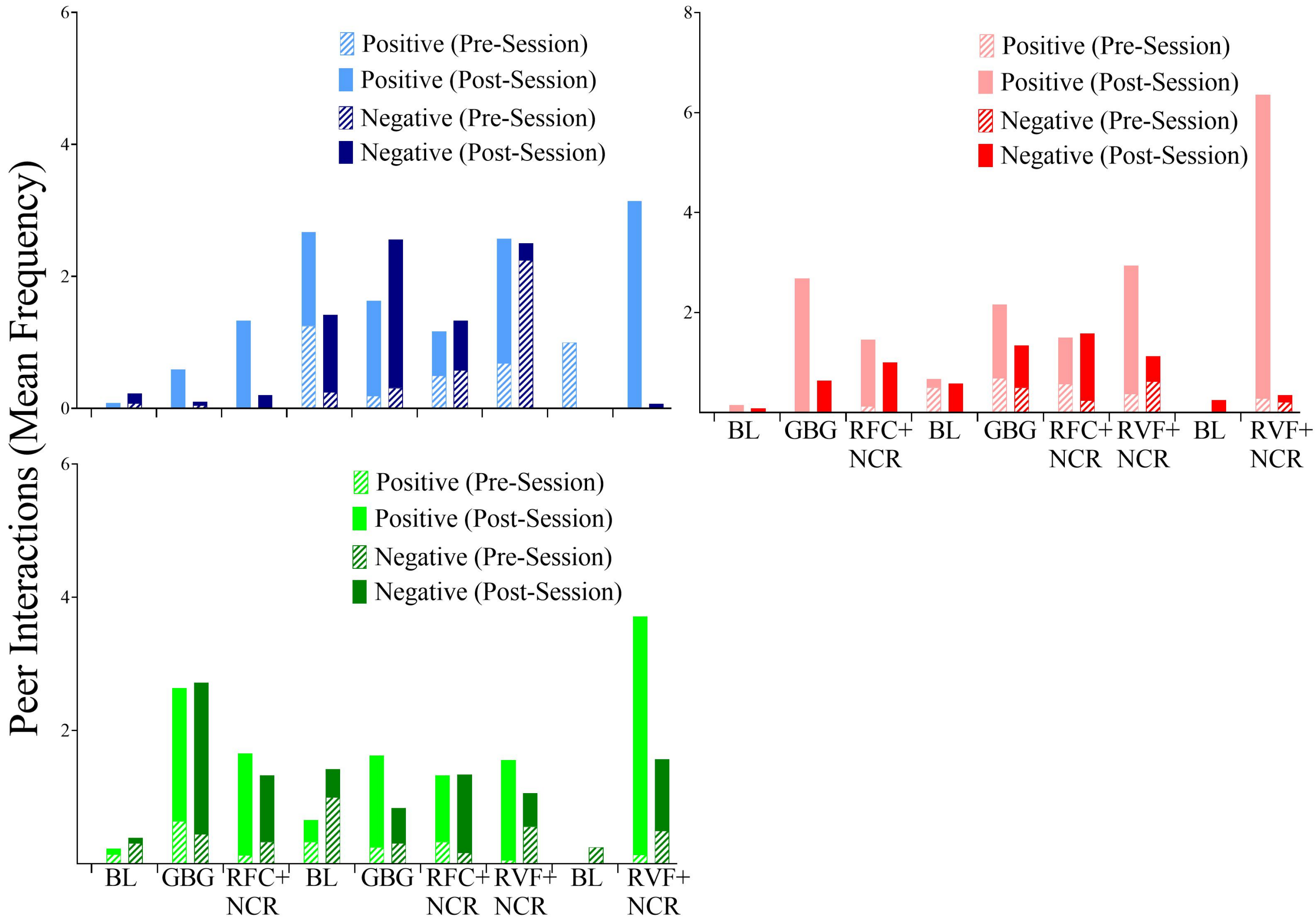


Figure 6. The layout depicts the mean frequency of positive and negative interactions for each phase for each of the teams. The top left graph depicts data for the Blue Team. The top right graph depicts data for the Red Team. The bottom graph depicts data for the Green Team. The light-colored bars denote positive interactions. The dark-colored bars denote negative interactions. The hatched section of each bar denotes mean interactions during the pre-session. The solid section of each bar denotes mean interactions during the post-session. BL=Baseline, GBG=Good Behavior Game, RFC+NCR= Rules + Vocal & Visual Feedback + Criterion + Noncontingent Reinforcement, RVF+NCR= Rules + Vocal Feedback + Noncontingent Reinforcement

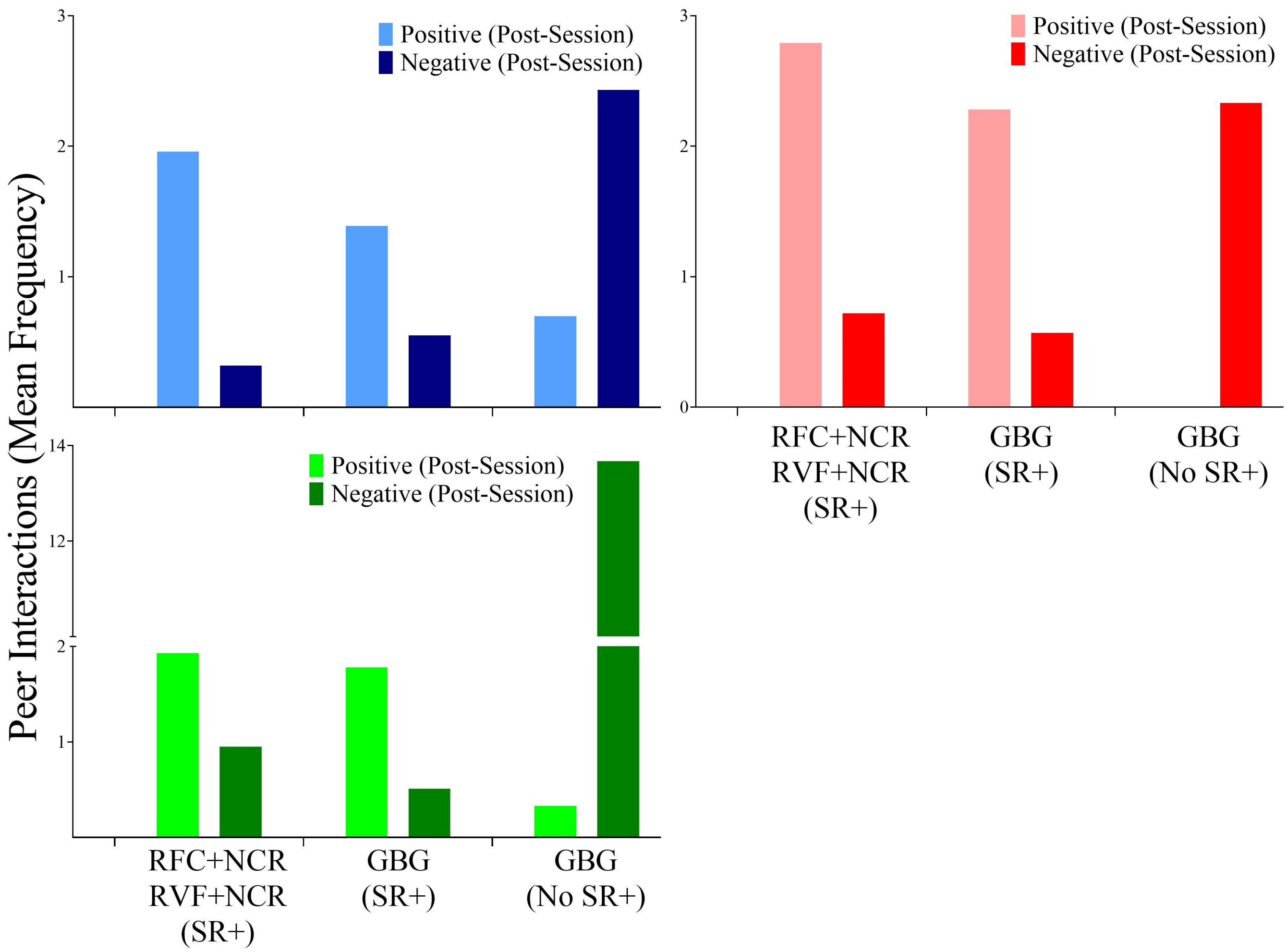


Figure 7. The layout depicts the mean frequency of positive and negative interactions for post-sessions associated with contingent and noncontingent delivery of rewards for each of the teams. The top left graph depicts data for the Blue Team. The top right graph depicts data for the Red Team. The bottom graph depicts data for the Green Team. The light-colored bars denote positive interactions. The dark-colored bars denote negative interactions. GBG=Good Behavior Game, RFC+NCR= Rules + Vocal & Visual Feedback + Criterion + Noncontingent Reinforcement, RVF+NCR= Rules + Vocal Feedback + Noncontingent Reinforcement

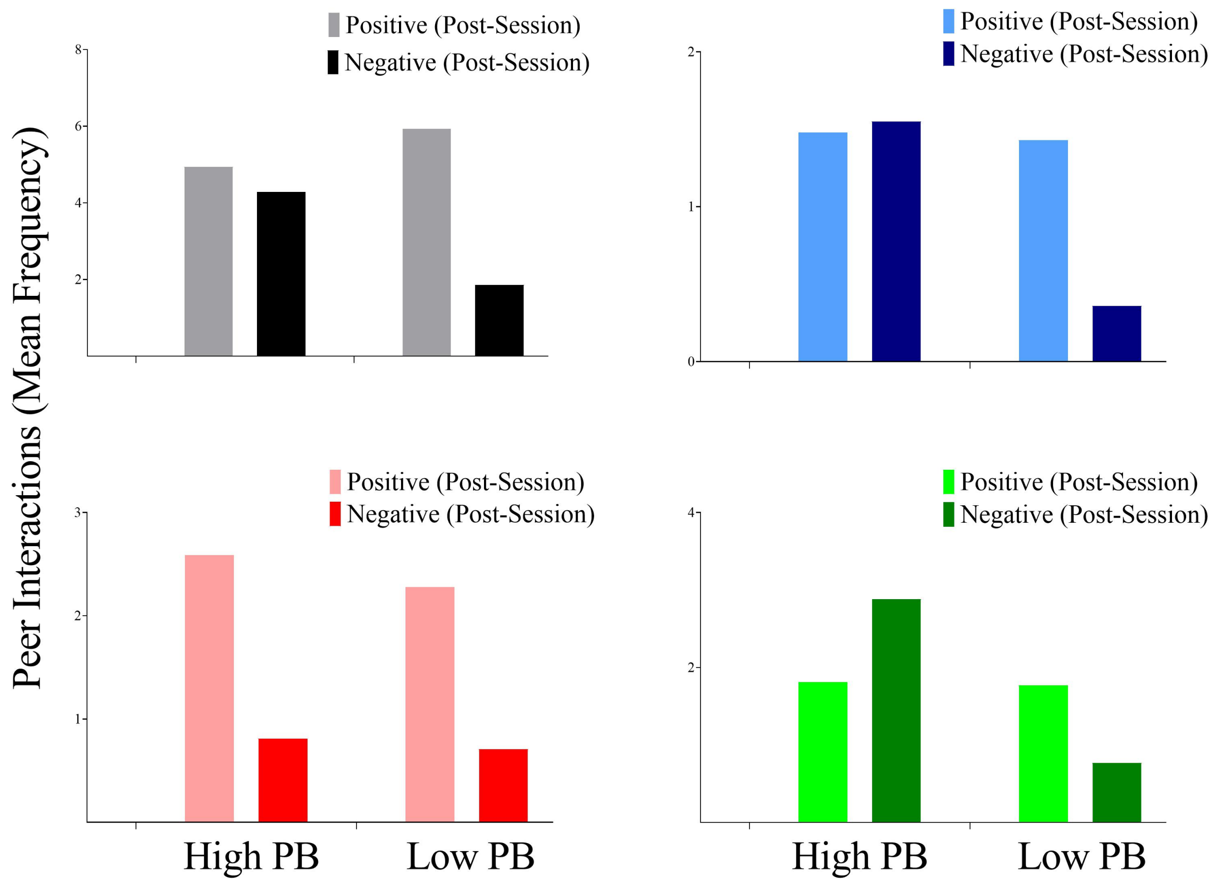
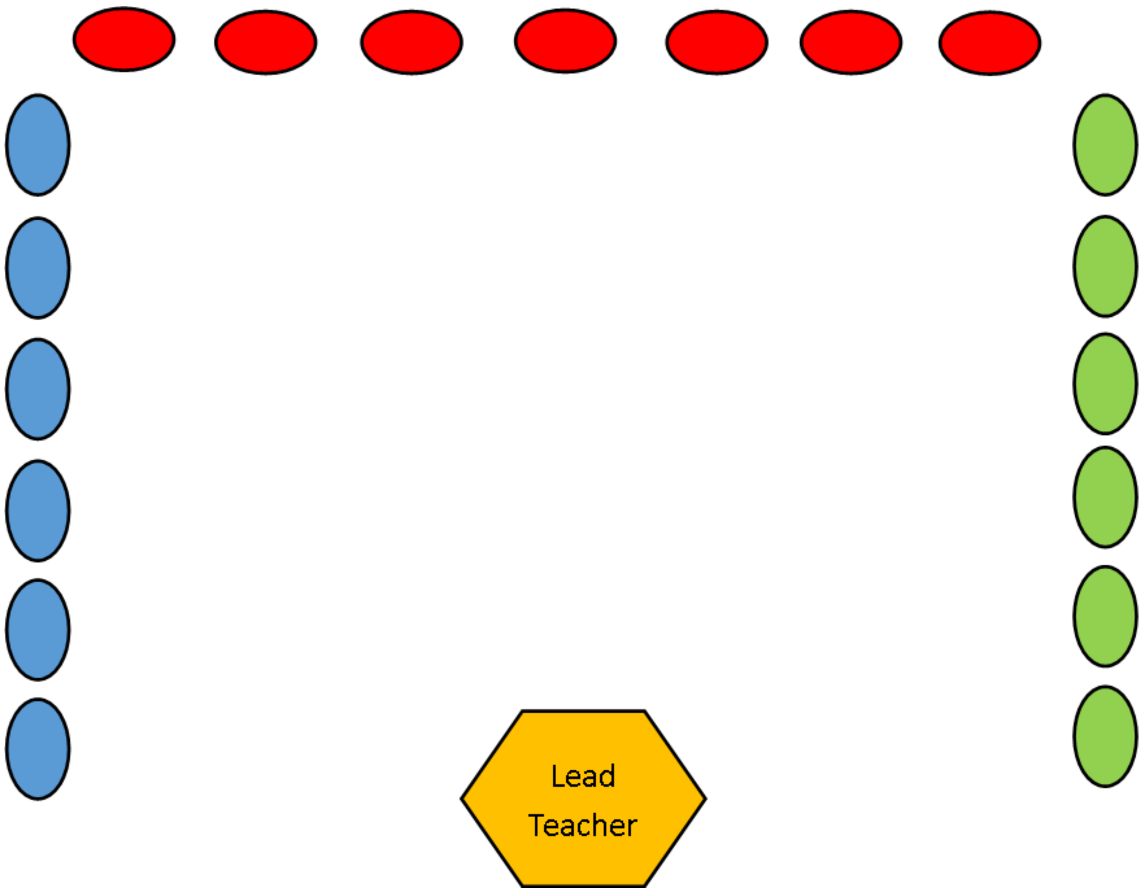


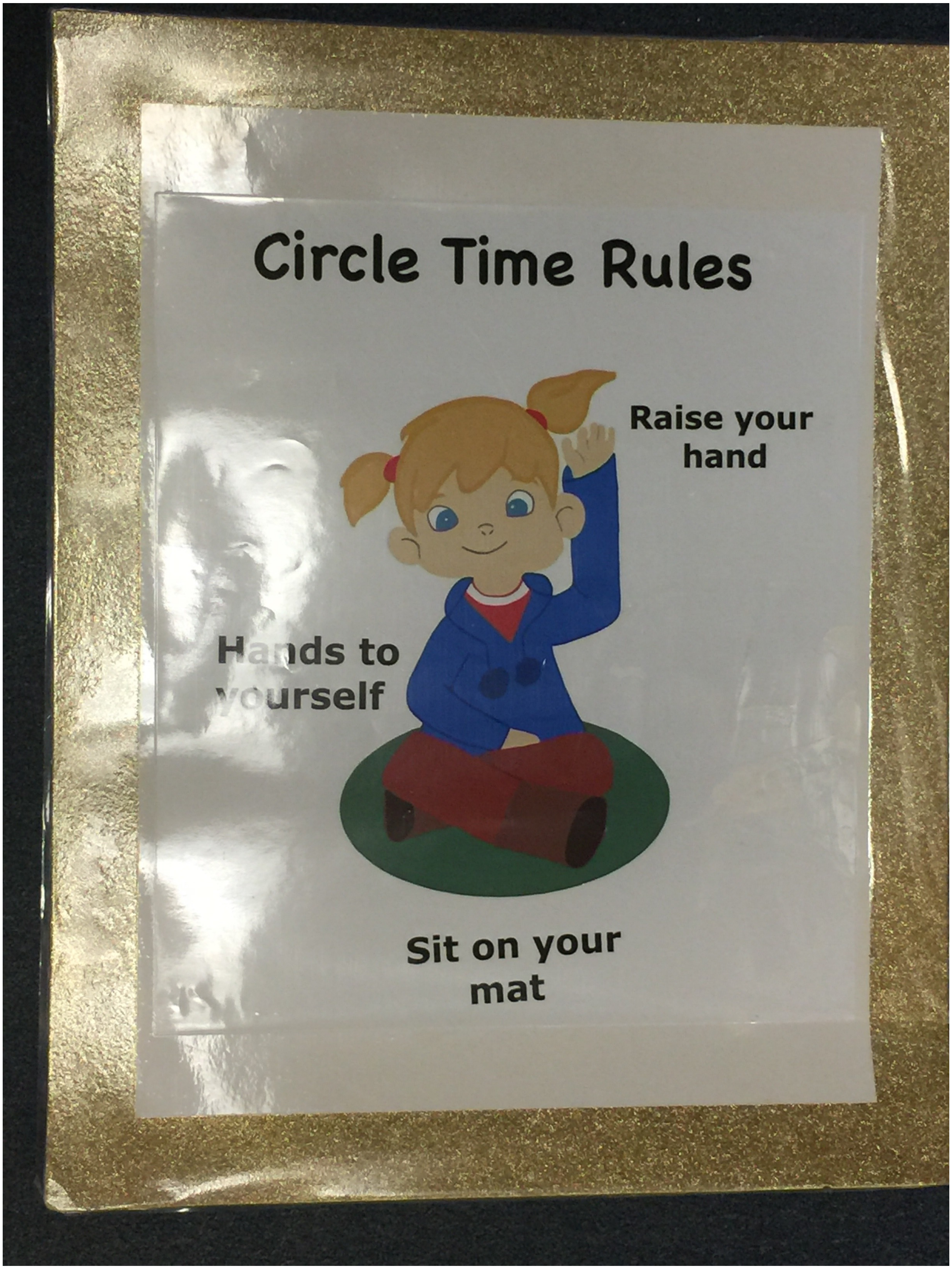
Figure 8. The layout depicts the mean frequency of positive and negative interactions during post-sessions associated with sessions in which high and low rates of disruptive behavior occurred for the class and each of the teams. The top left graph depicts data for the class. The top right graph depicts data for the Blue Team. The bottom left graph depicts data for the Red Team. The bottom right graph depicts data for the green. The light-colored bars denote positive interactions. The dark-colored bars denote negative interactions. High PB=sessions with less than an 80% reduction in disruptive behavior, Low PB=sessions with greater than or equal to an 80% reduction in disruptive behavior

Appendices

Appendix A



Appendix B.


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

Appendix C.

Appendix D.

Date:	Session #:	Circle Time	Observer:
Duration:	Phase:	AM PM	P / R
Group: Blue			Lead Teacher:

Child Initials	Target Bx	Intervals																				Totals
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
	IVB																					
	Out of Seat																					
	Touching																					
	IVB																					
	Out of Seat																					
	Touching																					
	IVB																					
	Out of Seat																					
	Touching																					
	IVB																					
	Out of Seat																					
	Touching																					
	IVB																					
	Out of Seat																					
	Touching																					
	IVB																					
	Out of Seat																					
	Touching																					

Team Rate:	
IVB:	
Out of Seat:	
Touching:	
Total: Bx:	

	Intitials	Date
Raw data entered:		

Experimenter Total:

Appendix E.

Date:	Session #:	Circle Time	Observer:
Group: Green	Phase:	AM PM	P / R
Pre-session / Post-session	SR+ Y N N/A		Lead Teacher:

Child Initials	Target Bx	Tally	Totals
	Pos Int		
	Neg Int		
	Pos Int		
	Neg Int		
	Pos Int		
	Neg Int		
	Pos Int		
	Neg Int		
	Pos Int		
	Neg Int		
	Pos Int		
	Neg Int		
		Initials	Date
	Raw Data Entered		

Anecdotal Notes (e.g., topographies & examples)

Appendix F.

Date:_________ **Session#:_________** **Phase: ________** **Experimenter:___________**

	GBG Component	
1	Presented rules poster board to children.	+ - Ø
2	Reviewed each rule on the poster board with the children.	+ - Ø
3	Set up rules poster board next to teacher or in a position clearly visible to all children.	+ - Ø
4	Uses dry-erase board to deliver hatch marks for rule violations.	+ - Ø
5	Announce which teams won at the end of circle.	+ - Ø
6	Deliver rewards to winning teams.	+ - Ø
	RFC+NCR Component	
1	Presented rules poster board to children.	+ - Ø
2	Reviewed each rule on the poster board with the children.	+ - Ø
3	Set up rules poster board next to teacher or in a position clearly visible to all children.	+ - Ø
4	Uses dry-erase board to deliver hatch marks for rule violations.	+ - Ø
5	**Does not** announce which teams won at the end of circle.	+ - Ø
6	Deliver rewards to all teams.	+ - Ø
	RVF+NCR Component	
1	Presented rules poster board to children.	+ - Ø
2	Reviewed each rule on the poster board with the children.	+ - Ø
3	Set up rules poster board next to teacher or in a position clearly visible to all children.	+ - Ø
4	Delivers vocal feedback for rule violation (**does not** use dry-erase board to deliver hatch marks for rule violations)	+ - Ø
5	**Does not** announce which teams won at the end of circle.	+ - Ø
6	Deliver rewards to all teams.	+ - Ø

+ Correctly implemented/omitted the component - Made an error when implementing the component

Ø Did not complete a component when it should have been implemented

Vocal Feedback Delivered

Red Team	*Blue Team*	*Green Team*