4-1-2003

The Clinical Application of the Good Behavior Game in Three Elementary Classrooms

Keri K. Ramaekers
University of Nebraska at Omaha

THE CLINICAL APPLICATION OF THE GOOD BEHAVIOR GAME IN THREE ELEMENTARY CLASSROOMS

Ed.S. Field Project

Presented to the

Department of Psychology

And the

Faculty of the Graduate College

University of Nebraska

In Partial Fulfillment

Of the Requirements for the Degree

Specialist in Education, School Psychology

University of Nebraska at Omaha

by

Keri K. Ramaekers

April 2003

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

ED.S. FIELD PROJECT ACCEPTANCE

Acceptance for the faculty of the Graduate College,
University of Nebraska, in partial fulfillment of the
Requirements for the degree Specialist in Education, School Psychology,
University of Nebraska at Omaha.

Committee

Lisa Kelly-Vance - Co-Chair

Mark Shriver - Co-Chair

Melodee Landis

Chairperson Lisa Kelly-Vance & Mark Shriver - co-chairs

Date May 8, 2003

THE CLINICAL APPLICATION OF THE GOOD BEHAVIOR GAME IN THREE ELEMENTARY CLASSROOMS

Keri K. Ramaekers

University of Nebraska at Omaha, 2003

Advisor: Mark D. Shriver, Ph.D.

The clinical application of a version of the Good Behavior Game (GBG) to reduce disruptive behaviors in 1 fourth grade and 2 fifth grade regular education classrooms is evaluated. A multiple-baseline across classrooms and changing criterion design was utilized. Upon implementation of the GBG, disruptive behaviors were reduced for each classroom as a whole and for two students, individually referred for problem behaviors. Furthermore, disruptive behavior rates conformed to the changes in criterion levels implemented in 1 classroom. A modest but positive relationship between treatment integrity and degree of effectiveness of the GBG was supported. Overall, the GBG is an effective classroom and individual behavior management technique and is feasible for use in routine school psychology practice.

Acknowledgements

I would like to thank my advisor and co-chair of my research committee, Dr. Mark Shriver, for his supervision and assistance with the design, implementation, and reporting of the project. I would also like to thank the other members of my research committee, Dr. Lisa Kelly-Vance (co-chair) and Dr. Melodee Landis, for their assistance and suggestions regarding the project.

I thank Kimberly Bell for the many hours she devoted to assisting with the data collection for the project. I am greatly in debt to her. I also thank the teachers and students who participated in the project for their time, cooperation, and assistance.

Finally, I thank my family for their support throughout my undergraduate and graduate studies. I especially thank my husband, Steve, my daughter, Taylor, and my mother-in-law, Vi, for all the sacrifices they made this past year to allow me to complete this project.

Table of Contents

The Clinical Application of the Good Behavior Game in Three Elementary Classrooms

One challenge facing school psychologists is the selection and implementation, often indirectly, of behavioral interventions. Often, focus is placed on "fixing" the behavior problems of an individual child, who is referred by his/her teacher (Elliot, Witt, Kratochwill, & Stoiber, 2002). This narrow focus fails to consider the broader context in which the behavior occurs (i.e., the classroom). Additionally, school psychologists often encounter several children within a classroom who are exhibiting similar types of inappropriate behavior. Disadvantages to using individual interventions in such cases include the amount of planning time, the amount of record keeping, and the skills and time necessary to implement interventions with multiple students simultaneously (Dolan et al., 1993). A more appropriate focus might be on management of the classroom as a whole.

When several students within a classroom need to be targeted for a behavioral intervention, one available classroom management technique is the Good Behavior Game (GBG) (Barrish, Saunders, & Wolf, 1969). Barrish et al. originally developed the GBG premised on the hypothesis that a student's disruptive behavior in a classroom is the result of reinforcement by peers; therefore, if group reinforcement for inhibiting the disruptive behavior is used, the disruptive behavior will be reduced (Embry, 2002). Thus, the GBG involves an interdependent group-oriented contingency, in which all members of the class must meet the goal for inhibiting disruptive behaviors in order for each member to receive reinforcement (Cooper, Heron, & Heward, 1987). Overall, the GBG is

considered a positive reductive technique, as reinforcement, rather than punishment, is used to reduce disruptive behavior. Specifically, differential reinforcement of low rates of behavior is used. Students are reinforced for not exceeding a criterion number of allowed disruptions.

The GBG, as originally introduced, involves establishing classroom rules, dividing the students into teams, visibly recording rule infractions by each team, and rewarding teams that meet a pre-established criterion for the number of rule infractions allowed (Barrish et al., 1969). The GBG has been reported to be equally acceptable to teachers as individual intervention techniques (Tingstrom, 1994).

Recently, Embry (2002) promoted the GBG as a "behavioral vaccine", comparing it to antiseptic hand washing to prevent disease and seat belt use to prevent injuries. Embry recommended that the game be used universally in elementary schools as a preventive measure to reduce the rates of disruptive behavior problems that are linked to severe developmental consequences (i.e., adolescent substance abuse, violence, and delinquency) and to reduce the number of special education placements.

Although the GBG has existed for over thirty years and has accumulated some research support, few professionals know of and use the technique and additional research is needed, particularly if the GBG is to be promoted as a "behavioral vaccine". First, little research has been conducted on the GBG in the last decade. Additionally, existing literature includes teacher reports of direct implementation of the GBG and reports by psychologists of implementation of the GBG primarily for research purposes. Few reports exist, however, on the use of the GBG in response to an actual teacher

referral to the school psychologist for behavioral or general classroom management concerns. Additional reports on the current use and indirect implementation of the GBG by school psychologists in a consultative role could strengthen confidence in the efficacy and feasibility of the GBG in current, routine school psychology practice. Finally, although a few studies have used a multiple baseline design to examine the GBG's effectiveness (e.g., Barrish et al., 1969, Grandy, Madsen, & De Mersseman, 1973; Patrick, Ward, & Crouch, 1998), only one study was found that utilized a multiple baseline design across classrooms and teachers (i.e., Salend, Reynolds, & Coyle, 1989). This study, however, was conducted in a residential setting with emotionally disturbed adolescents, rather than in a typical regular education classroom. Additional examinations of this nature, in a regular education setting, could strengthen the support for the efficacy of the GBG.

The current paper will provide an introduction to the GBG and its variations. Research related to the efficacy, generalization and maintenance of effects, and acceptability of the GBG will be summarized. Various rationales behind the GBG will also be presented. Finally, the advantages versus disadvantages of using the GBG will be considered.

The GBG and its Variations

Although sharing basic components, several variations of the GBG have been successfully employed. As originally introduced, the GBG involves selecting target behaviors to reduce in frequency and developing classroom rules for inhibiting engagement in such behaviors (Barrish et al., 1969). Students are then divided into two

teams, and the rules are explained. A mark is recorded on the chalkboard, corresponding to the appropriate team, each time a rule is broken by any member of a team. At the end of the period during which the game is played, the team(s) with the fewest marks or the team(s) not exceeding a pre-established criterion for the maximum number of rule infractions allowed is rewarded. In general, natural reinforcers other than attention (e.g., special activities) are used. In addition to daily rewards, rewards for meeting a weekly criterion for the maximum number of rule infractions allowed can be used.

Several variations on these basic components of the GBG have been employed. Studies have varied in whether the rules of the game are worded negatively or positively and whether or not students are allowed to assist in determining the rules and rewards (e.g., Barrish et al., 1969; Fishbein & Wasik, 1981; Grandy et al., 1973; Saigh & Umar, 1983; Salend et al., 1989). Fishbein and Wasik hypothesized that positively wording rules may increase student motivation, reduce student resentment, and improve the teacher's attention to positive behavior. Salend et al. used individualized rules and criterion levels for each team. Some researchers have also chosen not to use teams or to use more than two teams (Grandy et al., 1973; Salend et al., 1989). Some have also chosen not to provide immediate feedback in the form of marks on the chalkboard or to have an observer other than the teacher record marks (Grandy et al., 1973; Medland & Stachnik, 1972). Additionally, Fishbein and Wasik chose to record marks based on following the rules rather than rule infractions, using a momentary time sampling technique. With regard to rewards, a grab bag reinforcement system, in which a reward is randomly selected from a variety of potential rewards, has been used successfully (Salend et al.,

1989), as well as rewards administered by a teacher other than the teacher implementing the game (Fishbein & Wasik, 1981). Overall, applications of these variations on the original GBG have resulted in reductions in disruptive behaviors; thus, some evidence exists for the flexibility of the GBG in meeting the needs of a particular classroom.

Critical components.

Because many variations on the basic components of the GBG have been utilized, it is important to examine what components are critical to the success of the intervention. Although only a handful of studies have examined the issue, some conclusions can be drawn. First, the establishment and communication of classroom rules has been shown to be a critical component (Medland & Stachnik, 1972). Medland and Stachnik found that establishment and communication of rules, alone, resulted in a decrease in disruptive behavior below baseline levels. Further reductions were achieved with the addition of the other components of the game, however. In general, rules are recommended as the first step in the development of any classroom management system (Evertson, 1987; McGinnis, Frederick, & Edwards, 1995). Rules serve a preventative function, fostering an environment conducive to learning by establishing, and making concrete, general behavior standards and by providing structure.

In addition to classroom rules, the group contingency (rewarding the group for inhibiting disruptive behavior), the component the GBG was premised on, appears crucial to the intervention's success (Fishbein & Wasik, 1981; Harris & Sherman, 1973; Medland & Stachnik, 1972). As noted previously, the GBG involves an interdependent group contingency. Addition of the group contingency component, after rules have been

established, has produced further decreases in disruptive behavior below baseline levels (Medland & Stachnik, 1972). On the other hand, removal of the group contingency component from the GBG has been shown to result in increases in disruptive behavior to initial levels (Fishbein & Wasik, 1981; Harris & Sherman, 1973).

Finally, the criterion level, or the maximum number of rule infractions a team can engage in and still receive a reward, has been shown to be a component that affects the level of reduction in disruptive behavior (Harris & Sherman, 1973). In the Harris and Sherman study, levels of disruptive behavior conformed to the criterion level set for winning the game. In other words, when a lower criterion was set, greater reduction in inappropriate behavior occurred and vice versa.

In contrast to the above components, the necessity of the components of immediate, visual feedback provision (e.g., marks on the chalkboard) and the use of teams is questionable. Medland and Stachnik (1972), in their systematic component analysis, found greater decreases in inappropriate behavior when immediate visual feedback on performance was provided than when feedback was only given in the form of announcing the winner(s) at the end of the class period. In contrast, Harris and Sherman (1973), in their systematic component analysis, did not find any effect for the immediate feedback component.

With regard to the use of teams, Grandy et al. (1973) maintained the use of a group contingency but did not use teams when implementing the GBG. Rates of disruptive behavior were successfully reduced. Harris and Sherman (1973), in their systematic component analysis, however, found the team component critical. The authors

also implemented a "tie" component, however. For this component, if both teams exceeded the criterion on a given day, the teams could still win the game if a "tie" occurred in the number of infractions. The authors hypothesized that this component was responsible for the effectiveness of the team component. The "tie" component kept a team that exceeded the criterion from further increasing their level of disruptive behavior dramatically.

Research on the GBG

Efficacy. Existing evidence suggests the GBG might be effective across behaviors, grade levels, subjects, settings, student characteristics, and cultures (Embry, 2002; Saigh & Umar, 1983). In general, when the GBG is implemented, disruptive behaviors are reduced, and this effect is reversed when the GBG is removed (e.g., Barrish et al., 1969; Grandy et al., 1973; Medland & Stachnik, 1972; Saigh & Umar, 1983; Salend et al., 1989). Evidence suggests that the GBG can be effective in reducing a variety of behaviors, including out-of-seat behavior, talking without permission, general off-task behavior, negative verbal comments and cursing, offensive gestures, and physical aggression (e.g., Barrish et al., 1969; Fishbein & Wasik, 1981; Patrick et al., 1998; Saigh & Umar, 1983; Salend et al., 1989). Substantial and meaningful reductions in these behaviors have been reported. For example, decreases of 48 – 99% in out-of-seat behavior and talking without permission have been reported (Barrish et al., 1969; Harris & Sherman, 1973; Medland & Stachnik, 1972). Some studies report reductions in these two behaviors to near zero levels (e.g., Barrish et al., 1969; Grandy et al., 1973; Harris & Sherman, 1973; Medland & Stachnik, 1972).

Existing evidence also suggests that the GBG might be successfully employed across grade levels and subjects (e.g., Barrish et al., 1969; Dolan et al., 1993; Harris & Sherman, 1973; Salend et al., 1989). The game has typically been utilized with students in the fourth through sixth grades; however, some evidence exists supporting the game's efficacy with younger students (Dolan et al., 1993; Saigh & Umar, 1983), as well as with adolescents in a residential special education setting (Salend et al., 1989). The GBG has been used to reduce disruptive behavior in a variety of general education classes, including reading, English, mathematics, science, and spelling (Barrish et al., 1969; Grandy et al., 1973; Harris & Sherman, 1973; Medland & Stachnik, 1972).

The GBG might also be able to be implemented successfully in settings outside these typical general education classes. Fishbein and Wasik (1981) employed a unique variation of the GBG during library class. Using a momentary time sampling technique, the librarian scanned the class on four occasions during the period and recorded a point for a team if all members were engaging in the appropriate behaviors being targeted. The homeroom teacher, rather than the librarian, administered the reinforcement after the library period. Patrick et al. (1998) implemented the GBG in a physical education setting. The researchers successfully reduced inappropriate and increased appropriate social behaviors during a volleyball unit. Finally, Salend et al. (1989) implemented the GBG in a residential special education setting for emotionally disturbed adolescents. Target behaviors and criteria for winning were individualized across teams, to permit teachers to better address the behaviors of primary concern for individual students. A variety of disruptive behaviors were reduced.

The work by Salend et al. (1989), described above, also suggests the potential utility of the GBG across student characteristics. Additionally, Medland and Stachnik (1972) examined the effectiveness of the GBG with four individuals who exhibited the greatest rates of disruptive behavior in a regular education classroom and found that, after implementation of the GBG, these rates decreased to near zero level. Thus, students in both regular and special education placements, including students who exhibit high rates of disruptive behavior, might benefit from use of the GBG. Taken together, the above studies suggest that the GBG may be a flexible intervention technique.

In addition to evidence suggesting flexibility across behaviors, grade levels, subjects, settings, and student characteristics, some evidence supports the cross-cultural utility of the GBG (Saigh & Umar, 1983). Saigh and Umar effectively used the GBG to reduce out-of-seat, talking without permission, and aggressive behavior in a Sudanese classroom. Staff, students, and parents reported satisfaction with the GBG. Additionally, researchers have reported using the GBG with Caucasian and African American students, as well as with students from other, unspecified backgrounds (Dolan et al., 1993; Grandy et al., 1973).

Rationale behind the GBG. Researchers have offered a variety of rationales as to why the GBG has the potential to be an effective intervention technique. Medland and Stachnik (1972) hypothesized that, while focusing on the reduction of disruptive behavior, the GBG may also promote positive, competing behaviors, which might then be coincidentally reinforced. The increase in positive, competing behaviors may then increase the effectiveness of the game and the likelihood of generalization. Dolan et al.

(1993) provided a second rationale, suggesting that increased social awareness and a sense of competence may be fostered and may be important for maintenance and generalization of effects.

A rationale provided by several researchers is that the GBG contains an element of positive peer pressure (Dolan et al., 1993; Embry, 2002; Patrick et al., 1998; Salend et al., 1989). As Patrick et al. stated, "Focusing on the group rather than on specific individuals in the good behavior game allows those individuals who have fewer social skills to be in the presence of peers who can model correct behavior and who are present at the time of an inappropriate behavior to discourage it" (p. 151-152). As noted previously, the GBG was developed based on the hypothesis that disruptive behavior is the result of social reinforcement and that providing group reinforcement for inhibiting such behavior will reduce the rates of the disruptive behavior (Embry, 2002). Research support exists for the importance of decreasing peer reinforcement for disruptive behavior and increasing positive peer pressure (Embry, 2002). Overall, regardless of additional contributing factors, the most parsimonious argument for the potential of the GBG to be effective is likely Thorndike's "Law of Effect", which asserts that " . . . behavior varies as a function of its consequences (Embry, 2002, p. 279).

Generalization and maintenance of effects. Examination of research on the GBG reveals some evidence for generalization and maintenance of effects. In examining generalization, studies employing a reversal design, with a return to baseline conditions being instituted following implementation of the game, are useful. In using such a design, some researchers have found that rates of disruptive behavior return to initial baseline

levels after the game is withdrawn (e.g., Barrish et al., 1969; Salend et al., 1989). Others, however, have found that rates of disruptive behavior show only modest recovery, suggesting the potential for modest generalization of effects (Grandy et al., 1973; Medland & Stachnik, 1972).

Additional evidence regarding generalization of effects comes from the study by Fishbein and Wasik (1981). Although the authors did not provide specific data, it was noted that while the GBG was being implemented during library class, improvements in behavior were noted in the primary classroom, as well. The authors stated, however, that this generalization of effects might have been the result of the primary teacher delivering the rewards for winning the game in the library. Overall, although some evidence suggests a potential for generalization of effects, if one desires to reduce disruptive behavior during additional class periods or in additional settings, it is best to directly expand implementation of the GBG to such periods and settings. Some researchers have implemented such expansions with success (e.g., Dolan et al., 1993; Harris & Sherman, 1973).

With regard to maintenance, the studies reviewed suggest the potential for persistence of effects across as many as 13 weeks (e.g., Barrish et al., 1969; Fishbein & Wasik, 1981; Harris & Sherman, 1973). There is a severe lack of long-term follow-up data, however. Embry (2002) claimed that a few longitudinal studies suggest that use of the GBG during the elementary years can alter the developmental trajectory of behavior problems, preventing later substance abuse, violence, and delinquency. Clearly, more data on long-term maintenance of effects is needed.

Acceptability of the GBG. In choosing whether or not to implement any intervention, consideration should be given to the acceptability, by consumers, of the intervention (Elliot et al., 2002; Tingstrom, 1994). An interrelationship, although imperfect, is hypothesized by some to exist between acceptability, use, integrity, and effectiveness (e.g., Elliot et al., 2002; Gresham, 1989; Gresham & Gansle, 1993; Sterling-Turner, Watson, Wildmon, Watkins, & Little, 2001). It is hypothesized that if an intervention is acceptable, the likelihood is increased that consumers will use it with integrity and, thus, that the intervention will be effective.

In general, the GBG is reported to be acceptable to consumers (e.g., Barrish et al., 1969; Saigh & Umar, 1983; Salend et al., 1989; Tingstrom, 1994). Teachers have reported that the GBG is easier to implement than individual contingencies (Grandy et al., 1973). Group contingencies, in general, are also typically acceptable to students (Elliot et al., 2002).

Tingstrom (1994) empirically examined teacher acceptance (social validity) of the GBG in two analogue studies. Teachers were provided with written scenarios describing students, behavior problems, and either an individual intervention or the GBG. Teachers then rated the interventions according to acceptability, use, and effectiveness. Overall, teachers rated the traditional version of the GBG as equally acceptable as individual interventions. Additionally, the GBG was rated acceptable regardless of the age of students or severity of the behavior problems described. Finally, acceptability ratings were positively related to use and effectiveness ratings.

Summary of the Potential Advantages of the GBG

Overall, existing evidence suggests that the GBG might be a flexible, low cost, effective intervention strategy when several students in a classroom exhibit disruptive behaviors and might be an easier strategy than implementing multiple, simultaneous, individual interventions in such cases (Dolan et al., 1993; Salend et al., 1989; Tingstrom, 1994). There is some evidence that the GBG has worked across behaviors, grade levels, subjects, settings, student characteristics, and cultures (Embry, 2002; Saigh & Umar, 1983). Although treatment integrity is hypothesized to increase the likelihood of an intervention being effective, the GBG is reported to sometimes be effective even when treatment integrity is not perfect (Embry, 2002). This fact provides an important benefit when school psychologists are implementing the intervention indirectly, in a consultative manner. Harris and Sherman (1973) provided three reasons for the ease in implementation of the GBG, including the fact that the teacher does not have to differentially provide consequences for individual students, the GBG incorporates an easy system for recording a variety of behaviors of concern, and the teacher is the only observer required to implement the GBG.

Another potential benefit of the GBG is its reported acceptability to consumers (Barrish et al., 1969; Grandy et al., 1973; Saigh & Umar, 1983; Salend et al., 1989; Tingstrom, 1994). Furthermore, the GBG typically relies on the provision of natural, activity reinforcers. The use of activity reinforcers increases consumer acceptability (Embry, 2002). The fact that little consultation time is required, a characteristic of interventions preferred by consultees, according to Elliot et al. (2002), also increases the

potential for the acceptability of the GBG.

Finally, Elliot et al. (2002), in their discussion of appropriate methods for selecting intervention strategies, recommended that antecedent control strategies be included in interventions. Although the primary focus is on consequent control strategies, the GBG includes recommended antecedent control tactics, as well. For example, students are initially instructed in the rules of the game and are then reminded, either verbally or visually, of the rules at the beginning of each game period. Additionally, eye contact between the teacher and students is naturally fostered by the need for the teacher to monitor behavior.

Potential Disadvantages of the GBG

As with any intervention technique, there are potential disadvantages of the GBG to be considered. First, the teacher has to be alert to the behaviors of several students in addition to teaching (Barrish et al., 1969). Teachers should be monitoring classroom behavior as part of their daily practice, however. Additionally, monitoring the same behaviors across all students is likely easier than monitoring individualized target behaviors, as would be the case if individual interventions were used.

Other potential disadvantages exist, including the fact that students engaging in appropriate behaviors lose rewards sometimes, as a result of the performance of their team members (Salend et al., 1989). Additionally, potential concerns exist regarding those students who exhibit the highest rates of disruptive behaviors. The potential for negative peer pressure or punishment of these students by others has been expressed as a concern; however, no actual occurrences have been reported (Patrick et al., 1998; Salend

et al., 1989).

Related problems involve the students who exhibit high rates of disruptive behaviors accumulating large numbers of demerits for their team and/or refusing to play the game (e.g., Barrish et al., 1969; Harris & Sherman, 1973; Medland & Stachnik, 1972). Several potential solutions to these problems have been suggested, however. First, Barrish et al. dropped such students from their teams and recorded their behavior individually. Second, Medland and Stachnik added a preventive component that allowed the class to vote students who accumulated four or more demerits off of the team for the following day. This component never proved necessary to implement, as no students accumulated four or more demerits. Third, Harris and Sherman implemented an easy and highly promising preventive technique. As previously described, if both teams exceeded the criterion for the day, the teams could still win if a "tie" occurred; therefore, a team that exceeded the criterion was prevented from further increasing disruptive behaviors dramatically. With regard to students refusing to play the game, Harris and Sherman placed such students on a team of their own and instituted consequences for exceeding the criterion (i.e., staying after school).

Another potential disadvantage to the GBG is the lack of a demonstrated relationship between success of the GBG in reducing disruptive behaviors and improvements in academic performance (Dolan et al., 1993; Embry, 2002; Harris & Sherman, 1973; Patrick et al., 1998). As Patrick et al. emphasized, however "...the development of social skills does not require the rationale of improved student learning" (p. 153). The development of social skills is an important educational goal in itself.

Additionally, Starkweather-Lund and Shriver (2003) suggested that classroom management, in general, increases academic engagement (e.g., paying attention, raising one's hand to request assistance) and academic responding (e.g., answering teachers' questions, task completion, reading), which are positively related to academic achievement. In their study, Starweather-Lund and Shriver found that effective teacher commands paired with verbal praise resulted in improved compliance with teacher commands and increases in academic engagement and academic responding. It is possible that continued research on the GBG will reveal similar relationships. Finally, if nothing else, the reduction of disruptive behavior provides the context for other, academic interventions to be successful (Embry, 2002).

In summary, as with any intervention strategy, potential disadvantages of the GBG must be considered. The potential benefits of the GBG, however, appear to outweigh the potential disadvantages reported. Additionally, procedures intended to combat some of the potential disadvantages have been described and might prove useful.

Summary and Conclusions

In summary, when several students in a classroom exhibit behaviors of concern, a classroom-wide intervention may be warranted. A good choice for such an intervention might be the GBG. Existing evidence suggests its acceptability to consumers (e.g., Tingstrom, 1994) and its ease (e.g., Harris & Sherman, 1973) and flexibility (e.g., Fishbein & Wasik, 1981; Grandy et al., 1973; Salend et al., 1989) in implementation. Furthermore, existing evidence suggests that the GBG might be efficacious across behaviors, grade levels, subjects, settings, student characteristics, and cultures (e.g.,

Barrish et al., 1969; Embry, 2002; Fishbein & Wasik, 1981; Medland & Stachnik, 1972; Patrick et al., 1998; Saigh & Umar, 1983). Components critical to the success of the GBG appear to be the establishment and communication of classroom rules (Medland & Stachnik, 1972), the group contingency component, and the criterion level (e.g., Harris & Sherman, 1973). The most parsimonious explanation for the effectiveness of the GBG is likely the "Law of Effect" (Embry, 2002). Some evidence also exists for modest generalization and maintenance of the effects of the GBG (e.g., Fishbein & Wasik, 1981; Grandy et al., 1973; Medland & Stachnik, 1972). In general, the potential advantages of using the GBG appear to outweigh any potential disadvantages to its use.

In conclusion, although the existing evidence suggests the GBG might be a useful intervention technique, little research on the game has occurred in the past decade and few reports exist on the indirect implementation of the GBG in response to an actual teacher referral to the school psychologist. Additionally, few reports exist on the effectiveness of the GBG in reducing disruptive behavior for individuals. More information is needed in this regard to strengthen confidence in the efficacy and feasibility of using the GBG in current, routine school psychology practice. Furthermore, only one study was found that utilized a multiple baseline design across classrooms and teachers (i.e., Salend et al., 1989), and this study was not conducted in a regular education setting. Additional examinations of this type, in a regular education setting, could further strengthen the evidence for the GBG's effectiveness in a typical classroom environment.

Purpose of the Current Study

The current study reports on the clinical application of the GBG in three elementary classrooms. In two of the classrooms, individual student referrals initiated contact with the author. Two individual students, Darren and Andrew, identified as having Attention Deficit Hyperactivity Disorder and receiving special education services, were referred for disruptive behavior concerns in their regular education classrooms. Both students had a history of behavior problems and resulting suspensions and detentions. Andrew had experienced multiple school transfers as a result of behavior problems. Both students were being treated for ADHD symptoms and behavior concerns by an outside agency, where the author was doing practicum work and became involved in the students' cases.

Upon initial consultation with Darren and Andrew's classroom teachers, it was evident, through qualitative behavioral observations, that several students in the classrooms were engaging in the behaviors of concerns. The author and her colleagues believed that improving the classroom environment might improve Darren and Andrew's behavior. Furthermore, the teachers expressed a desire for assistance in managing several students' behaviors. Thus, a focus on classroom management rather than individual interventions appeared warranted. Finally, the third classroom involved a self-initiated referral by the teacher specifically for classroom management assistance. All three teachers indicated that the daily reading period was one of specific concern.

A multiple baseline design across classrooms and teachers was used to examine the effects of the GBG on reducing the frequency of disruptive behaviors below baseline

levels for each class, as well as for the two individuals described above. A changing criterion design was used in one classroom, as well, to provide further information on the effectiveness of the GBG.

Additionally, data on treatment integrity is reported. When implementing an intervention indirectly, in a consultative manner, it is important to examine the degree to which it is implemented as intended, referred to as treatment integrity (Gresham, 1989; Gresham & Gansle, 1993; Jones, Wickstrom, & Friman, 1997; Sterling-Turner et al., 2001). In their review of studies published between 1980-1990, in seven journals renowned for reporting on behavior interventions, Gresham and Gansle found that only 14.4% reported data on treatment integrity. Although conflicting results exist regarding the relationship of treatment integrity and effectiveness of an intervention (e.g., Gresham, 1989; Sterling-Turner et al., 2001; Wickstrom, Jones, LaFleur, & Witt, 1998), Gresham and Gansle's meta-analysis of the studies that did report treatment integrity data revealed at least a moderate but positive relationship between treatment integrity and successful intervention outcomes. The key issue remaining, however, is that the assessment of treatment integrity is necessary for establishing a functional relationship between an intervention and any resulting behavior change (internal validity) (Gresham, 1989; Gresham & Gansle, 1993). Thus, if the effectiveness of an intervention is to be evaluated, as in the current study, data on treatment integrity is crucial.

Expected Results

A decrease in the frequency of target behaviors below baseline levels was expected for each classroom, upon implementation of the GBG. Likewise, a decrease in

the frequency of target behaviors below baseline levels was expected for Darren and Andrew, upon implementation of the GBG. It was expected that the reduction in target behaviors would occur only at the point of intervention in a setting, without a concurrent reduction in settings where the GBG had yet to be introduced. It was further expected that the frequency of target behaviors would conform to any changes in the criterion level, based on the findings of Harris and Sherman (1973). Similar results were expected with expansion of the GBG to additional subjects. Finally, the goal for treatment integrity was to ultimately establish and maintain 100% adherence to the intervention components.

Method

Participants

Participants included staff and students in three regular education classrooms. Data was recorded for each classroom as a whole. Additionally, data was recorded for two individuals, referred to the author for behavior concerns.

Classroom A. Classroom A was a fifth grade classroom. Staff included a female, Caucasian teacher and her male, African American assistant. The teacher had completed a bachelor's degree plus 18 hours of graduate studies in education and had 20 years of teaching experience. Her assistant had five years of experience as a teacher's assistant. Students in classroom A included 10 boys and 8 girls. Approximately 83% of the students were African American, with a minority from Caucasian, Hispanic, and Asian backgrounds.

An individual student referral initiated contact between the author and the teacher in classroom A. Upon initial consultation, however, the teacher expressed concerns over

general classroom management. It was determined that a classroom wide intervention might be appropriate.

Classroom B. Classroom B was a fourth grade classroom. Staff included a female, Caucasian teacher, a female, Caucasian student teacher, and a female, African American assistant. The teacher had completed a bachelor's degree in education and her master's degree in the area of administration and supervision and had 17 years of teaching experience. The current school year was the first year of experience in the classroom for both the student teacher and the assistant. The student teacher was to complete her bachelor's degree in education at the end of the current semester. Students in classroom B included 8 boys and 8 girls. Approximately 88% of the students were African American, with a minority from Hispanic and Asian backgrounds.

As in classroom A, an individual student referral initiated contact between the author and the teacher in classroom B. Again, upon initial consultation, however, the teacher expressed concerns over general classroom management. It was determined that a classroom wide intervention might be appropriate for this classroom, as well.

Classroom C. Classroom C was a fifth grade classroom. The teacher initiated contact with the author by requesting assistance with classroom management. Staff included the male, African American teacher and his female, African American Assistant. The teacher had completed a bachelor's degree plus 21 hours of graduate studies in education and had four years of teaching experience. Prior to beginning his teaching career, this individual worked in social services for 11 years. His assistant had two years of experience as a teacher's assistant. Students in classroom C included 13 boys and 10

girls. Approximately 90% of the students were African American, with a minority from Caucasian and Hispanic backgrounds.

Individuals. Darren was a 10-year-old, African American male in classroom A. Andrew was a 9-year-old, Hispanic male in classroom B. Referrals regarding these two individuals initiated the author's contact with classrooms A and B. Original referrals for services were initiated by the parents/guardians of these individuals, who then provided written consent for treatment. These individuals had been identified through a multidisciplinary team evaluation as in need of Special Education services under a verification of Other Health Impairments. Both individuals were diagnosed with Attention Deficit Hyperactivity Disorder (ADHD). Darren was prescribed Wellbutrin and Adderall to control symptoms of ADHD. Andrew was prescribed Tomaxitine to control symptoms of ADHD.

Setting

The GBG was implemented in three regular education classrooms in an urban school district in the Midwest. The GBG was conducted during each class' daily reading period, which lasted approximately one hour. When asked, each teacher had identified this period as the period of primary concern for initiating an intervention. Typical activities during this period included teacher reading, small and large group reading assignments, individual seatwork, and class lessons, discussions, and assignments. Generally, students were seated at their desks, arranged in rows, except during small group work.

Dependent Variables and Measures

Selection and definition of target behaviors. Behavioral observation procedures were used to collect data. Target behaviors were chosen and operationalized with the assistance of the classroom teachers. The teachers were asked to report all behaviors of concern in their classroom and then asked to prioritize these behaviors. The teachers were then asked to select the disruptive behaviors that, if reduced, would have the most significant and positive impact on the learning environment.

Overall, the teachers were remarkably consistent in their reports. Target behaviors selected by each teacher included *out of seat without permission*, *talking without permission*, and *failure to follow directions the first time given*. The author and teachers then collaboratively defined the behaviors, with examples and nonexamples of each initially selected and then further clarified as questions arose.

In each classroom, *out of seat without permission* meant that no part of the student's body was touching his/her desk or chair, and the teacher had not given permission for leaving the desk. Examples included leaving one's desk to go sharpen a pencil, leaving one's desk to give an item to a friend, and leaving one's desk to get a drink or tissue, without the teacher's explicit permission. Nonexamples included leaving one's desk to put an assignment in the assignment box, as implicit permission was given for this behavior in each class.

Talking without permission was defined, in each classroom, to mean vocalization of a word(s) without the teacher's explicit permission. Examples included blurting out an answer without raising one's hand and being called on and nonacademic conversation

with a peer. Nonexamples included speaking when classroom, choral responding was expected or when engaging in small group work with other students and vocalizations other than word(s) (e.g., grunting, humming).

Failure to follow directions the first time given was defined, in each classroom, as failing to follow the teacher's verbal instruction before he/she repeated the instruction. Examples included the teacher giving a verbal command, such as "Get out your book", and the student failing to comply the first time asked. Nonexamples included failure to follow written directions (e.g., existing rules posted, directions on a worksheet).

In addition to the above three target behaviors, the teacher in classroom B selected two additional target behaviors. These target behaviors included *throwing objects* and *speaking impolitely. Throwing objects* was defined as sending any object through the air by a rapid motion of the arm. Examples included throwing school supplies or other objects at the teacher or a peer and throwing an object in the trash from a distance greater than 1 foot. Nonexamples included throwing something in the trash from a distance of 1 foot or less.

Speaking impolitely was defined to mean cursing or referring in an uncomplimentary way to someone's mother or father. Examples included cursing or using phrases beginning with "Your mama..." or "Your daddy..." Nonexamples included arguing with the teacher or a peer and using a harsh tone of voice.

Recording target behaviors. Some researchers have utilized a partial interval recording system to collect baseline and progress monitoring data when implementing the GBG (e.g., Barrish et al., 1969; Harris & Sherman, 1973). In this study, however, an

event recording system was utilized to record the frequency of target behaviors, similar to Salend et al. (1989). Because teachers used event recording in implementing the GBG, the teachers' daily data was then able to be utilized to evaluate the effectiveness of the game for the classroom as a whole. Event recording was also used to collect data on the frequency of the target behaviors for Darren and Andrew, producing data that is comparable to the class data.

Interobserver agreement. In addition to recording the frequency of target behaviors, interobserver agreement was measured. The establishment of interobserver agreement is important for demonstrating that the target behaviors are adequately operationalized and for providing support for the ability of others to replicate the procedures and results. As previously discussed, one of the disadvantages of the GBG is that the teacher has to be alert to the behaviors of students in addition to teaching (Barrish et al., 1969). As a result, if calculated between the teacher and an independent observer, interobserver agreement would likely be low. The independent observer would likely have the opportunity to witness a greater number of occurrences of the target behavior, as he/she would not have to simultaneously attend to teaching. The differences in observations would not be reflective of inadequate operationalizations of the target behaviors but of differing opportunities for observation. Because of these facts, the current study measured interobserver agreement in a manner similar to Salend et al. (1989) in that agreement between two, independent observers was calculated, rather than agreement between the teacher and an independent observer, as the majority of researchers have done. Like Salend et al., the observers used an event recording system to

document the frequency of target behaviors.

Treatment integrity. Finally, because the intervention was implemented in an indirect manner, treatment integrity was monitored. Assessment of treatment integrity is essential for establishing a functional relationship between the GBG and any behavior change (Gresham, 1989; Gresham & Gansle, 1993). Additionally, as noted, treatment integrity has been found to be moderately and positively, though not perfectly, related to an intervention's effectiveness (Elliot et al., 2002; Gresham & Gansle, 1993). Finally, information regarding correct implementation of an intervention is needed before proper and informed decisions regarding the need for any modifications to treatment procedures can be made (Elliot et al., 2002).

In the current study, two methods were used to assess treatment integrity. First, means of obtaining information on treatment integrity recommended by several researchers include using component checklists and direct observation (Elliot et al., 2002; Gresham, 1989; Gresham & Gansle, 1993; Jones et al., 1997; Wickstrom et al., 1998); therefore, the current study utilized a checklist of the GBG components to be implemented by the classroom teacher on a daily basis (refer to Appendix A). The components included reminding the class of the game and of the rules (verbally and/or visually), reminding the class of the criterion, tallying rule infractions on the chalkboard, totaling the tallies at the end of the period, and awarding the winning team(s). An independent observer completed the checklist. Second, the number of days that the intervention was implemented, compared to the number of days implementation was expected, was examined, by reviewing the number of days the teacher recorded

intervention data, in order to provide supplementary information on treatment integrity.

Design

Because differing amounts of time were required to establish baseline levels of target behaviors in each classroom, and because the GBG was begun at different times in the classrooms, it was possible to use a multiple baseline design across classrooms (and, thus, teachers) to examine the efficacy of the GBG in reducing target behaviors below baseline levels. A multiple baseline design is useful when reversal is not desired and/or practical, as is often the case in an applied setting (Cooper et al., 1987). A multiple baseline design involves "the sequential application of the independent variable across technically different [settings]" (Cooper et al., 1987, p. 197). Specifically, after an initial baseline phase, the intervention is introduced in one setting, while baseline conditions continue in the remaining settings. It is predicted that, if the intervention were not introduced, baseline trends in behavior would continue. The extended baseline conditions in the remaining settings allow verification of this prediction. Any change in behavior trends after introduction of the intervention is then attributed to the function of the intervention. As the intervention is then sequentially introduced in the remaining settings, replication of the intervention's effect is then achieved.

Overall, the use of a multiple baseline design increases the confidence that any reduction in target behaviors is the result of the GBG (Patrick et al., 1998). If disruptive behavior is reduced only at the point of intervention in a setting, without a concurrent reduction in settings where the intervention has yet to be introduced, internal validity is demonstrated.

In addition to a multiple baseline design, a changing criterion design was employed, in which the criterion, or goal, for the frequency of target behaviors was periodically changed as success was demonstrated. With a changing criterion design, each criterion phase serves as a baseline for the following phase (Cooper et al., 1987). As the behavior conforms to changes in the criterion, replication of the effects of the intervention and experimental control is demonstrated.

Procedures

Baseline data. As noted, the GBG was implemented during each class' daily reading period. As described, target behaviors were selected with the teachers' assistance. Each teacher then collected baseline data on the frequency of occurrence of the target behaviors for their class as a whole. Using an event recording system, the teacher tallied each instance of a target behavior. The class was not informed that the teacher was recording these behaviors. During the baseline phase, the teacher maintained current instructional and behavioral management procedures. The author and a fellow researcher collected baseline data on Darren and Andrew, using the same system.

Baseline collection continued in classrooms A and B until a reliable trend could be established for the classroom, as a whole, *and* for Darren and Andrew. In classroom C, a decreasing trend was noted; however, due to the teacher's concerns and eagerness to begin an intervention, a clinical decision was made to discontinue the baseline phase after 3 days, despite the decreasing trend.

Preparing for the GBG. For classrooms A and B, once a stable or increasing trend in target behaviors was noted during baseline, for the class as a whole and for the targeted

individuals, the GBG was initiated. For classroom C, as noted, a clinical decision was made to initiate the GBG, despite a decreasing trend in the baseline data.

In preparing to initiate the GBG in each class, classroom rules were developed based on the target behaviors and were posted in each classroom. For the current study, the rules were worded positively (e.g., Fishbein & Wasik, 1981; Grandy et al., 1973). For example, for the target behavior *out of seat without permission*, the rule was "stay in your seat, unless you have permission to get up." Positive wording of the rules was chosen in an attempt to maintain a positive intervention approach and a focus on desired behavior (e.g., McGinnis et al., 1995). Additionally, Evertson (1987) expressed concern that negative wording might "...convey negative expectations, discourage student responsibility, and lead to teacher-student conflicts that defeat the goals of a positive, productive class" (p. 61).

In addition to developing rules, preparation included selecting an initial criterion for the number of rule infractions allowed. For classrooms A and C, the initial criterion for the number of rule infractions allowed to win the game was established by determining the mean number of instances of the target behaviors for the class as a whole during the baseline phase and reducing that number by 10%. It was believed that this initial criterion would be an important but achievable reduction in disruptive behaviors. Cooper et al. (1987) recommended an initial criterion at or slightly below the baseline mean when using differential reinforcement of low rates of behavior.

Because of extremely high rates of target behaviors recorded during the baseline phase in classroom B, however, it was believed that a 10% reduction in target behaviors

would not be meaningful or challenging in this classroom. Instead, the initial criterion for this classroom was established by using the same criterion as in classroom A. At this point, this criterion had demonstrated success in classroom A.

After rules were developed and initial criterion levels established, each class was divided into two teams, with the teacher's assistance. Although conflicting evidence exists regarding the necessity of teams (Grandy et al., 1973; Harris & Sherman, 1973), a decision was made to retain this original component of the GBG. Even if the component were to be proven unnecessary, no evidence exists to suggest its use would adversely impact the results.

In dividing each class into teams, an attempt was made to ensure that the students who exhibited the highest levels of disruptive behavior were distributed evenly between the two teams (e.g., Saigh & Umar, 1983). Additionally, consideration was given to facilitating ease in data recording. For example, if possible, students were divided by rows, so the teacher could easily identify to which team a student belonged.

Finally, a grab bag reward system was prepared (e.g., Salend et al., 1989). A grab bag reward system was chosen to attempt to ensure that at least some rewards would prove reinforcing for each individual student and to attempt to avoid satiation on a single reward. The students and teachers assisted in determining the rewards for their classroom, in order to attempt to insure the selection of rewards that were truly reinforcing for students and feasible for teachers to provide. With the assistance of the author, the students and teachers brainstormed possible rewards. The students and teachers had to agree on rewards to be included in the grab bag system. Rewards were then written on

cards and placed in a bag.

Natural reinforcers, other than teacher attention, were used (e.g., Barrish et al., 1969). Examples included extra recess time, fewer homework problems, reading to the Kindergarteners, and eating lunch in the room with the teacher. Each time a team won, the teacher or a student from the team randomly selected a reward from the grab bag for the whole team to receive.

Implementation of the GBG. On the first day of implementation of the GBG, the author introduced the game to the class and modeled the procedures. No other study was reviewed in which a consultant modeled the procedures for the teacher. It was hypothesized that initially modeling the game would promote treatment integrity by the teacher. Sterling-Turner et al. (2001) found that direct training, such as modeling, in the implementation of an intervention procedure resulted in greater treatment integrity than the provision of written or verbal descriptions alone.

Following the initial session, the teacher and/or assistants implemented the game on a daily basis. Teachers were provided with a general description of the steps of the GBG and a list of procedures to be completed on a daily basis (refer to Appendixes B and C). Daily procedures to be implemented by the teacher included reminding the class of the game, rules, and criterion, tallying rule infractions on the chalkboard according to teams, summing the tally marks at the end of the period, and rewarding the winning team(s). The teacher also recorded the total number of infractions for the class on a calendar, for data keeping purposes. Teams competed only against the criterion, not each other. A team could win if the criterion was not exceeded. Both teams could win on a

given day.

In addition to the daily data collected by the teacher, the author and a fellow researcher attempted to collect weekly progress monitoring data on Darren and Andrew. Interobserver agreement was checked approximately weekly, as well. Suspensions, detentions, field trips, and scheduling conflicts resulted in missing data for some weeks. The author and fellow researcher independently observed and recorded rule infractions, using event recording, during one reading period in each class. Behavior definitions were reviewed and questions addressed periodically.

Finally, the author monitored treatment integrity, using the component checklist. At least weekly monitoring of treatment integrity is recommended (Elliot et al., 2002). The current study assessed treatment integrity, using the component checklist, twice each week until 100% integrity was demonstrated. It was believed that by initially monitoring treatment integrity twice each week, problems could be remedied more quickly than if weekly monitoring was used, increasing the proportion of time that the intervention was implemented as intended and, thus, increasing confidence in the study's results. Following demonstration of 100% integrity, weekly monitoring was conducted to attempt to ensure that treatment integrity was maintained. In addition, the number of days that the GBG was implemented, compared to the number of days implementation was expected, was monitored at least weekly, by reviewing the number of days the teacher recorded intervention data.

When integrity was less than 100%, feedback was provided, as recommended by Jones et al. (1997). In general, the teacher was verbally reminded of the steps he/she had

forgotten. Additionally, the author modeled the game a second time in classroom C on day 10 of the intervention phase because of treatment integrity concerns.

The data collected on the GBG was closely monitored and changes in the intervention made, as appropriate. First, the data was examined approximately every two weeks to determine if a change in the criterion level was warranted. If the class had met the criterion on a majority of the days the GBG was implemented *and* the teacher believed that a change in the criterion was appropriate, the criterion was lowered.

Using these procedures, the criterion was only lowered in class A. It was lowered from 30 to 20 to 18 and, finally, to 16. Although review of the data for classroom B indicated the class had met the criterion on a majority of the days, the teacher was not yet willing to lower the criterion. In classroom C, concerns with treatment integrity resulted in a decision not to change the criterion. The insurance of proper implementation was desired before changes were made.

In addition to considering a criterion change, if a reduction in disruptive behaviors was achieved and maintained during reading class, consideration was given to expansion of the GBG to another subject(s), with approval of the teacher. The same procedures were applied during the additional period(s). Again, only classroom A expanded the GBG to another subject, the science period. Classroom B's teacher did not wish to expand the game to another subject until after the winter break, at which time the author's services were ended. In classroom C, concerns over treatment integrity again resulted in a decision not to expand the intervention.

Analysis

In congruence with past research (e.g., Patrick et al., 1998; Salend et al., 1989), methods of analysis included visual analysis and comparison of means and ranges. All values reported were rounded to the nearest whole number. Visual analysis included examination of trends and changes in levels of the target behaviors. Comparisons were made between the ranges and mean frequencies of target behaviors during baseline and treatment phases for each classroom as a whole and for the two individuals of interest. The percentage of change in the mean frequency of target behaviors from baseline to treatment phases is reported, as well. Also reported is the number of days that the criterion for the number of rule infractions allowed was met.

In addition to the frequency of target behaviors, interobserver agreement and treatment integrity were examined. The percentage interobserver agreement was calculated by dividing the smaller frequency by the larger frequency and multiplying the result by 100 (Salend et al., 1989). The range and overall mean percentage agreement is reported. Treatment integrity is reported as the percentage of components, on the component checklist, adhered to. The range and overall mean percentage of treatment integrity is reported. Additionally, the number of days that the intervention was implemented, compared to the number of days implementation was expected, is reported as a percentage.

Results

Interobserver Agreement

Classrooms A, B, and C. When the number of rule infractions was recorded for

classroom A, as a whole, interobserver agreement ranged from 81% - 100%, with a mean of 91% (refer to table 1). Because Thanksgiving break interfered with the scheduled observation, data on interobserver agreement was unable to be collected for week 4 of the intervention.

For classroom B, interobserver agreement ranged from 54% - 100%, with a mean of 84% (refer to table 1). Examination of the data reveals that 54% agreement is an outlier, however. If this data point is excluded from analysis, percentage agreement for classroom B ranged from 78% - 100%, with a mean of 91%. It is noted that data on interobserver agreement was unable to be collected on week 4 of the intervention in this classroom, however, because of a field trip during the normal reading period during which observations were scheduled and Thanksgiving break making rescheduling of the observation difficult.

Finally, for classroom C, interobserver agreement ranged from 77% - 94%, with a mean of 88% (refer to table 1). Data on interobserver agreement was unable to be collected on week 4 and week 7 of the intervention, however, because of special activities resulting in reading class not occurring on the scheduled observation days. Furthermore, holiday breaks made rescheduling of these two observations difficult.

Individuals. When the number of rule infractions was recorded for Darren, interobserver agreement ranged from 73% - 100%, with a mean of 87% (refer to table 1). Data on interobserver agreement for Darren is unavailable for week 2 of the intervention phase because of a detention on the scheduled observation day.

For Andrew, interobserver agreement ranged from 90% - 100%, with a mean of

98% (refer to table 1). No data on interobserver agreement for Andrew is available for weeks 3 and 4 of the intervention phase because of a suspension, a field trip, and detentions leaving few opportunities to arrange an observation.

Overall Results

Overall, in classrooms A and B, a clear decrease below baseline levels in the number of rule infractions was achieved when the GBG was implemented. The lower frequency of rule infractions continued over the course of 7 weeks and 5 weeks, respectively. With regard to treatment integrity, 100% compliance with treatment components was established and generally maintained throughout the remainder of the intervention. In classroom C, a slight decrease below baseline levels in the number of rule infractions was achieved when the GBG was implemented. With regard to treatment integrity, 100% compliance with treatment components was established but not maintained throughout the remainder of the intervention. The lower frequency of rule infractions continued over the course of 7 weeks in this classroom.

With regard to Darren and Andrew, a fairly small amount of data was obtained, because of unforeseen events. The data obtained revealed a decrease, upon implementation of the GBG, below baseline levels in the number of rule infractions exhibited by each individual. The reduced frequency of rule infractions was maintained for Andrew over the course of the intervention. For Darren, a return to baseline rates of behavior was observed during week 2 of the intervention.

Frequency of Rule Infractions for Each Classroom

Classroom A. Overall, classroom A exhibited decreasing numbers of rule

infractions in reading and science class during the period that the GBG was played. Visual inspection of the data (refer to figure 1) reveals an increasing trend in the number of target behaviors exhibited in reading class during the baseline phase. The mean number of target behaviors witnessed by the teacher during the baseline phase was 34 (range = 22 – 47).

In contrast to the baseline phase, when the GBG was implemented during the reading period, a decreasing trend was observed in the number of rule infractions exhibited. The class decreased the number of rule infractions to a mean of 13 (range = 0 – 30). Thus, when compared to the baseline mean, a 62% decrease in the number of rule infractions was achieved.

Overall, the number of rule infractions exhibited conformed to changes in the established criterion for "winning" the game. The goal for the number of rule infractions was initially set at 30 (15 per team), 10% lower than the baseline mean. It was later dropped to 20, 18, and finally 16. The number of rule infractions for the class, as a whole, exceeded the established criterion on only two occasions. On each of these occasions, one team lost the game.

When the total number of rule infractions, for both teams, is examined, the class met the goal for the number of rule infractions allowed on at least 31 out of 34 days that the game was played. The teacher forgot to record the data on day 16 of the intervention; it is uncertain whether the goal was achieved on this day. When the number of rule infractions is examined separately for each team, both teams won the game on 28 out of the 33 days for which data is available. Of the remaining 5 days, one team lost on each

day.

On day 15 of implementation of the GBG during the reading period, the game was expanded to science class. Overall, the mean number of rule infractions during science, while implementing the GBG, was 20 (range = 9 – 42). Initial levels of rule infractions during science class exceeded current levels in reading class, where the GBG had been in effect for 15 periods. The number of rule infractions exhibited during science quickly exhibited a decreasing trend, however, and conformed fairly closely to the number of infractions occurring during reading class. Nonetheless, the number of rule infractions exhibited by the final day was higher in science class than in reading class.

In further examining the data for the science period, the class, as a whole, met the goal for the number of rule infractions allowed on 9 of the 16 days the game was played. Both teams won on 8 days, only one team won on 7 days, and both teams lost on only 1 day.

Classroom B. Overall, classroom B exhibited decreasing numbers of rule infractions in reading class during the period that the GBG was played. Visual inspection of the data (refer to figure 2) reveals a stable trend but high levels of target behaviors during the baseline phase. The mean number of target behaviors witnessed by the teacher during the baseline phase was 213 (range = 131 – 318).

In contrast to the baseline phase, when the GBG was implemented during the reading period, a sharp and rapid decrease in the level of rule infractions was achieved. The number of rule infractions then exhibited a stable trend for the remainder of time the GBG was implemented. Overall, the class decreased the number of rule infractions to a

mean of 29 (range = 8 – 69). Thus, when compared to the baseline mean, an 86% decrease in the number of rule infractions was achieved. Additionally, the trend in the number of rule infractions exhibited conformed closely to the criterion of 30 (15 per team).

Further examination reveals that, during the reading period, the class, as a whole, met the goal for the number of rule infractions allowed on 16 of the 23 days the game was played. When the number of rule infractions is examined separately for each team, both teams won on 11 of the 23 days. Of the remaining 12 days, one team lost on 7 days and both teams lost on 5 days. The criterion level was never lowered and the GBG was not expanded into other periods of the day in classroom B.

Classroom C. The reduction in the number of rule infractions exhibited by classroom C was not as large as in classrooms A and B. Visual inspection of the data (refer to figure 3) reveals a decreasing trend in the numbers of target behaviors witnessed by the teacher during the baseline phase. The mean number of target behaviors witnessed by the teacher during the baseline phase was 14 (range = 12 – 16).

Upon beginning the GBG, an immediate decrease in the level of target behaviors occurred. On the first day of implementation, only 4 rule infractions were noted. In general, a stable trend of approximately 9 rule infractions was observed for the remainder of the time the GBG was in effect. Overall, the class decreased the number of rule infractions to a mean of 11 (range = 0 - 18). Thus, when compared to the baseline mean, a 21% decrease in the number of rule infractions was achieved.

The initial criterion for the number of rule infractions allowed in classroom C was

set at 12 (6 per team), 10% lower than the baseline mean. The class, as a whole, met the goal for the number of rule infractions allowed on 18 of the 21 days the game was played. When the number of rule infractions is examined separately for each team, both teams won on 17 of the 21 days. Of the remaining 4 days, one team lost on 3 days and both teams lost on 1 day. The criterion was never lowered, and the GBG was not expanded into other periods of the day in classroom C.

Frequency of Rule Infractions for Individuals

Darren. Very little data is available for Darren because this student was transferred to another classroom two weeks after implementation of the GBG. Visual inspection of the data (refer to figure 4) reveals an increasing trend in the numbers of target behaviors witnessed by an independent observer during the baseline phase. The mean number of target behaviors witnessed during the baseline phase was 45 (range = 30 – 57).

Upon implementation of the GBG, an immediate decrease in the level of target behaviors occurred. On day 3 that the game was played, an independent observer recorded only 6 rule infractions for Darren. On day 10, the independent observer recorded 57 rule infractions for Darren, a return to baseline level.

Andrew. As with Darren, only a small amount of data is available for Andrew. Opportunities for observation were limited by detentions, suspensions, and absences. Specifically, no data is available during weeks 3 and 4 of implementation of the GBG because of a suspension, a field trip, and detentions leaving few opportunities to arrange an observation.

Visual inspection of the data (refer to figure 5) reveals a stable trend in the numbers of target behaviors witnessed by an independent observer during the baseline phase. The mean number of target behaviors witnessed during the baseline phase was 14 (range = 10 –19). Upon implementation of the GBG, an immediate decrease in the level of target behaviors was achieved, and this lower level was maintained through week 5. Overall, Andrew decreased the number of rule infractions engaged in to a mean of 1 (range = 0 - 2), representing a 99% decrease below the baseline mean.

Treatment Integrity

Classroom A. In classroom A, examination of the direct observation data indicates that the mean percentage of treatment components adhered to was 100% (range = 100% - 100%) (refer to table 2). Additionally, the GBG was implemented during reading class on 100% of the days expected. The teacher forgot, however, to record the data for the author for one of these days; therefore, data on the number of rule infractions for day 16 of the intervention phase is missing (refer to figure 1). The students in classroom A verified that the GBG was conducted on this day.

With regard to science class, the GBG was implemented on 84% of the days expected. The GBG was not implemented on three days. The teacher's explanation for why the game was not implemented on these days was that she forgot. Additionally, it is noted that implementation on day 34 of the intervention phase was not expected, as science class did not occur on this day due to a special end of semester celebration.

Classroom B. In classroom B, the mean percentage of treatment components adhered to was 89% (range = 63% - 100%) (refer to table 2). In this classroom, 100%

compliance with the checklist components was achieved by the second observation period. Data on treatment integrity was unable to be collected on week 4 of the intervention in this classroom, however, because of a field trip during the normal reading period. The GBG was implemented later in the day on this occasion. The author was unable to be present to observe at this time.

In examining the data, the GBG was implemented in classroom B on 100% of the days expected. Examination of figure 2 reveals missing data for day 21 of the intervention phase; however, a field trip on this day resulted in reading class being canceled. Because the reading period did not occur, implementation of the GBG on this day was not expected.

Classroom C. In classroom C, the mean percentage of treatment components adhered to was 82% (range = 63% - 100%) (refer to table 2). Data on treatment integrity was unable to be collected on week 4 and week 7 of the intervention in this classroom, however, because of special activities resulting in reading class not occurring on the scheduled observation day. In this classroom, 100% compliance with the checklist components was achieved by the second observation period; however, this level of compliance was not maintained.

As previously noted, when treatment integrity was less than 100%, the teacher was verbally reminded of the components forgotten. Additionally, the GBG was modeled a second time on day 10 of the intervention phase in this classroom.

The GBG was implemented on only 62% of the days expected in this classroom. The GBG was not implemented on 8 of the 29 days expected. When asked why the game

was not implemented on these days, the teacher explained that he simply forgot. Additionally, examination of figure 3 reveals missing data for days 9, 15, and 27. On day 15, a class field trip resulted in reading class not being held. On days 9 and 27, reading class was also canceled due to guest speakers coming into the classroom. Because reading class was not held on days 9, 15, and 27, implementation of the GBG was not expected.

Discussion

Overall Conclusions

The purpose of the current study was to examine the effectiveness of the GBG in reducing disruptive behavior rates for classrooms, as a whole, and for individuals reported to be engaging in relatively high rates of disruptive behavior, compared to classroom peers. Review of the data reveals that the GBG was demonstrated to be an effective intervention technique for classroom and individual behavior management. Upon implementation of the GBG, rates of disruptive behavior were immediately reduced in each classroom and for each individual. The effects of the GBG were maintained across as many as 7 weeks for each classroom and across 5 weeks for Andrew.

Support for the Effectiveness of the GBG

Classrooms A, B, and C. Examination of the data indicates that the GBG was effective in producing reduced rates of disruptive behavior in each classroom. Upon implementation of the GBG in each classroom, disruptive behaviors were immediately reduced below baseline levels.

Overall, rates of disruptive behavior were reduced by 21%-86%. The differences

in the degrees of reduction in each classroom can be partially explained by the initial levels of disruptive behavior in each room. The largest reduction (86%) occurred in the classroom with the highest baseline rates of disruptive behavior and vice versa. Although variability in disruptive behavior rates is evident during the intervention phase, lower levels of disruptive behavior, compared to baseline, were generally maintained throughout. Exceeding the criterion rate was rare in each classroom.

The use of a multiple baseline design strengthens conclusions regarding a functional relationship between the GBG and reduced behavior rates (Cooper et al., 1987). During the reading period in classrooms A and B, reductions in disruptive behavior occurred at the point of implementation of the GBG. Although a decreasing trend in target behavior rates was evident during the baseline phase in classroom C, support for a functional relationship is derived from the immediate reduction in the level of target behaviors, upon implementation of the GBG.

The use of a changing criterion design in classroom A also provides support for a functional relationship between the GBG and reduced behavior rates. In both reading and science class, disruptive behavior rates generally conformed to the four criterion levels. Thus, the effects of the GBG were replicated and experimental control demonstrated (Cooper et al., 1987).

Anecdotal reports from staff members in each classroom indicated overall satisfaction with the GBG and the resulting reductions in disruptive behavior. Each teacher expressed a desire to continue to expand the GBG. Instructions were left for continuing the intervention after the author's services were complete. Instructions

included reviewing the data every two weeks. If disruptive behavior rates were at or below the criterion level on a majority of the days, then a decision could be made to either lower the criterion or expand the GBG to another subject. The criterion was not to be lowered below 8 (4 per team), however, in order to maintain reasonable and achievable expectations.

Darren and Andrew. The GBG was also demonstrated to be effective in managing individual behavior. For Andrew, an immediate reduction in disruptive behavior rates was achieved and maintained with implementation of the GBG. A 99% reduction in target behaviors for Andrew represents a highly significant and meaningful change.

For Darren, although a return to baseline levels of disruptive behavior was witnessed on day 10 of the intervention phase, support for the GBG's effectiveness is still evident. First, after an increasing trend in target behaviors was witnessed during the baseline phase, an immediate 87% reduction in target behaviors was witnessed upon implementation of the GBG. Additionally, the data for day 10 do not appear to accurately represent Darren's typical behavior following implementation of the GBG. Day 10 possessed some unique characteristics. First, on this day, Darren had been informed that he was soon to be transferred to another classroom. Additionally, on this day, Darren's team had exceeded the criterion for winning the game. Although Darren engaged in high numbers of rule infractions throughout the period, after the criterion was exceeded, a dramatic increase in the number of rule infractions was witnessed. Darren commented that his team had already lost, so it did not matter anymore.

Information from anecdotal teacher reports and classroom data support the

conclusion that the data for Darren on day 10 are not representative and that the GBG was effective in reducing disruptive behaviors for Andrew and Darren. Darren's teacher reported that Darren's behavior on day 10 was atypical, compared to what had been witnessed on previous days when the GBG was in effect. The number of rule infractions witnessed on day 3 of the intervention phase was more representative, according to the teacher. Overall, the teacher reported that Darren's levels of disruptive behavior had improved during reading after implementation of the GBG. Additionally, examination of the classroom data reveals that the class exceeded the criterion for the number of rule infractions allowed on only two days; therefore, Darren, as witnessed by the teacher, must have met the criterion himself on at least all but 2 days that the game was played. As reported, when Darren was observed to exceed the criterion, his team also exceeded the criterion.

Andrew's teacher reported that the data collected for Andrew was representative of his typical behavior. Additionally, examination of the classroom data reveals that the class met the criterion for the number of rule infractions allowed on 16 of the 23 days the game was played; therefore, Andrew, as observed by the teacher, must have met the criterion on these days, as well.

Thus, the individual and classroom data, in addition to anecdotal teacher reports indicate that the GBG can be an effective behavioral intervention technique for individuals with a history of high rates of disruptive behavior. Additionally, teachers reported satisfaction with the effects of the GBG.

Treatment Integrity

Any time an intervention is implemented indirectly, in the context of a consultative relationship, the potential for imperfect adherence to procedures always exists; therefore, treatment integrity is important to assess if conclusions are to be made regarding a functional relationship between the intervention and any changes in behavior (Gresham, 1989; Gresham & Gansle, 1993). Additionally, a moderate and positive relationship is hypothesized to exist between treatment integrity and the effectiveness of an intervention (e.g., Gresham, 1989; Gresham & Gansle, 1993). It is unknown, however, how much a treatment protocol can be violated before the intervention becomes ineffective (Gresham, 1989; Jones et al., 1997). The current study provides some support for a positive, though imperfect, relationship between treatment integrity and the effectiveness of an intervention.

The data regarding implementation of the intervention components and the data regarding the percentage of expected days that the GBG was implemented indicate acceptable levels of treatment integrity in classrooms A and B, but not in classroom C. Overall, less reduction in disruptive behavior rates was achieved in classroom C, in which treatment integrity was poor, than was achieved in classrooms A and B. Treatment integrity was not perfect in classroom B and in classroom A's science class, however. Despite this fact, the data strongly support the effectiveness of the GBG in these two classrooms. Additionally, although lower levels of reductions were achieved, disruptive behaviors in classroom C were reduced. Thus, again, an important and positive relationship between treatment integrity and effectiveness appears to exist, although the

relationship is not perfect.

Implications for School Psychology

General implications. Overall, use of the GBG appears appropriate and feasible in the routine practice of school psychology, given certain considerations. The current study demonstrated the effectiveness of the GBG in reducing disruptive behavior rates. The intervention proved to be low-cost and easy to implement. It was also an easy technique for teachers to understand. Limited knowledge of behavioral principles is required.

With regard to time requirements, the GBG was generally a time-efficient intervention for teachers to implement, as well. Recording of behaviors required more time in classroom B, in which higher rates of disruptive behaviors were evident. The teacher reported difficulty with the recording during the baseline phase; however, after rates were reduced with implementation of the GBG, the recording became more manageable.

In contrast to the time-efficiency of the GBG for teachers, the procedures used in the current application of the GBG proved time consuming for the author. As a practicum student at the time, the author's caseload was lower than that of a full-time employed school psychologist. As a result, interobserver agreement and treatment integrity measures were able to be collected frequently. For a full-time school psychologist wishing to implement the GBG, collection of interobserver agreement data might not be feasible or necessary. It is recommended that treatment integrity be monitored; however, less frequent checks might prove necessary. Overall, however, little consultation time was required to implement the GBG. Consultation time was similar to any case in which

behavior problems must be defined, analyzed, treated, and monitored.

Specific implications for classroom and individual management. The GBG might be a good intervention choice if several students within a classroom are exhibiting similar behaviors of concern. In such cases, a single, classroom intervention serves as a time and resource efficient method for addressing multiple intervention targets simultaneously. Additionally, as demonstrated in the current study, the GBG can be effective in managing the behavior of individual students with a history of high rates of disruptive behaviors. Use of a classroom management technique rather than an individual intervention addresses the potential impact of the classroom context on the individual's behavior. Some issues must be considered before selecting the GBG as an intervention technique, however.

First, school psychologists should consider the fit between the GBG and existing classroom practices (Detrich, 1999). One potential reason for the poor treatment integrity and less dramatic results achieved in classroom C is that the intervention differed from existing classroom practices, in which the teacher's primary method of addressing disruptions was to ignore them. Although the intervention is easy and time-efficient to implement, discipline and commitment, by both the teacher and school psychologist, to implementation of the procedures is necessary. It is believed that the GBG will prove a good fit in the majority of classrooms, however. In the author's experience, most teachers naturally attend to and provide consequences for disruptive behaviors, in some fashion.

An additional consideration, particularly in selecting the GBG as an intervention strategy for an individual with a history of high rates of disruptive behavior, is the

potential for negative peer pressure or punishment of students exhibiting the highest rates of behaviors (Patick et al., 1998; Salend et al., 1989). Such negative peer pressure and punishment was witnessed on a few occasions in the current application of the GBG. Students would verbally reprimand the student exhibiting the highest rates of disruptive behaviors or complain about that student being on their team, blaming him/her for loss of rewards. Teachers instructed students to refrain from such punishment and reminded them that the GBG was a team effort and that they should encourage each other.

Similarly, the issue of how to address the students accumulating the large numbers of demerits for their team arose in classrooms B and C (e.g., Barrish et al., 1969; Harris & Sherman, 1973; Medland & Stachnik, 1972). On day 9 of implementation of the GBG in each classroom, the suggestion was made, in accord with procedures used by Barrish et al., that students accumulating more than four demerits in a period might have to be removed from their respective teams. Additionally, different from the procedures used by Barrish et al., it was suggested that such students would have to "earn" their way back onto the team, and thus to being eligible for rewards, by accumulating no more than four demerits during the period. The threat of these procedures, alone, was apparently sufficient in controlling behaviors because the teachers never had to implement the procedures.

As a result of these observations, school psychologists are urged to develop plans to address such concerns and to implement such plans at the initiation of the GBG. One potential preventative technique for addressing these concerns is the implementation of the "tie" component, discussed previously (Harris & Sherman, 1973). The author was

unaware of this technique prior to beginning the GBG in the current study. Until further empirical examination of methods to address the concerns is conducted, however, it is recommended that school psychologists select the method that appears best suited to the classroom in which they are intervening.

Another consideration in selecting the GBG as an intervention strategy is the fact that intervention effects might not generalize or be maintained long-term. Although not directly assessed in the current study, anecdotal teacher reports indicated that, in general, the GBG's effects did not generalize to other periods during which the GBG was not being implemented. Furthermore, although Darren and Andrew's behavior improved during their daily reading periods, both students continued to receive detentions and suspensions for behaviors occurring during periods in which the GBG was not in effect. As a result of continued behavior concerns, Darren's parents ultimately requested he be transferred to another classroom before the GBG could be expanded to additional periods. Also as a result of continued behavior concerns, the multidisciplinary team was considering transferring Andrew to a yet another new school, although it is the author's understanding that this decision has yet to be made.

The continued behavior concerns for Darren and Andrew illustrate the time and patience required when implementing the GBG. Gradual and direct expansion is necessary for managing behaviors across class periods. Darren and Andrew's teachers expressed beliefs that these students' behaviors could be managed with continued expansion of the GBG. Overall, continued empirical examination of methods for promoting generalization is needed, however.

Finally, although reduced disruptive behavior rates were maintained over the course of the current study, no long-term maintenance data was gathered. It is uncertain if effects would be maintained over a longer period of time or upon cessation of the GBG.

Limitations

Limitations related to the applied nature of the study. Although possessing several strengths, the current study also includes some limitations that deserve discussion. A majority of the limitations derive from the fact that the study reports on the clinical application of the GBG. Conducting research on the GBG was not the initial goal. Providing excellent psychological services to the teachers and students was the goal. As a result, decisions were made and some procedures used that would not have likely been the case had the primary intent been to conduct research.

For example, a clinical decision was made to begin the GBG in classroom C, despite a decreasing trend in the baseline data, because the teacher remained concerned with the level of disruptive behavior and was eager to begin an intervention and because two independent observers were able to witness much higher numbers of disruptive behaviors. Although it is impossible to know with certainty whether the decreasing baseline trend would have continued and achieved reductions in disruptive behaviors would have occurred without implementation of the GBG, it is again noted that a rapid decrease in the level of target behaviors was achieved upon implementation of the GBG and that behavior rates continued to generally conform to the criterion level.

Another limitation of the current study is that measurement of baseline rates of target behaviors in classroom A's science class did not occur. It is impossible to know

whether reductions in disruptive behavior would have occurred without direct implementation of the GBG. According to anecdotal reports from the teacher, target behavior rates were not decreasing in science class prior to implementation of the GBG. Furthermore, as previously noted, rates of disruptive behavior in science class upon initial implementation of the GBG were higher than the current rates in reading class, suggesting that generalization of the GBG's effects to the science period did not occur until the GBG was directly implemented in that period. Finally, despite the lack of baseline information, further support for a functional relationship between the GBG and disruptive behavior rates in science comes from the fact that disruptive behavior rates in science generally conformed closely to the changes in criterion levels.

A third limitation resulting from the applied nature of the study is the small amount of data available for Darren and Andrew. Suspensions, detentions, and absences made scheduling of observations difficult. As previously noted, however, anecdotal teacher reports and the classroom data provide confirmation of the conclusion that the GBG was effective in reducing disruptive behaviors for these students.

Another limitation involves the fact that planned expansions to additional subjects and changes in criterion levels were unable to be implemented in classrooms B and C. Although classroom B won the game on a majority of the days the GBG was implemented, the criterion level was never changed and expansion to other subjects did not occur because the teacher was not yet willing to implement these changes. She was afraid of rushing things and wished to wait until after the winter break to implement any changes. The author's services were complete upon the beginning of winter break,

however. The author had persuaded the teacher to try lowering the criterion to 20 (10 per team) on day 24 of the intervention phase; however, the teacher was absent that week, and the substitute did not wish to implement the game. Expansion and criterion changes were not initiated for classroom C because of treatment integrity concerns in this classroom.

The concerns regarding treatment integrity in classroom C represent another limitation to the current study. In contrast to classrooms A and B, the goal of establishing and maintaining 100% adherence to the intervention components was not achieved for classroom C, despite feedback provision to the teacher and modeling of the GBG a second time on day 10 of the intervention phase. Although 100% adherence was witnessed on two occasions, including the first observation subsequent to the second modeling of the procedures, the remainder of observations indicated less than perfect adherence. Tallying demerits and rewarding the winning team(s) were the only components adhered to during 100% of the observations in this classroom. Each of the other components was implemented inconsistently. Of particular concern is the fact that, on more than one occasion, the teacher would verbally note target behaviors but fail to record them on the chalkboard. Additionally, as reported, the GBG was only implemented on 62% of the expected days in this classroom. It is possible that modeling the procedures additional times would have improved the maintenance of treatment integrity in this classroom.

It is uncertain why the teacher in classroom C did not implement the GBG as intended. Gresham (1989) listed six factors related to treatment integrity, including the

motivation of the treatment agent and the complexity, time requirements, material and resource requirements, required number of treatment agents, and effectiveness (perceived and actual) of the intervention. Anecdotally, the teacher reported satisfaction with the GBG and its effects, even calling the author and her fellow researcher his "fairy godmothers". He also expressed a desire to continue the GBG beyond the time of the author's services, and, as in the other classrooms, was given instructions for doing so. Additionally, the GBG is generally perceived as easy to implement (Harris & Sherman, 1973) and required little time, human, or material resources to implement in the current classroom. The intervention differed, however, from existing classroom practices, in which the teacher's primary method of addressing disruptions was to ignore them. Detrich (1999) stated that one variable related to treatment integrity is the fit of an intervention to existing classroom practices. As previously noted, it is possible that this factor was related to the poor treatment integrity, and poor treatment integrity is hypothesized to be the reason for the less dramatic effects of the GBG in this classroom.

An additional issue related to treatment integrity involves the potential for inflated estimates in the data. Again, because the current study involved a clinical application of the GBG, modeling and feedback were used in an attempt to promote treatment integrity (Jones et al., 1997; Sterling-Turner et al., 2001). Thus, a pure measure of integrity, without such promotion, was not obtained. As a result, treatment integrity estimates may be inflated. It is uncertain with what level of integrity teachers would implement the GBG if such training and feedback were not provided. This issue is one that requires additional empirical evaluation.

Another fact that deserves mention is that, as mentioned by Barrish et al. (1969), teachers were not able to witness all instances of disruptive behavior. This fact is evident in the individual versus classroom data. In examining the data, higher numbers of disruptive behaviors are often reported for Darren than are reported for his classroom as a whole on a particular day. This fact is the result of the independent observer being able to attend to one individual and observe more instances of disruptive behavior than the teacher, who had to simultaneously teach the lesson and attend to the behavior of multiple students. Despite the teachers' inability to witness all behaviors, the GBG proved effective, a characteristic that makes the GBG particularly useful as a classroom management strategy.

Finally, it is possible that the values for interobserver agreement in the current study are slightly inflated due to the fact that behavior definitions were reviewed periodically. Specifically, review of the definitions could have resulted in increased observer drift. Observer drift involves "...the tendency of observers to change the manner in which they apply the definitions of behavior over time" (Kazdin, 1977, p. 143). Observers can "drift" together, such that they apply different definitions over time yet maintain high levels of agreement (Lipinski & Nelson, 1974). Agreement with additional observers would be lower, however. Because the procedures for reviewing definitions in the current study were not systematic, it is impossible to empirically examine the effects in this case. This fact represents a limitation in the current study. Future researchers should address this issue systematically. Again, had the primary intent of the current study been to conduct research, procedures for collection of interobserver agreement data

would likely have been more systematic.

Additional limitations. An additional potential limitation to the current study is the possibility of expectancy effects impacting the data (e.g., Kazdin, 1977; Shuller & McNamara, 1976). An expectancy effect is "...the effect of prior knowledge upon the observation process" (Shuller & McNamara, 1976, p. 519). The teachers and author were aware of the students whose behaviors were of concern and were knowledgeable of the expected outcomes of the intervention. This fact may have resulted in the teachers and author observing what they expected. Specifically, higher rates of disruptive behaviors were expected during the baseline phases and lower rates after implementation of the GBG. Additionally, Kazdin noted that expectancy effects are likely when feedback is provided, including feedback in the form of observers being exposed to the data as it is being collected. The teachers and author were aware of trends in the data as they collected it. Shuller and McNamara (1976) empirically addressed the issue of potential expectancy effects, however, and reported different results. Trained observers were given different descriptions or labels of subjects and then asked to record six behaviors and to report their subjective impressions of subjects. The researchers found that the descriptions affected the observers' subjective impressions (expectations) of subjects but not the accuracy of the behavior recordings.

Finally, another limitation, always to be considered in research, is the issue of external validity. The current study was conducted on a limited sample. Results and conclusions must be cautiously applied to other students, teachers, grades, settings, and geographical locations.

Suggestions for Future Implementation and Research

Several suggestions are offered for future implementation of and research on the GBG. First, more data is needed on the effectiveness of the GBG in reducing disruptive behaviors for individuals. More frequent assessment is recommended. Teachers could be requested to record data for the individuals of concern, in addition to classroom data, on a daily basis. In the current application, the author wished to keep demands on the teacher as low as possible. Little additional effort would be required, however, if only one or two individuals were monitored. Individual data could be gathered by asking the teacher to separately tally individual rule infractions, perhaps by inconspicuously recording tallies on a piece of paper at his/her desk or on a piece of masking tape affixed to his/her hand. By avoiding distinguishing marks on the chalkboard, this method would avoid singling out the individual and potentially further encouraging negative peer pressure or punishment of the individual.

Second, more data on the generalization and long-term maintenance of effects is needed. Direct assessment of generalization is recommended by directly recording behavior rates during periods in which the GBG is not implemented. Efforts to develop and empirically examine methods of promoting generalization are also encouraged.

Third, the current study did not assess the relationship between reductions in disruptive behavior and academic performance. It is recommended that future researchers continue to examine the relationship between classroom management of behaviors and academic engagement, academic responding, and academic achievement. In particular, the potential for academic engagement and academic responding to serve as mediators

between classroom management and academic achievement (Starkweather-Lund & Shriver, 2003) should continue to be investigated, in general and in the context of the GBG.

With regard to treatment integrity, it is recommended that researchers and clinicians directly assess treatment integrity when implementing the GBG (Wickstrom et al., 1998). It appears that there was at least a modest and positive relationship between treatment integrity and effectiveness of the GBG in the current study. Further, empirical examination is warranted, however. Additionally, continued empirical examination of the effects of procedures designed to promote treatment integrity (e.g., modeling and feedback) is encouraged.

Another recommendation for future researchers is that two independent observers be used to assess interobserver agreement, as in the current study and the study by Salend et al. (1989). Again, the teachers were not able to witness as many instances of disruptive behaviors as the independent observers because of the simultaneous need to teach; thus, differing opportunities for observation existed.

Finally, issues related to students accumulating large numbers of demerits, and punishment of such students, continue to require examination. As noted, it is recommended that clinicians and researchers implementing the GBG develop plans to address such concerns and implement such plans at the initiation of the GBG. Similarly, it is also recommended that clinicians and researchers implementing the GBG address the potential problem of rates of disruptive behavior rapidly accelerating after the criterion has been exceeded. Specifically, the further examination of preventative efforts to

address these two, related concerns is recommended. In particular, further empirical examination of Harris & Sherman's (1973) "tie" component is recommended.

Summary

In summary, the current study provides support for the effectiveness and feasibility of the GBG in routine school psychology practice. Reductions in rates of disruptive behaviors were witnessed in three elementary classrooms. Additionally, modest evidence was obtained suggesting the effectiveness of the GBG in reducing disruptive behavior rates for individuals referred for behavior concerns.

The GBG was an easy, low-cost, time-efficient intervention for the classrooms. Little consultation time was required. Teachers reported overall satisfaction with the GBG and its effects. All three teachers reported desires to continue the GBG. Overall, if several students within a classroom need to be targeted for similar behavior concerns, consideration of the GBG is recommended. Issues requiring further empirical examination, however, keep the author from promoting the GBG as a "behavioral vaccine" (Embry, 2002) at this point.

References

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.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied Behavior Analysis.* Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Detrich, R. (1999). Increasing treatment fidelity by matching interventions to contextual variables within the educational setting. *School Psychology Review, 28* (4), 608-620.

Dolan, L. J., Kellam, S. G., Brown, C. H., Werthamer-Larsson, L., Rebok, G. W., Mayer, L. S., Laudolff, J., Turkkan, J. S., Ford, C., & Wheeler, L. (1993). The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology, 14* (3), 317-345.

Elliot, S. N., Witt, J. C., Kratochwill, T. R., & Stoiber, K. C. (2002). Selecting and evaluating classroom interventions. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 243-286). Bethesda, MD: National Association of School Psychologists.

Embry, D. D. (2002). The good behavior game: A best practice candidate as a universal behavioral vaccine. *Clinical Child and Family Psychology Review, 5* (4), 273-297.

Evertson, C. M. (1987). Managing classrooms: A framework for teachers. In D. C. Berliner & B. V. Rosenshine (Eds.), *Talks to teachers* (pp. 54-74). New York: Random House.

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.

Grandy, G. S., Madsen, C. H., Jr., & De Mersseman, L. M. (1973). The effects of individual and interdependent contingencies on inappropriate classroom behavior. *Psychology in the Schools, 10* (4), 488-493.

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18* (1), 37-50.

Gresham, F. M., & Gansle, K. A. (1993). Treatment integrity of school-based behavioral intervention studies: 1980-1990. *School Psychology Review, 22* (2), 254-272.

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.

Jones, K. M., Wickstrom, K. F., & Friman, P. C. (1997). The effects of observational feedback on treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 12* (4), 316-326.

Kazdin, A. E. (1977). Artifact, bias, and complexity of assessment: The ABCs of reliability. *Journal of Applied Behavior Analysis, 10* (1), 141-150.

Lipinski, D. U., & Nelson, R. (1974). Problems in the use of naturalistic observation as a means of behavioral assessment. *Behavior Therapy, 5* (3), 341-351.

McGinnis, J. C., Frederick, B. P., & Edwards, R. (1995). Enhancing classroom management through proactive rules and procedures. *Psychology in the Schools, 32* (3), 220-224.

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

Patrick, C. A., Ward, P., & Crouch, D. W. (1998). Effects of holding students accountable for social behaviors during volleyball games in elementary physical education. *Journal of Teaching in Physical Education, 17* (2), 143-156.

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

Salend, S. J., Reynolds, C. J., & Coyle, E. M. (1989). Individualizing the good behavior game across type and frequency of behavior with emotionally disturbed adolescents. *Behavior Modification, 13* (1), 108-126.

Shuller, D. Y., & McNamara, J. R. (1976). Expectancy factors in behavioral observation. *Behavior Therapy, 7* (4), 519-527.

Starkweather-Lund, A. R., & Shriver, M. D. (2003). *Training teachers to give effective commands: Effects on student compliance, academic engagement, and academic responding.* Manuscript in preparation, University of Nebraska Medical Center.

Sterling-Turner, H. E., Watson, T. S., Wildmon, M., Watkins, C., & Little, E. (2001).

Investigating the relationship between training type and treatment integrity. *School Psychology Quarterly, 16* (1), 56-67.

Tingstrom, D. H. (1994). The good behavior game: An investigation of teachers' acceptance. *Psychology in the Schools, 31,* 57-65.

Wickstrom, K. F., Jones, K. M., LaFleur, L. H., & Witt, J. C. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13* (2), 141-154.

Appendix A

Component Checklist for Assessing Treatment Integrity

TREATMENT INTEGRITY DATA **Date of Observation:**

Classroom:

Remind Class of Game and Classroom Rules

Opportunity	*Adherence (Y or N)*

Remind Class of Criterion

Opportunity	*Adherence (Y or N)*

Remind Class of Reward

Opportunity	*Adherence (Y or N)*

Mark Demerits on Board During Class

Opportunity	*Adherence (Y or N)*

Tally Demerits at End of Instructional Period

Opportunity	*Adherence (Y or N)*

Award Winning Team/Teams

Opportunity	*Adherence (Y or N)*

Appendix B

General Description of Steps of the GBG Provided to Teachers

Steps to The Good Behavior Game:

1. Establish classroom rules for behavior and explain the rules to the class.
2. Divide the classroom into two teams.
3. Establish a goal for the number of rule infractions allowed (I will assist you based on the data collected so far).
4. Make a tally mark on the chalkboard, under the appropriate team name, each time any member of a team breaks a rule.
5. At the end of the period, add up the tally marks for each team.
6. Reward the team(s) that does not exceed the number of rule infractions allowed (Be thinking of potential rewards that would be feasible in your classroom and that could be provided to one or both teams, depending on who wins on a given day).

Appendix C

List of Daily Procedures for Implementing the GBG Provided to Teachers

The Good Behavior Game: Steps to Be Completed on a <u>DAILY</u> Basis:

1. Remind the class of the game and rules.
2. Remind the class of the goal.
3. Remind the class of the rewards.
4. Tally rule infractions on the board.
5. Add up the tallies at the end of the period.
6. Reward the winning team(s)!

Table 1

Percent Interobserver Agreement

Week	Class A	Class B	Class C	Darren	Andrew
Baseline	91	54	94	73	90
1	94	78	94	100	100
2	90	89	87		100
3	81	97	89		--
4	--	--	--		--
5	90	100	87		100
6	90		77		
7	100		--		

Note. Dashes indicate that data was not obtained for that week.

Table 2

Percent of Treatment Components Adhered To

Week	Class A	Class B	Class C
1	100	63	63
1	100	100	100
2	100	100	63
3	100	100	100
4	100	--	--
5	100	83	83
6	100		83
7	100		--

Note. Dashes indicate that data was not obtained for that week.

Figure Captions

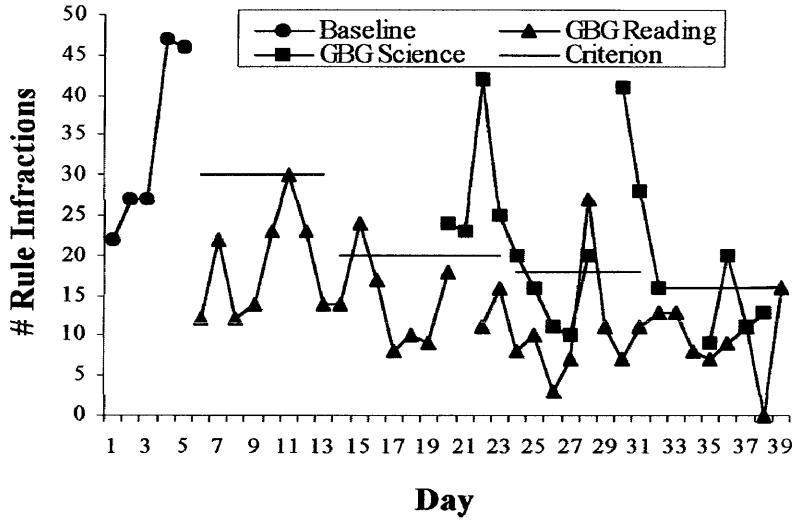
Figure 1. Number of rule infractions during baseline and intervention phases for classroom A.

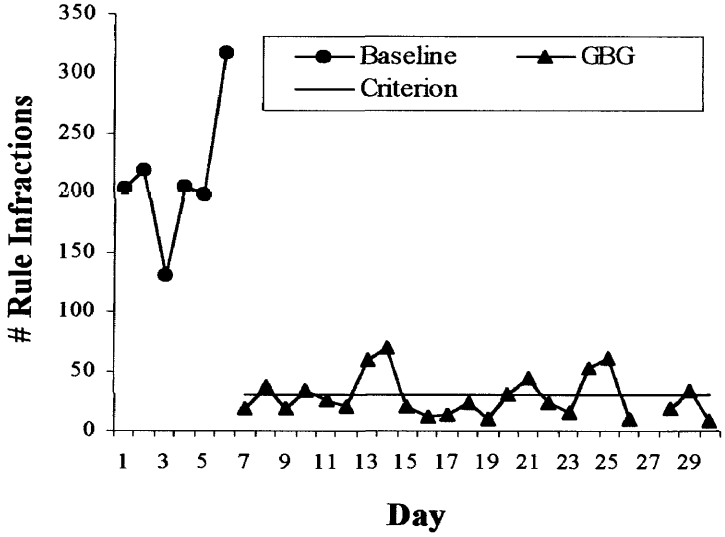
Figure 2. Number of rule infractions during baseline and intervention phases for classroom B.

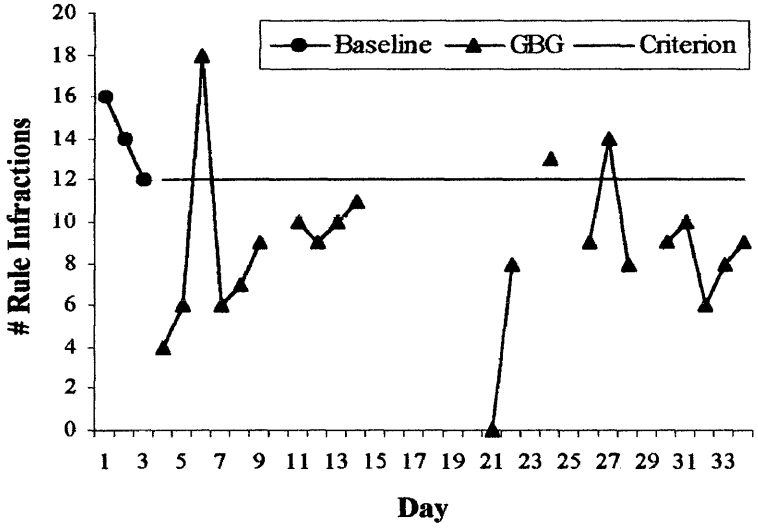
Figure 3. Number of rule infractions during baseline and intervention phases for classroom C.

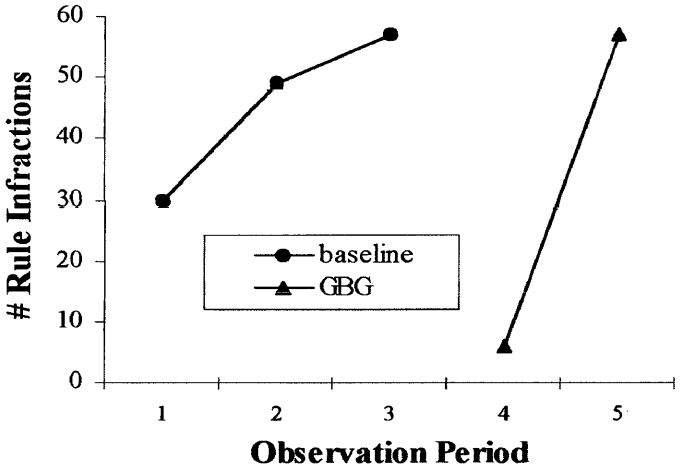
Figure 4. Number of rule infractions during baseline and intervention phases for Darren.

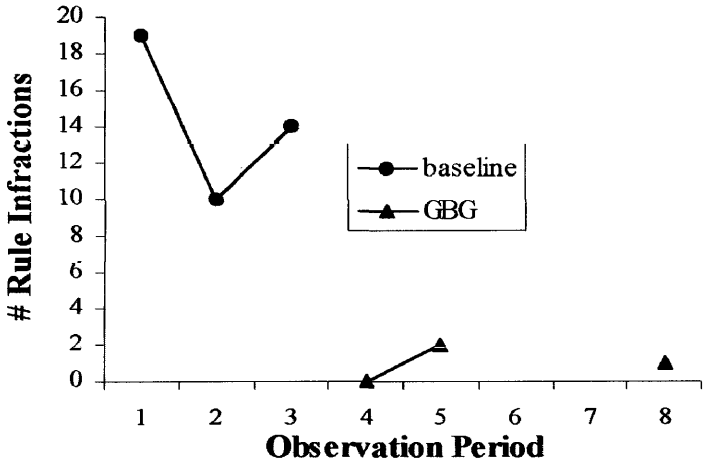
Figure 5. Number of rule infractions during baseline and intervention phases for Andrew.


Baseline
GBG Reading
GBG Science
Criterion
Rule Infractions
Day
0
5
10
15
20
25
30
35
40
45
50
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39


350
300
250
200
150
100
50
0
Rule Infractions
Baseline
GBG
Criterion
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29
Day


Baseline
GBG
Criterion
Rule Infractions
Day
0
2
4
6
8
10
12
14
16
18
20
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33


60
50
40
30
20
10
0
Rule Infractions
baseline
GBG
1
2
3
4
5
Observation Period


20
18
16
14
12
10
8
6
4
2
0
Rule Infractions
baseline
GBG
1
2
3
4
5
6
7
8
Observation Period