Received: 13 July 2020 | Revised: 6 January 2021 | Accepted: 20 January 2021

DOI: 10.1002/pits.22496

RESEARCH ARTICLE

WILEY

Adapting the good behavior game for special education classrooms

Emily A. Groves | Richard J. May | Rachel E. Rees | Jennifer L. Austin

School of Psychology, University of South Wales, Pontypridd, Wales, UK

Correspondence
Emily A. Groves, School of Psychology, University of South Wales, CF37 1DL, UK.
Email: emily.groves@southwales.ac.uk

Abstract

The good behavior game (GBG) is a classroom management intervention that employs a group contingency to support appropriate behavior and reduce classroom disruptions. To date, the majority of GBG research has included participants of typical development in mainstream education classrooms or alternative schools. The current study evaluated the GBG across two classrooms in a special education school in Wales. Children included those with both intellectual and physical disabilities. Although some adaptations were made to accommodate individual student needs, those adjustments were minimal and required little additional effort from teachers. Results were evaluated using a single-case withdrawal design and indicated that the GBG was effective at reducing disruptive and off-task behavior in both classrooms.

KEYWORDS
classroom management, good behavior game, learning disabilities, special educational needs

Across educational contexts, teachers report disruptive behavior as a primary concern affecting both the quality of the classroom learning environment and job satisfaction (Ofsted, 2014, 2019). Deficiencies in classroom management skills exacerbate these problems (Department for Education, 2013; Derrington, 2008) and highlight the need for efficient and effective strategies that take minimal time and resources to learn and subsequently implement. The good behavior

Rachel E. Rees is now at the Behaviour Clinic, Swansea.

game (GBG; Barrish et al., 1969) is an example of one such strategy. Often employed as a Tier 1 intervention, the GBG has proven itself effective (see Flower et al., 2014b and Tingstorm et al., 2006 for reviews), acceptable (Tingstorm, 1994), and resistant to treatment integrity failures (Joslyn & Vollmer, 2020).

The GBG combines a number of features to support good classroom behavior. Teachers begin by setting clear expectations for behavior, which become the "rules" of the game. Teachers then divide the class into teams and an interdependent group contingency (Litow & Pumroy, 1975) is used to establish shared responsibility for behavior. Rule violations result in the accrual of points for the team, regardless of how many team members broke the rule. To win the game and receive a reward, teams must earn no more than a specified number of points. A variation of the GBG, often referred to as the caught being good game (CBBG; e.g., Wright & McCurdy, 2011), involves awarding points for rule-following and providing rewards to teams who earn at or above a specified criterion. A specific advantage of this version of the game is that it places focus on appropriate behavior rather than problem behavior.

1 | EVIDENCE BASE FOR THE GBG

Although the GBG has garnered over 50 years of research demonstrating its effectiveness across a range of school settings and target behaviors, the majority of this research has been conducted in mainstream education classrooms with typically developing students (Bowman-Perrott et al., 2016; Flower et al., 2014b). There is an emerging literature on applications of the GBG in alternative education settings (Groves & Austin, 2017, 2019; Joslyn et al., 2014, 2019; Rubow et al., 2018; Salend et al., 1989; Sy et al., 2016), but applications in other special education settings are relatively scarce.

The lack of GBG research across a broader spectrum of special education needs (SEN) is unfortunate, given that SEN schools and classrooms face many of the same behavioral challenges as mainstream schools. Despite access to specialist resources and increased staffing, the UK Office for Standards of Education, Children's Services, and Skills (Ofsted, 2005) reported that schools catering to pupils with autistic spectrum disorders (ASD) and intellectual disabilities (ID) tended to regard all their pupils as challenging, with teachers reporting the occurrence of problem behavior on a daily basis. In Wales, recent reports from the Chief Inspector of Education and Training described how pupils in special schools often make limited progress towards their individual behavior targets (Estyn, 2016, 2017, 2018). Concerns regarding the extent of problem behavior within SEN schools have deepened in recent years, with Ofsted (2018) reporting that pupils receiving SEN support are five times more likely to be permanently excluded from school than their mainstream counterparts. Many such exclusions are due not to severe behaviors such as aggression or property destruction, but to low-level, persistent disruptions (Department for Education, 2019; Welsh Government, 2018). These findings highlight the need for additional evidence-based strategies to support both teachers and students in addressing problem behavior. Although individualized interventions might be necessary for more severe behavior, the use of more general, class-wide strategies may help teachers deal with low-level, daily disruptions that interfere with learning.

2 | GBG IN SPECIAL EDUCATION

A notable addition to the GBG evidence base is Breeman et al. (2016), who evaluated the GBG across 11 special education schools over the course of one school year. The schools catered for students with a range of disorders, including ASD, attention deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder. The study employed a cluster randomized controlled design, in which classrooms were assigned to either the GBG intervention group or an education-as-usual control group. The authors used a response cost version of the GBG, in which teams were given a certain number of cards at the outset of the game. During the game, cards were removed from teams contingent upon rule violations. To win the game and receive a reward, teams needed to have at least one card remaining at the end of the game. Given the population, the authors altered games to include fewer class rules than is typical, shorter game lengths, and smaller teams.

Effects of the GBG on students' behavioral and emotional problems were evaluated via the problem behavior at school interview (PBSI; Erasmus Medical Center, 2000), which was completed by teachers at the beginning and end of the school year. Results indicated no significant differences between GBG and control classrooms with respect to PBSI ratings. However, given that significant increases in problem behavior were reported for the control group, it is possible the GBG may have prevented the worsening of problem behavior.

The findings of Breeman et al. (2016) are inconsistent with other studies evaluating the GBG in alternative and special education classrooms (Joslyn et al., 2014, 2019; Rubow et al., 2018; Salend et al., 1989; Sy et al., 2016), which have reported improvements in behavior during GBG phases. However, studies reporting positive effects typically have employed direct observation strategies to measure the impact of the GBG. Although indirect measures like the ones employed in Breeman et al. provide insight into teachers' experience of the GBG, direct observation might provide a more focused evaluation that may be less prone to bias.

3 | POTENTIAL BARRIERS TO GBG IMPLEMENTATION IN SPECIAL EDUCATION

It is possible that the relative lack of GBG research across a broader scope of SENs is due to concerns regarding its adaptability for students with particular types of learning needs. Although the GBG allows teachers the flexibility to set rules and contingencies that are suited to their students' capabilities (Groves & Austin, 2017), some teachers (and researchers) may assume that the complexity of those contingencies will make application of the game difficult for some students (e.g., those with IDs). Particular physical disabilities (e.g., blindness, deafness) also would require further adaptation of procedures and might be viewed as cumbersome. To date, we know of no GBG studies that have specifically targeted classrooms with students requiring support for these types of SENs.

Breeman et al. (2016) reported that only minimal adaptations were required for the GBG to be implemented across a range of SENs. Unfortunately, the study produced equivocal outcomes in terms of behavior improvement. Therefore, it is unknown whether those adaptations were sufficient to produce meaningful behavior changes. Further, given that Breeman et al. reported no participants with ID, it is difficult to discern the types of adaptations that might be needed to support these students.

4 | PURPOSE

The purpose of Study 1 was to pilot implementation of the GBG in a classroom in an SEN school and to determine whether a GBG with modifications similar to those used in Breeman et al. (2016) were sufficient to produce behavior improvements for children with intellectual and physical disabilities. Study 2 sought to replicate and extend the findings of Study 1 by assessing direct effects on behavior in another SEN classroom, as well as assessing the teacher's perceptions of those effects. To promote a stronger focus on positive behavior during the GBG, we employed the CBGG version of the game across both studies.

5 | STUDY 1

5.1 | Method

5.1.1 | Participants and setting

The study took place in a SEN school for children aged 3–19 years with moderate, severe, and profound IDs, specific learning disabilities (e.g., dyslexia, dyspraxia), ASD, and emotional and behavioral disorders (EBD). A class including 8

students aged 14–15, who had a diagnosis of intellectual and/or physical disability, participated in the study. Three children in the class were identified by their teacher as engaging in high rates of problem behavior. Observations of all students in the class were conducted by the researcher, which verified the teacher's nominations. Gethin (aged 15) had a diagnosis of ID and was severely sight impaired. Carys (aged 14) and Dewi (aged 14) both had a diagnosis of ID. All participants engaged in high rates of calling out and off-task behavior during independent work. As these were the deemed the most disruptive pupils, they served as proxies for the behavior of the entire class. Data were collected during literacy and numeracy lessons. Each lesson lasted approximately 1 h and the GBG was played during the portion of the lesson in which students completed individual, teacher-assisted work tasks (usually 40–50 min).

5.1.2 | Response definitions and measurement

Response definitions were developed following discussion with the teacher and direct observation of the children in the classroom. *Call outs* were defined as vocalizations to a teacher without being called upon to speak. *Off-task behavior* was defined as the student's body and gaze not being oriented towards the teacher, whiteboard or materials, and pencil not in hand or near paper (during writing tasks), for at least 3 consecutive seconds. If a student touched or manipulated materials while on task, they were not scored as off-task.

The frequency of call outs was recorded and presented as rate per minute. Off-task behavior was recorded using partial-interval time-sampling (Fiske & Delmolino, 2012) and presented as a percentage of intervals in which behavior occurred. Observations were between 15 and 30 min in duration and divided into 10 s intervals. All participants were observed during each interval and the interval was scored if any of the students engaged in a target behavior. Observations were paused if a target student left the room and resumed once the student returned. Observations did not occur if any of the three target students were absent. Observations were conducted two to three times per week, with no more than two sessions being conducted per day. If two sessions were conducted in the same day, observers allowed at least 2 h between sessions.

5.1.3 | Interobserver agreement (IOA)

A second, independent observer collected data during 24% of observation sessions for the purposes of determining IOA. IOA for call outs was calculated on an exact count per interval basis by dividing the number of intervals with 100% agreement by the total number of intervals and multiplying by 100. IOA for off-task behavior was calculated on an interval-by-interval basis by dividing the number of intervals with agreement by the total number of intervals and multiplying by 100. For call outs, the mean IOA was 91% (range: 74%–100%). For off-task behavior, the mean IOA was 91% (range: 83%–99%).

5.2 | Experimental procedures

5.2.1 | Baseline

During baseline, no programmed intervention was in place and the teacher was asked to respond to students' disruptive behaviors as she typically would. Following baseline, the class teacher attended a 2 h, department-wide training on the GBG, which was led by the first author. The training was attended by six class teachers and 12 teaching assistants. The training included step-by-step instructions on how to implement the GBG, including instruction and video-modeling of GBG implementation. The trainer met with the teacher participating in the study to discuss how the game would be implemented in her class and the modifications that might be necessary for her

students. These included providing pictures to correspond with written rules, providing individualized copies of the rules (including rules in Braille), using a proximal visual display of each team's points, and comprising teams of no more than three students.

5.2.2 | Good behavior game

The teacher divided the class into small teams of 2–3 students. Each team sat together at a table and included one target student. Team composition remained the same throughout the study. At the outset of the first GBG session, the teacher told the students that they would be playing a game in which they had the opportunity to earn soccer "goals" for following classroom rules and that they could earn a surprise reward at the end of the lesson if they achieved a certain number of goals (rewards were revealed at the end of the game and included such things as small edibles, 5 min of tablet time, or extra break time). The teacher then explained the three class rules to the students using examples and nonexamples. The rules included *stay on task*, *request attention appropriately*, and *be kind to others*. The rules and corresponding pictures were printed on a poster and displayed at the front of the class. Each student also was provided with a copy of the rules to place on their table. Gethin's rules were written in Braille. The teacher used a soccer theme for awarding points and referred to students as "players" during the game. The teacher used a timer to remind her to check in with teams every 4 min and determine if the team had earned a goal. If all members of the team had followed the rules for the preceding interval, the teacher placed a cutout of a laminated soccer ball on a paper soccer net that was attached to each team's table and provided praise. If a team had not followed the rules, the teacher did not provide a soccer ball and reminded the students of the rules.

At the start of each session, the teacher reminded the students of the GBG rules by referring to the pictures on the poster and announced how many goals were required to win the game. This criterion for winning varied depending on the length of the game (a higher criterion was used in longer games), but was usually set at between 60% and 90% of the total number of goals available during that session. For example, for a 40-min session in which 10 goals were available, the criterion was between 6 and 9 goals. The goal criterion never exceeded the total number of goals available during that session. Approximately 5 min before the end of the lesson, the teacher announced that the game was over and calculated the number of goals that each team had achieved and revealed the surprise reward. Teams that had achieved the criterion for that session received that game's designated reward.

5.3 | Experimental design and data analysis

A single-case, withdrawal design was used to evaluate the effects of the GBG on the dependent variables. A withdrawal design involves repeated introduction and withdrawal of an independent variable to demonstrate experimental control (Ledford & Gast, 2018). Specifically, an ABA design was employed, in which 'A' denoted a baseline phase with no intervention and "B" denoted the intervention phase. Phase changes occurred once stable responding had been observed for both dependent variables. The experimental design was originally intended to include a second intervention phase (i.e., ABAB); however, due to the school's reorganization of classrooms and teachers, data collection ended before the implementation of a second treatment phase.

Visual and quantitative analyses were conducted to determine effects of the GBG on call outs and off-task behavior. Changes in trend, level, and stability across baseline and intervention sessions were assessed via visual analysis (Kratochwill et al., 2010). Means were calculated for each measure across conditions and were subsequently used to describe overall levels of problem behavior as low, moderate, or high. Low level was defined as less than 30%, moderate level was defined as 30%–60%, and high level was defined as more than 60%. In addition, percentage nonoverlapping data (PND) and Tau-U effect size calculations were conducted. PND (Scruggs

et al., 1987) is a calculation of the percentage of data points in phase B (i.e., intervention phase) exceeding the single highest phase A (i.e., baseline) data point. When targeting behaviors for reduction, the calculation for PND is reversed, such that the percentage of data points in phase B below the single lowest phase A data point is calculated. When interpreting PND scores, a score of more than 70% suggests an effective intervention, 50%–70% suggests questionable effectiveness, and less than 50% suggests no observed effect (Parker et al., 2011a). Tau-U (Parker et al., 2011b) provides an index of nonoverlap between phases combined with trend from within the intervention phases. Tau-U scores were calculated for each phase comparison using an online calculator (available at www.singlecaseresearch.org). When interpreting Tau-U scores, scores of 0–.65 suggests weak effects, scores of .66–.92 suggest medium effects, and scores of .93–1.0 suggest large or strong effects (Parker & Vannest, 2009).

5.4 | Treatment integrity

During all intervention sessions, treatment integrity was recorded via a 10-item checklist. The checklist included a list of all steps that were necessary to implement the game (e.g., scoreboard visible to all students, teacher provided points for rule following, teacher provided rewards to winning teams). Treatment integrity averaged 93% (range: 80%–100%). On the occasions when treatment integrity was not 100%, errors included the teacher failing to remind students of the rules before starting the game and not delivering the reward immediately after the end of the game. A second independent observer recorded treatment integrity data during 33% of sessions. Agreement for treatment integrity data was calculated by dividing the number of checklist agreements by the total number of checklist items and multiplying by 100. Agreement for treatment integrity was 100%.

5.5 | Results and discussion

The results of Study 1 are displayed in Figure 1. The top panel displays the rate of calling out behavior. During baseline, the target students engaged in an average of 1.58 call outs per minute (range: 0.6–1.58). The GBG was introduced in Session 6 and call outs reduced to an average of 0.16 responses per min (range: 0.05–0.4). The GBG was then withdrawn in Session 12 and the rate of call outs initially remained low (Sessions 12–14). During the last three sessions of the withdrawal phase, rate of calls out increased to similar levels observed in the initial baseline ($M = 0.6$; range: 0.1–1.8).

Percentage of intervals with off-task behavior is displayed in the bottom panel of Figure 1. During baseline, students engaged in moderate but variable levels of off-task behavior ($M = 46\%$; range: 25%–60%). Upon introduction of the GBG, off-task behavior reduced substantially, to an average of 4% during the GBG phase (range: 1%–7%). A reduction in variability was also observed during GBG sessions, such that levels of responding remained consistently low during the GBG phase. The GBG was withdrawn in Session 12 and, similarly to call outs, off-task behavior initially remained low and gradually increased towards to latter end of the phase ($M = 18\%$; range: 2%–55%). The results of the PND and Tau-U effect size calculations suggest that the GBG intervention was very effective. Comparisons between the initial baseline and GBG condition suggest a large effect size in terms of decreasing call outs ($PND = 100\%$; $Tau\text{-}U = -1$) and off-task behavior ($PND = 100\%$; $Tau\text{-}U = -1$). Effect size calculations did not, however, consider the withdrawal to baseline phase due to the lack of a second treatment phase.

The results of Study 1 suggest that the GBG was effective at reducing call outs and off-task behavior in a classroom for adolescents with IDs and physical disabilities. Rate of call outs and percentage of intervals with off-task behavior decreased substantially and immediately following the introduction of the GBG. When the GBG was removed, problem behavior gradually increased to preintervention levels. Although the analysis was conducted in only one SEN classroom, the data suggest that the GBG is a viable intervention for students with IDs. Like Breeman et al. (2016), we made only minor modifications to the game to address the students' cognitive and

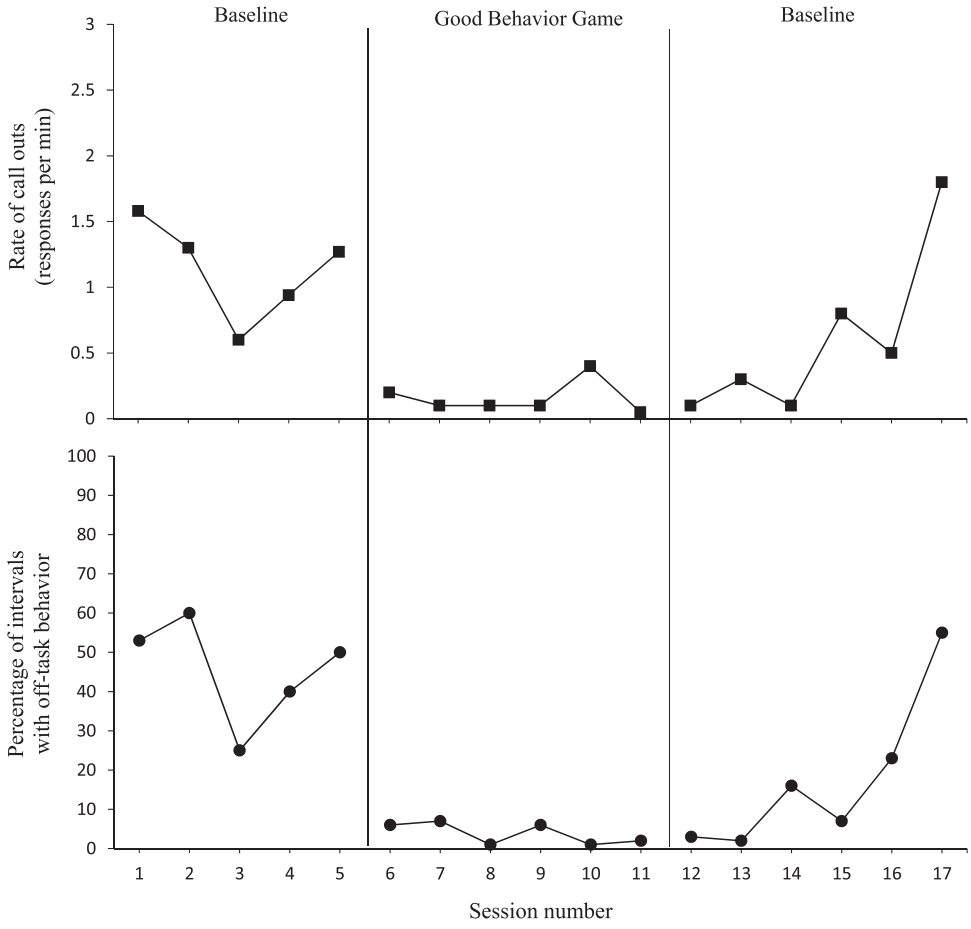

FIGURE 1 Percentage of Intervals with Problem Behavior across Students. The top panel displays intervals with calling out and the bottom panel displays intervals with off-task behavior across baseline and GBG phases

physical disabilities. Specifically, we ensured that written rules were augmented with pictures and that the rules appeared as a proximal stimulus prompt on each student's desk (in Braille, where needed). We made the concept of points less abstract by using a soccer theme in which students earned goals on a net, which was a concept familiar to all the students. As interdependent contingencies with large groups can make it more difficult to discern why a criterion might be met or not met, we used small team sizes to make teams more easily discernible to the students and to increase the likelihood they would understand why their team had earned (or missed) a goal. Unlike Breeman et al., we did not use short game durations. The teacher opted to use game lengths that matched the length of individual work tasks, so that rewards were delivered near the end of the session and did not disrupt the flow of the lesson.

Despite encouraging outcomes, the premature end to the study meant that the effects of the GBG could not be replicated in a second intervention phase, which would have strengthened experimental control. Additionally, data were collected on all target students during every interval (i.e., if any student engaged in off-task behavior during an interval, the interval was scored). Although this type of whole-group measure is common when analyzing the behavior of groups of students, recording individual student behavior would provide a more sensitive measure of the effects of the IV.

The purpose of Study 2 was to replicate and extend the findings of Study 1. An ABAB withdrawal design was used to enhance experimental control and data were collected on individual participants' disruptive and off-task behavior. We used fewer modifications to procedures to determine whether a more "standard" GBG would produce positive effects on behavior. Additionally, Study 2 sought to examine the social validity (Wolf, 1978) of the GBG procedures by assessing teacher's perceptions of behavior change and acceptability of the intervention.

6 | STUDY 2

6.1 | Method

6.1.1 | Participants and setting

Study 2 was conducted in the same school as Study 1, in a class comprised of eight students aged 12–15 who had a diagnosis of intellectual and/or physical disability. Harri (aged 14), Evan (aged 14), and Ieuan (aged 15) were identified by their teacher as engaging in higher rates of problem behavior (e.g., leaving their seats without permission, off-task behavior) than their classmates. Classroom observations by the researcher verified the teacher's nominations. All three students had a diagnosis of ID. Data were collected during 1 h literacy lessons, and the GBG was played during the portion of the lesson in which students completed individual teacher-assisted work tasks (usually 40–60 min).

6.1.2 | Response definitions and measurement

Target behaviors were selected based on discussions with the class teacher and observations in the classroom. *Inappropriate sitting* included being out of seat, sitting on legs/knees, and/or swinging on the chair. *Off-task behavior* included the student's body or gaze not being oriented towards the teacher, whiteboard, or work materials, and pencil not in hand or near paper (during writing tasks) for at least 3 s consecutively. If student touched or manipulated materials whilst on-task, this was not scored as off-task. Data were recorded on all dependent variables for each student.

Partial-interval time-sampling was used to collect data on target behaviors and data were presented as the percentage of interval in which the target behavior occurred. Observations lasted 12–24 min (depending on the duration of individual work time) and were divided into 10 s intervals. One student was observed per interval and the observer rotated across students. Observations were conducted three to four times per week, with no more than two sessions being conducted per day. If two sessions were conducted in the same day, observers allowed at least 2 h between sessions.

6.1.3 | Interobserver agreement

A second, independent observer was present during 30% of sessions. IOA was calculated on an interval-by-interval basis, by dividing the number of intervals with agreement by the total number of intervals and multiplying by 100. For inappropriate sitting, the mean IOA was 94% (range: 84%–100%). For off-task behavior, the mean IOA was 92% (range: 80%–99%).

6.2 | Experimental procedures

6.2.1 | Baseline

The teacher was asked to respond to problem behaviors as she typically would. Following baseline, the teacher and a teacher's assistant attended a 1 h 30 min training with the first author that was identical in format to the training conducted in Study 1. The teacher selected the game rules during the training, with consideration of the problem behaviors that the teacher had identified before baseline. The teacher also asked to include three additional rules that had not been targeted initially. The rules of the game were *stay on task*, *stay in your seat and sit safely*, *use materials appropriately*, *hands up to talk*, and *keep noise down*. The teacher and researcher agreed that the entire class would serve as a single team.

6.2.2 | Good behavior game

The teacher explained that the class would be playing a game, in which everyone had to work together to follow the rules and win the game. The written rules and corresponding pictures were displayed on a poster at the front of the class. The scoreboard was depicted as a race-track with a start and finish line, which appeared on the electronic whiteboard. The teacher explained that she would check on the team several times during the game and would move the race car forward if everyone on the team had been following the rules. She further explained that to win the game and earn a reward, the car had to cross the finish line before the game ended.

Games lasted either 45 or 60 min, depending on the length of the lesson. The number of spaces on the racetrack varied as a function of lesson duration. In a 45 min game, the race-track consisted of 15 spaces. In a 60 min game, it had 20 spaces. The criterion for winning was set at 80% of the total points available (e.g., 12 and 16 points for the 45 and 60 min games, respectively). Rewards included 5 min of tablet time or extra break time. The teacher used a timer to remind her to assess behavior every 3 min. If all students had been following the rules for the preceding interval, the teacher used an electronic clicker to move the car forward and provided brief praise. If the team had not followed the rules, the teacher did not move the car and reminded the students of the rules. As an incentive for the class to continue adhering to the classroom rules once they had met or surpassed the finish line, the teacher awarded each student a bonus "Dojo" point for each space the car moved past the finish line. Dojo points were accumulated throughout the week and were exchanged for a reward on a Friday afternoon.

6.3 | Experimental design and data analysis

As in Study 1, a single-case withdrawal design was used to evaluate the effects of the GBG on the dependent variables. Phase changes occurred once a stable pattern of responding had been observed for both dependent variables. The data analysis strategy was identical to the one employed in Study 1 and employed both visual analysis and calculation of effect sizes using PND and Tau U.

6.4 | Treatment integrity

Treatment integrity data were collected during all GBG sessions using a checklist identical to the one used in Study 1. Treatment integrity averaged 97% (range: 90%–100%). When treatment integrity was not 100%, errors included not reminding the students of the rules before starting the game and failing to provide feedback to students when a point was not earned.

6.5 | Social validity

At the end of the study, the teacher completed a researcher-developed 22-item questionnaire, which included a 5-point Likert scale. A higher score indicated greater agreement with the statement. The questionnaire assessed the extent to which the teacher was involved in designing the intervention, if she enjoyed implementing the procedures, the ease/difficulty in implementing the procedures, and the impact of the procedures on the behavior in the classroom. The teacher was also asked if she would continue to implement the procedures in the future and if she would recommend the GBG to other teachers.

All students in the class (including nontarget students) completed an 8-item questionnaire during brief, individual meetings with a researcher. Response options included yes, no, and maybe. The researcher read each question and the answer options during the meeting, and subsequently recorded the student's response. Questions on the students' questionnaire asked if they enjoyed playing the GBG, if they felt they behaved better during the game, if they found the rules easy to understand, and if they would like to continue playing the game in the future.

6.6 | Results and discussion

Figure 2 displays the percentage of intervals with inappropriate sitting and off-task behavior for Harri. During the pretreatment baseline, Harri engaged in variable levels of inappropriate sitting, averaging 21% (range: 3%–44%), and moderate levels of off-task behavior (M = 51%; range: 33%–73%). Upon implementation of the GBG, Harri's inappropriate sitting immediately reduced to zero or near zero levels (M = 2%; range: 0%–10%), whilst off-task behavior also reduced substantially, averaging 5% (range: 0%–16%). The GBG was withdrawn in session 13 and an immediate increase was observed in inappropriate sitting (M = 20%; range: 3%–30%) and off-task behavior (M = 41%; range: 19%–50%). When the GBG was reintroduced during the final phase of the study, inappropriate sitting reduced to zero in all sessions, whilst off-task behavior reduced to zero or near zero (M = 1%; range: 0%–3%). Harri was absent during session 10.

Figure 3 displays the percentage of intervals with inappropriate sitting and off-task behavior for Evan. During the pretreatment baseline, Evan engaged in variable levels of inappropriate sitting (M = 20%; range: 3%–43%), and moderate to high levels of off-task behavior (M = 58%; range: 40%–77%). The GBG was introduced in session 7 and Evan's appropriate sitting reduced to 2% (range: 0%–5%), whilst off-task behavior reduced to an average of 8% (range: 5%–13%). The GBG was withdrawn in session 13, and inappropriate sitting gradually increased to pretreatment levels (M = 12%; range: 2%–23%), whilst off-task behavior increased to an average of 29% (range: 19%–40%), with an increasing trend. When the GBG was reintroduced, inappropriate sitting and off-task behavior decreased to a low and stable level (M = 4%, range: 2%–6% and M = 4%, range: 0%–8%, respectively). Evan was absent during session 12.

Figure 4 displays the percentage of intervals with inappropriate sitting and off-task behavior for Ieuan. During the pretreatment baseline, Ieuan engaged in low levels of inappropriate sitting (M = 20%; range: 12%–27%) and high but variable levels of off-task behavior, averaging 56% (range: 30%–83%). Upon introduction of the GBG in session 7, inappropriate sitting reduced to an average of 7% (range: 4%–11%), and off-task behavior reduced to an average of 14% (range: 0%–6%), with some variability. The GBG was then withdrawn and inappropriate sitting and off-task behavior returned to similar levels observed in the pretreatment baseline (M = 14%, range: 10%–19% and M = 44%, range: 30%–57%, respectively). Upon reintroduction of the GBG, inappropriate sitting returned to a low and stable level (M = 4%; range: 3%–4%), whilst off-task behavior reduced to an average of 15% (range: 6%–25%), with a decreasing trend. Ieuan was absent during session 13.

Table 1 displays the results of the PND and Tau-U effect size calculations. Phase comparisons between baseline and intervention and withdrawal to reimplementation suggest that the GBG was effective to very effective at reducing inappropriate sitting for Harri (*PND* = 80%; *Tau-U* = −.70 and *PND* = 100%; *Tau-U* = −1,

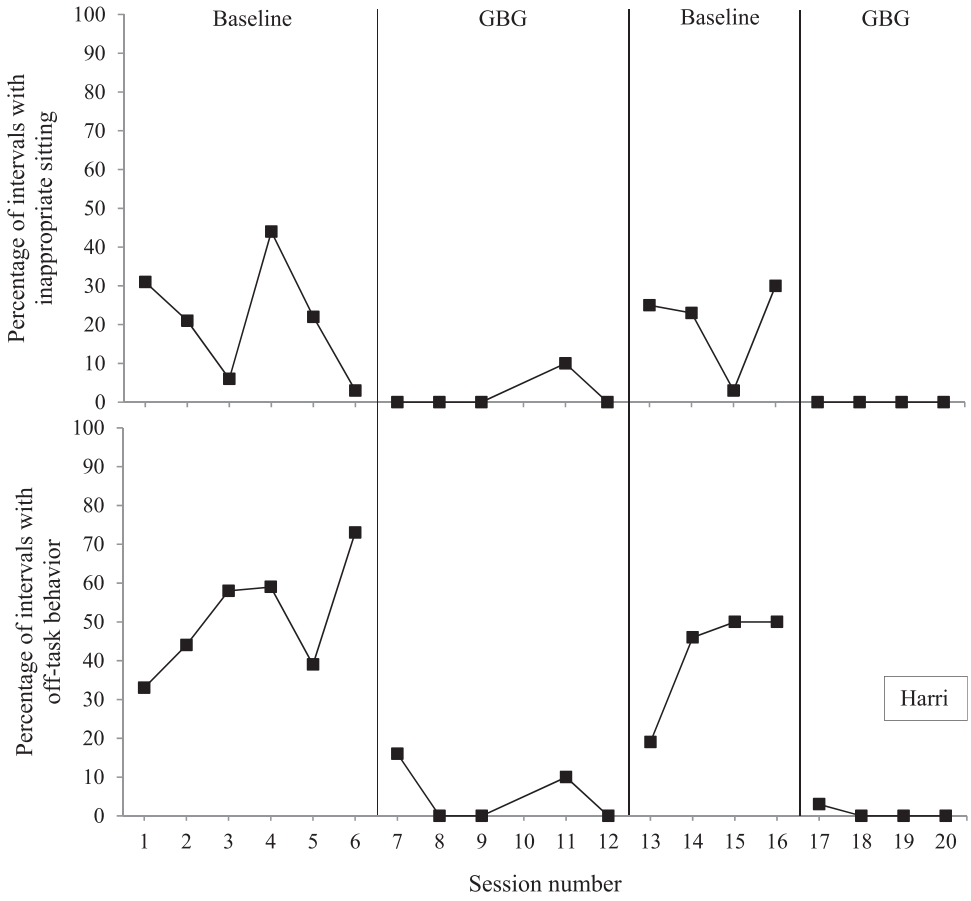

FIGURE 2 Percentage of Intervals with Problem Behavior for Harri. The top panel displays intervals with inappropriate sitting and the bottom panel displays intervals with off-task behavior across baseline and GBG phases. GBG, good behavior game

respectively) and Ieuan (*PND* = 100%; *Tau-U* = −1 and *PND* = 100%; *Tau-U* = −1, respectively). For Evan, there was nonagreement between PND and Tau-U calculations. Evan's Tau-U comparisons between baseline and intervention and withdrawal to reimplementation for inappropriate sitting suggest medium to strong effects (*Tau-U* = −1 and *Tau-U* = −.75, respectively). However, PND scores for the same comparisons suggest weak or no observed effect (*PND* = 60% and *PND* = 0%, respectively). Phase comparisons between baseline and intervention and withdrawal to reimplementation suggest that the GBG was very effective at reducing off-task behavior for Harri (*PND* = 100%; *Tau-U* = −1 and *PND* = 100%; *Tau-U* = −1, respectively), Evan (*PND* = 100%; *Tau-U* = −1 and *PND* = 100%; *Tau-U* = −1, respectively), and Ieuan (*PND* = 100%; *Tau-U* = −1 and *PND* = 100%; *Tau-U* = −1, respectively).

During the social validity assessment, the teacher stated that the GBG addressed the most pressing behavioral issues in her classroom, she felt sufficiently involved in the development of the game, she found it easy to implement, and it did not interfere with academic instruction. When asked questions regarding outcomes of the GBG, the teacher said she noticed meaningful decreases in her students' problem behavior, and felt she spent less time addressing those behaviors when using the intervention. When asked to what degree she felt the GBG was suitable for SEN classrooms such as her own, the teacher stated that she felt the procedures were very suitable and that she would continue to use the game in the future. The teacher also stated that she felt having students work in teams toward a common goal was a fair strategy for managing problem behavior in the classroom.

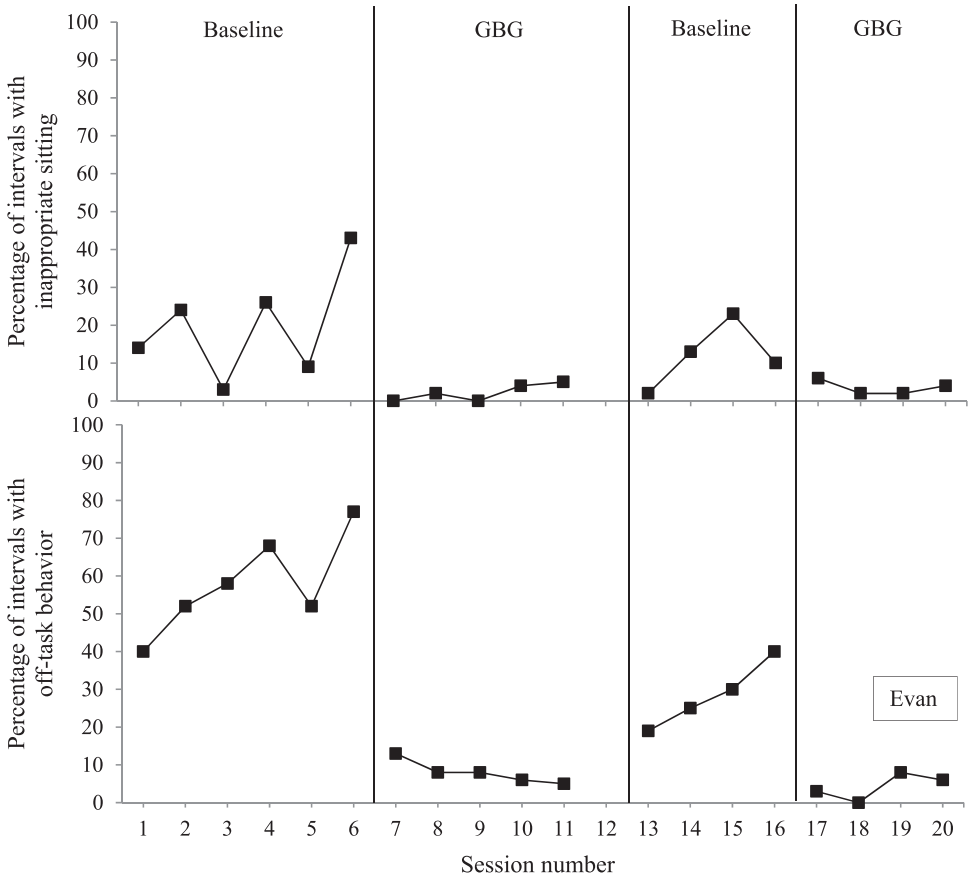

FIGURE 3 Percentage of Intervals with problem behavior for Evan. The top panel displays intervals with inappropriate sitting and the bottom panel displays intervals with off-task behavior across baseline and GBG phases. GBG, good behavior game

Table 2 displays the results of the participants' social validity questionnaires. All students said that they enjoyed playing the GBG and they found the rules easy to understand, whilst most students said that they felt their behavior was better when they played the game and they would like to continue playing the game in the future. However, when asked if they felt they got more schoolwork done when playing the game, results were mixed. When asked if they felt that earning points in teams was a fair strategy, all students agreed that it was.

The results of Study 2 provide additional evidence of the GBG's viability as a classroom management strategy for students with ID. Visual analysis of the data showed that inappropriate sitting and off-task behavior reduced in all participants during sessions in which the game was played. For the most part, effect size calculations supported these findings. PND comparisons between baseline and intervention and withdrawal to reimplementation suggested a weak or no effect on Evan's inappropriate sitting, likely due to data variability (Ledford & Gast, 2018). Furthermore, Evan's levels of inappropriate sitting reached a therapeutic floor, which also affected PND calculations.

Despite these discrepancies, the results are encouraging. Replication of effects across a similar classroom and with a stronger research design bolster the findings of Study 1. Further, Study 2 demonstrated that effects were produced with fewer modifications to the game. Although we continued to use pictures to augment the rules and displayed team scores using a concrete concept, we did not provide individualized copies of the rules or proximal scoreboards. Further, we included additional rules and larger teams. Throughout the game and during the social

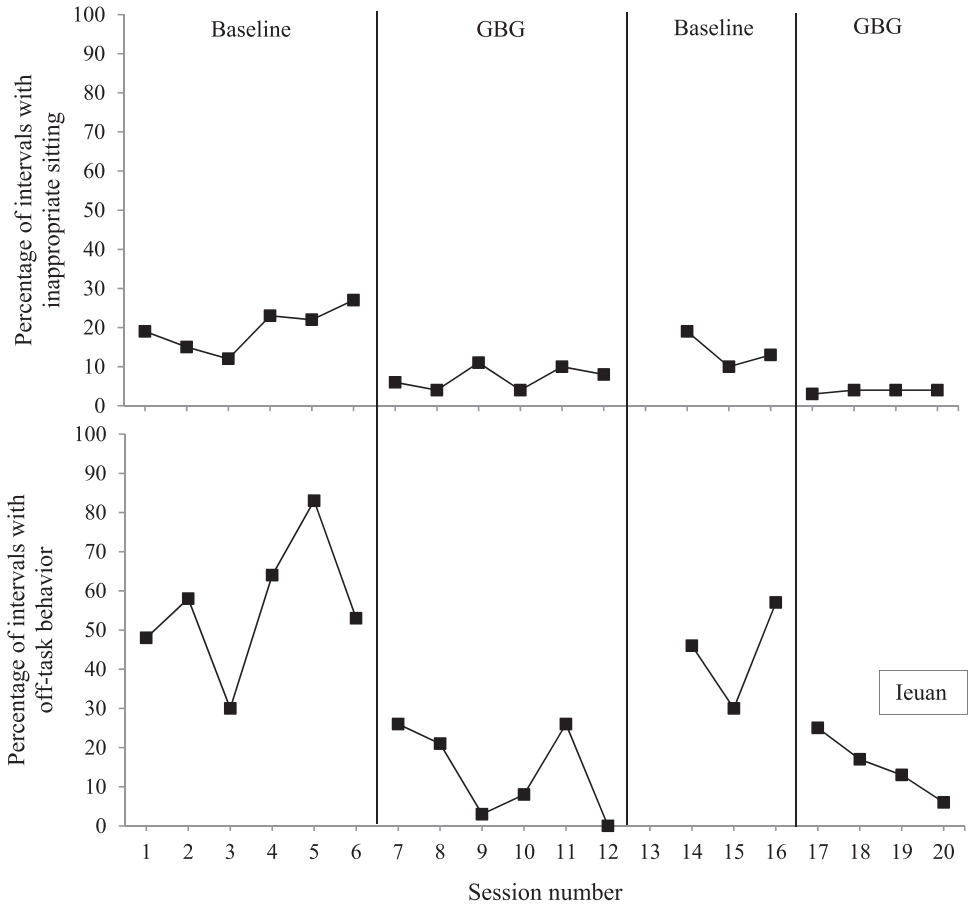

FIGURE 4 Percentage of Intervals with problem behavior Ieuan. The top panel displays intervals with inappropriate sitting and the bottom panel displays intervals with off-task behavior across baseline and GBG phases. GBG, good behavior game

validity interviews, students communicated a clear understanding of game play, suggesting adaptations to the original versions of the game (including the CBGG) may not be necessary for students with lower cognitive ability.

6.7 | General discussion

The purpose of the current studies was to evaluate the effects of the GBG on disruptive behavior in special education classrooms for student with intellectual and physical disabilities. We made adaptations to the game to suit students needs in both studies, but made fewer adaptations in Study 2. Results showed that the GBG effectively reduced problem behavior and that minimal adaptations were required for students to understand the contingencies and win the game. The teacher's agreement in Study 2 that the GBG was an appropriate strategy for SEN classrooms bolsters these conclusions.

In both experiments, minimal procedural adaptations were required to produce behavior change. However, additional adaptations may be necessary for students with lower cognitive ability than the students in our study. Given the potential of the GBG to both improve classroom behavior and promote positive social interactions among peers (Groves & Austin, 2019), further research on adapting the GBG for students with more severe

TABLE 1 Results of effect size calculations

	Effect sizes			
	Baseline to intervention		Withdrawal to reimplementation	
	PND	Tau-U	PND	Tau-U
Harri				
Inappropriate sitting	80%	−.70	100%	−1
Off-task behavior	100%	−1	100%	−1
Evan				
Inappropriate sitting	60%	−1	0%	−.75
Off-task behavior	100%	−1	100%	−1
Ieuan				
Inappropriate sitting	100%	−1	100%	−1
Off-task behavior	100%	−1	100%	−1

Abbreviation: PND, percentage nonoverlapping data.

disabilities might prove incredibly useful to SEN teachers and their pupils. It is possible that the original version of the game, in which points are delivered immediately after rule infractions (rather than after a period of time engaged in rule-following) might be a more effective arrangement. Picture schedules (e.g., Pierce et al., 2013) to illustrate contingencies could also be explored as a useful adaptation.

The study adds to the growing body of literature on the use of the GBG in alternative and special education classrooms, and the findings are consistent with the majority of research conducted in those settings (Groves & Austin, 2017; Joslyn et al., 2014, 2019; Rubow et al., 2018; Salend et al., 1989; Sy et al., 2016). However, our

TABLE 2 Results of participants' social validity questionnaires

Statement	Number of responses		
	Yes	Maybe	No
Did you enjoy playing the good behavior game?	7	0	0
Is your behavior better when you play the GBG?	5	2	0
Does playing the game help you get more schoolwork done?	3	2	2
Did you find the rules of the game easy to understand?	7	0	0
Everyone in your team has to follow the rules of the game to earn points. Do you think this is fair?	7	0	0
Would you like to continue playing the game in class?	6	0	1
Statement	Responses		
What did you like best about playing the game?	Receiving a reward (5)		
	Earning points (1)		
	Doing my work (1)		
What is your least favorite part of playing the game?	No answer (4)		
	Waiting for rewards (1)		
	Not earning points (1)		
	I find it distracting (1)		

Abbreviation: GBG, good behavior game.

results are inconsistent with Breeman et al. (2016), who found that the GBG neither improved nor worsened behavior problems in SEN classrooms. The disparity in findings between the current study and Breeman et al. may be a result of different measurement approaches, rather than differences in GBG effectiveness per se. In the current study and in much of the extant literature, repeated, direct observation measures were employed. Breeman et al. used teacher and student reports to determine the effectiveness of the GBG on problem behavior, which were conducted at two specific points in time (just before the start of intervention and at the end of the school year). Due to the timing of these reports, it is possible that teachers and students were referring to behavior more generally, and not referring to specific periods in which the GBG was played. Studies employing single-case designs and direct observations tend only to report data on specific times of day (typically when the teacher reports that behavior is particularly problematic). Occurrence of problem behavior outside of these observations (as well as behaviors not targeted by the intervention), typically are not considered. Further, previous research has demonstrated that the GBG effects do not tend to generalize to subsequent contexts in which the GBG is not played (Donaldson et al., 2015; Pennington & McComas, 2017). Thus, it is possible that more global teacher reports may underestimate the GBG as an effective classroom management strategy for common, low intensity behavior problems. This may also account for the disparity between recent meta-analyses, which have reported positive effects for the GBG when analyzing single-case designs (e.g., Bowman-Perrott et al., 2016) and less compelling effects when the analysis was targeted at larger-scale studies (Smith et al., 2019).

Despite somewhat contradictory findings across the GBG literature, studies employing single-case designs have repeatedly demonstrated that the intervention is effective and well-liked by teachers. The ease with which teachers can learn the strategy and implement it with integrity is another important selling point. Although individual training formats, in which the researcher trains the teacher in a one-to-one situation, are typical in previous GBG studies (Flower et al., 2014a; Lannie & McCurdy, 2007; Mitchell et al., 2015; Wahl et al., 2016), the teacher in Study 1 was trained as part of a large group training. However, measures of the integrity with which she implemented the intervention were commensurate with Study 2's teacher, who participated in an individualized training. School-wide trainings more cost effective and time efficient, and the current results provide preliminary evidence that these formats can be used without a loss to GBG integrity.

ACKNOWLEDGMENTS

The authors thank Wayne Murphy and Alec Clark for their support in conducting the study and Venessa Lewis for her assistance with data collection. This research was funded by The Tai Federation.

CONFLICT OF INTERESTS

The other authors declare that there are no conflict of interests.

ORCID

Emily A. Groves https://orcid.org/0000-0002-4460-2455
Jennifer L. Austin http://orcid.org/0000-0001-8751-4941

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. https://doi.org/10.1901/jaba.1969.2-119

Bowman-Perrott, L., Burke, M. D., Zaini, S., Zhang, N., & Vannest, K. (2016). Promoting positive behavior using the Good Behavior Game: A meta-analysis of single-case research. *Journal of Positive Behavior Interventions*, *18*, 180–190. https://doi.org/10.1177/1098300715592355

Breeman, L. D., van Lier, P. A. C., Wubbels, T., Verhulst, F. C., van der Ende, J., Maras, A., Struiksma, A. J., Hopman, J. A. B., & Tick, N. T. (2016). Effects of the Good Behavior Game on the behavioral, emotional, and social problems of children

with psychiatric disorders in special education settings. *Journal of Positive Behavior Interventions*, *18*, 156–167. https://doi.org/10.1177/1098300715593466

Department for Education. (2013). Teacher voice omnibus: May 2013 survey - pupil behaviour. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/210297/DFE-RR304.pdf

Department for Education. (2019). Permanent and fixed period exclusions in England: 2017 to 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/820773/Permanent_and_fixed_period_exclusions_2017_to_2018_-_main_text.pdf

Derrington, C. (2008). Behaviour in primary schools: Final report. http://www.channel4.com/documentaries/media/D/dispatches/Dispatches_Behaviour%20in%20Primary%20Schools_%20Survey.doc

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

Erasmus Medical Center. (2000). *Problem behavior at school interview*, Rotterdam, The Netherlands: Author.

Estyn. (2016). *The annual report of her majesty's chief inspector of education and training in Wales: 2015-2016*. Retrieved from: https://www.estyn.gov.wales/sites/default/files/documents/ESTYN_Annual%20Report%202015_2016_English.pdf

Estyn. (2017). *The annual report of her majesty's chief inspector of education and training in Wales: 2016-2017*. Retrieved from: https://www.estyn.gov.wales/sites/default/files/documents/Estyn_Annual%20Report_2018_English_Accessible.pdf

Estyn. (2018). *The annual report of her majesty's chief inspector of education and training in Wales: 2017-2018*. Retrieved from: https://www.estyn.gov.wales/sites/default/files/documents/ESTYN_Annual%20Report_Accessible_English__2018.pdf

Fiske, K., & Delmolino, L. (2012). Use of discontinuous methods of data collection in behavioral intervention: Guidelines for practitioners. *Behavior Analysis in Practice*, *5*, 77–81. https://doi.org/10.1007/BF03391826

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

Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., Vega, R. Jr (2014b). Effects of the Good Behavior Game on challenging behaviors in school settings. *Review of educational research*, *84*, 546–571. https://doi.org/10.3102/0034654314536781

Groves, E. A., & Austin, J. L. (2017). An evaluation of interdependent and independent group contingencies during the Good Behavior Game. *Journal of Applied Behavior Analysis*, *50*, 552–566. https://doi.org/10.1002/jaba.393

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

Joslyn, P. R., & Vollmer, T. R. (2020). Efficacy of teacher-implemented Good Behavior Game despite low treatment integrity. *Journal of Applied Behavior Analysis*, *53*(1), 465–474. https://doi.org/10.1002/jaba.614

Joslyn, P. R., Vollmer, T. R., & Hernández, V. (2014). Implementation of the Good Behavior Game in classrooms for children with delinquent behavior. *Acta de Investigación Psicológica*, *4*, 1673–1682. https://doi.org/10.1016/S2007-4719(14)70973-1

Joslyn, P. R., Vollmer, T. R., & Kronfli, F. R. (2019). Interdependent group contingencies reduce disruption in alternative high school classrooms. *Journal of Behavioral Education*, *28*, 423–434. https://doi.org/10.1007/s10864-019-09321-0

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single case designs technical documentation. In What Works Clearinghouse: Procedures and standards handbook (version 2.0).*

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

Ledford, J. R., & Gast, D. L. (2018). *Single case research methodology: Applications in special education and behavioral sciences*, New York, NY: Routledge.

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

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

Office for Standards in Education, Children's Services, and Skills. (2005). *Managing challenging behaviour*. Retrieved from http://www.ofsted.gov.uk/resources/managing-challenging-behaviour

Office for Standards in Education, Children's Services, and Skills. (2014). *Below the radar: Low-level disruption in the country's classrooms.* Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/379249/Below_20the_20radar_20-_20low-level_20disruption_20in_20the_20country_E2_80_99s_20classrooms.pdf

Office for Standards in Education, Children's Services, and Skills. (2018). *The annual report of her majesty's chief inspector of education, children's services and skills 2017/18*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/761606/29523_Ofsted_Annual_Report_2017-18_041218.pdf

Office for Standards in Education, Children's Services, and Skills. (2019). Teacher well-being at work in schools and further education providers. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819314/Teacher_well-being_report_110719F.pdf

Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, *40*, 357–367. https://doi.org/10.1016/j.beth.2008.10.006

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011a). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, *35*, 303–322. https://doi.org/10.1177/0145445511399147

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011b). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*, 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Pennington, B., & McComas, J. J. (2017). Effects of the Good Behavior Game across classroom contexts. *Journal of Applied Behavior Analysis*, *50*, 176–180. https://doi.org/10.1002/jaba.357

Pierce, J. M., Spriggs, A. D., Gast, D. L., & Luscre, D. (2013). Effects of visual activity schedules on independent classroom transitions for students with autism. *International Journal of Disability, Development and Education*, *60*, 253–269. https://doi.org/10.1080/1034912X.2013.812191

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

Salend, S. J., Reynolds, C. J., & Coyle, E. M. (1989). Individualizing the Good Behavior Game across type and frequency of behavior with emotionally disturbed adolescents. *Behavior Modification*, *13*, 108–126. https://doi.org/10.1177/01454455890131007

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single subject research: Methodology and validation. *Remedial and Special Education*, *8*, 24–33. https://doi.org/10.1177/074193258700800206

Smith, S., Barajas, K., Ellis, B., Moore, C., McCauley, S., & Reichow, B. (2019). A meta-analytic review of randomized controlled trials of the Good Behavior Game. *Behavior Modification*, 1–26. https://doi.org/10.1177/0145445519878670

Sy, J. R., Gratz, O., & Donaldson, J. M. (2016). The Good Behavior Game with students in alternative educational environments: Interactions between reinforcement criteria and scoring accuracy. *Journal of Behavioral Education*, *25*, 455–477. https://doi.org/10.1007/s10864-016-9257-0

Tingstorm, D. H. (1994). The Good Behavior Game: An investigation of teachers' acceptance. *Psychology in the Schools*, *31*, 57–65. https://doi.org/10.1002/1520-6807

Tingstorm, D. H., Sterling, H., & Wilczynski, S. (2006). The Good Behaviour Game: 1969-2002. *Behaviour Modification*, *30*, 225–253. https://doi.org/10.1177/0145445503261165

Wahl, E., Hawkins, R. O., Haydon, T., Marsicano, R., & Morrison, J. Q. (2016). Comparing versions of the Good Behavior Game: Can a positive spin enhance effectiveness? *Behavior Modification*, *40*, 493–517. https://doi.org/10.1177/0145445516644220

Welsh Government. (2018). Permanent and fixed-term exclusions from schools in Wales, 2016/17. Retrieved from https://gov.wales/sites/default/files/statistics-and-research/2018-12/181002-permanent-fixed-term-exclusions-from-schools-2016-17-en.pdf

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*, 203–214. https://doi.org/10.1901/jaba.1978.11-203

Wright, R. A., & McCurdy, B. L. (2011). Class-wide positive behavior support and group contingencies: Examining a positive variation of the Good Behavior Game. *Journal of Positive Behavior Interventions*, *14*, 173–180. https://doi.org/10.1177/1098300711421008

How to cite this article: Groves, E. A., May, R. J., Rees, R. E., Austin, J. L. (2021). Adapting the good behavior game for special education classrooms. *Psychology in the Schools*, 1–17. https://doi.org/10.1002/pits.22496