


Effects of the Good Behavior Game on Students' Academic Engagement in Remote Classrooms During the COVID-19 Pandemic

Journal of Positive Behavior
Interventions
1–13

© Hammill Institute on Disabilities 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10983007231168400
jpbj.sagepub.com



**Bridget O. Hier, PhD, BCBA^{1,2}, Connor K. MacKenzie, BS²,
Tory L. Ash, MS^{2,3}, Samantha C. Maguire, MS², Kaytlin A. Nelson, MS²,
Emily C. Helminen, MS² , Emily A. Watts, MS²,
Erin S. M. Matsuba, MS², Ellen C. Masters, MS²,
Carly C. Finelli, MA/AC^{2,4}, Joshua J. Circe, MS², Taylor J. Hitchings, BA²,
Alec R. Goldstein, MS², and William E. Sullivan, PhD⁵**

Abstract

This multiple-baseline design study examined the effects of the Good Behavior Game (GBG) on class-wide academic engagement in online general education classrooms. Teachers in three third- through fifth-grade classrooms implemented the GBG remotely during the COVID-19 pandemic. Treatment integrity was supported using aspects of implementation planning and by providing emailed performance feedback. Teachers' perceived usability and students' perceived acceptability of the GBG were assessed. Visual analysis results indicated two clear demonstrations of an effect, but experimental control was limited by smaller and delayed effects in one classroom. Statistical analyses of the data suggest that implementing the GBG was associated with moderate to strong, statistically significant improvements in students' academic engagement in all three classrooms. Teachers reported that the GBG was usable in their online classrooms, and students reported finding the intervention acceptable to participate in remotely. These results provide initial support for further examining the effectiveness and social validity of using the GBG to improve elementary students' academic engagement during remote instruction.

Keywords

classroom management, remote learning, online instruction, intervention acceptability, usability

Academic engagement is crucial for student success and has been defined behaviorally as engaging in passive or active behaviors that promote learning, such as participating in class, reading to oneself, or attending to the teacher (e.g., Fallon et al., 2020; Murphy et al., 2020). Student- and teacher-reported engagement levels correlate with grades in elementary school (Furrer & Skinner, 2003) and may have long-lasting positive impacts on graduation rates up to 14 years later (Alexander et al., 1997). Unfortunately, during the COVID-19 pandemic, students' levels of disengagement increased globally across multiple age groups (e.g., Novianti & Garzia, 2020; Zaccoletti et al., 2020). These reports of increased disengagement coincided with an unprecedented, national shift to online education in the wake of COVID-19.

Unfortunately, past research on the impact of online education and intervention is limited (McPherson & Bacow,

2015); however, online platforms offer potential benefits that may prove useful beyond the scope of a global pandemic, such as increased accessibility and flexibility (e.g., Dhawan, 2020). Recent studies found that traditionally in-person strategies were feasible and effective at increasing student engagement (Jia et al., 2022) and academic

¹State University of New York-Cortland, USA

²Syracuse University, NY, USA

³University of Wisconsin-Madison, USA

⁴Lancaster Central School District, NY, USA

⁵SUNY Upstate Medical University, Syracuse, NY, USA

Corresponding Author:

Bridget O. Hier, Department of Psychology, State University of New York at Cortland, P.O. Box 2000, Cortland, NY 13045, USA.

Email: bridget.hier@cortland.edu

Action Editor: Allison Bruhn

performance (Klein et al., 2022) when implemented online, suggesting that the benefits of existing evidence-based interventions may carry over to virtual settings. Thus, research examining the usability and effects of online interventions targeting academic engagement are both timely and relevant to educators who may extend their use of online education platforms beyond the context of the COVID-19 pandemic.

The Good Behavior Game

The Good Behavior Game (GBG) is an evidence-based behavior management strategy that teachers can use to improve academic engagement in their classrooms (Fallon et al., 2020; Murphy et al., 2020). The GBG is implemented as a class-wide intervention that uses an interdependent group contingency by (a) separating students into teams, (b) providing points to teams for displaying problematic behaviors, and (c) rewarding the team with the fewest points (Barrish et al., 1969). The GBG has consistently demonstrated large, positive effects on students' behavior across prekindergarten through grade 12 (TauU = 0.82, confidence interval [CI] = [0.78, 0.87]), and those effects can be enhanced with the combined use of daily and weekly rewards (Bowman-Perrott et al., 2016).

Since its conception, numerous modifications to the GBG have been empirically tested. One common modification is to award points for desired behaviors rather than undesired behaviors (e.g., Fallon et al., 2020; Murphy et al., 2020). Additional modifications have included changes to the setting where the game is implemented. While the GBG has not yet been evaluated online, positive effects have been shown in non-traditional classroom settings (e.g., Joslyn et al., 2014), libraries (Fishbein & Wasik, 1981), and lunchrooms (Grasley-Boy & Gage, 2022). These studies highlight the GBG's flexibility for implementation in various contexts.

Behavior-Specific Praise

An important component of modified GBG procedures that award points for desired behavior is the delivery of behavior-specific praise (BSP). Whereas general praise (GP) is any positive behavior-contingent feedback delivered without stating a behavior (e.g., "Good job!"; Lastrapes et al., 2018), BSP denotes a specific social or academic behavior (e.g., "Great job starting your work!"; Sallése & Vannest, 2020). Following several demonstrations that BSP improves student behavior and engagement (e.g., Rathel et al., 2014), a systematic review concluded that BSP is a potentially evidence-based practice with large between-case standardized mean difference effect sizes (range = 1.42–3.01; Royer et al., 2019).

Despite the effectiveness of praise (Spilt et al., 2016), teachers tend to underuse it during instruction (Beaman &

Wheldall, 2000) and often require supports to increase BSP (Rathel et al., 2014). Fortunately, because BSP is provided when points are awarded for positive behavior during the GBG, implementing the GBG can increase teachers' use of BSP (e.g., Lastrapes et al., 2018). In fact, multilevel modeling results from a large intervention study (Spilt et al., 2016) point to praise as a key active ingredient in the GBG. In that study, the GBG indirectly affected child outcomes through increased teacher praise and fewer reprimands.

Considerations of Treatment Integrity, Acceptability, and Usability in the GBG

Because educators often struggle to implement interventions as intended (Noell et al., 2005), researchers have considered different methods to monitor treatment integrity of the GBG (Joslyn et al., 2020). Most commonly, procedural checklists are developed and scored dichotomously to ensure that all planned steps of the intervention were adhered to (e.g., Fallon et al., 2020). When treatment integrity is low, performance feedback can be used to improve implementation (Fallon et al., 2015). Online classrooms, however, present unique challenges to the traditional delivery of feedback, as one cannot physically enter the classroom and pull a teacher aside to consult. Fallon et al. (2018) sent integrity data, praise for correct GBG implementation, and error correction via email and found that teachers' integrity improved, suggesting that feedback can effectively be provided outside of traditional face-to-face meetings.

In addition to treatment integrity considerations, examining students' perceptions of acceptability and teachers' perceived usability of the GBG may inform future intervention efforts, as higher acceptability and usability ratings are associated with more positive treatment outcomes (Eckert et al., 2017; Neugebauer et al., 2016). The GBG is considered acceptable and usable when implemented in person (e.g., Grasley-Boy & Gage, 2022); however, research has yet to examine perceptions of the GBG's acceptability and usability in an online format.

Purpose of the Current Study

Although the GBG has been effectively applied across educational settings (e.g., Fishbein & Wasik, 1981; Grasley-Boy & Gage, 2022), no studies have examined its effectiveness in an online classroom setting. Given the ubiquity of online learning during the COVID-19 pandemic, as well as its potential use beyond the pandemic, it is important to examine the GBG's generalizability to online platforms. Thus, the purpose of this study was to extend the GBG literature base to examine its effects on students' academic engagement in online general education classrooms while remotely supporting teachers' treatment integrity—including their use of BSP—through emailed performance

feedback. We also examined teachers' and students' perceptions of the GBG's usability and acceptability, respectively. Our research questions were as follows:

Research Question 1: Did the GBG improve students' engagement in online general education classrooms? Based on meta-analytic results that showed targeting on-task behavior had a positive effect on students' outcomes across studies and settings ($TauU = 0.59$; Bowman-Perrott et al., 2016), we predicted that implementation of the GBG would increase students' academic engagement.

Research Question 2: To what extent did teachers find the GBG usable in online general education classrooms? Based on previous research demonstrating that teachers find the GBG and its variations to have adequate acceptability (e.g., Mitchell et al., 2015), an aspect of usability (Briesch et al., 2013), we predicted that the teachers would perceive the GBG to be moderately usable.

Research Question 3: To what extent did students find the GBG acceptable in online general education classrooms? We predicted that the students would perceive the GBG to be moderately to highly acceptable based on previous research (Grasley-Boy & Gage, 2022; Mitchell et al., 2015).

Method

Participants and Setting

This study was conducted during the spring of the 2020–2021 school year. We recruited classrooms from a medium-sized elementary school, serving kindergarten through sixth grade, in a small city in upstate New York. The most recent publicly available school-level demographic data (New York State Education Department [NYSED], 2019) indicate that 403 students were enrolled in the school, with 50% identified as male. Approximately 87% of the students were identified as White, 6% as Hispanic or Latino, 1% as Asian or Native Hawaiian or Other Pacific Islander, 1% as Black, and 3% as Multiracial. In addition, 4% of the students were English Language Learners, 14% were identified as having a disability, and 51% were economically disadvantaged (as defined by NYSED, 2019).

A prior to the COVID-19 pandemic, all instruction at the participating school was in person. In response to the global pandemic, caregivers were given the option of hybrid instruction or fully remote instruction. There was one fully remote classroom at each grade level; we only recruited those classrooms. Following Institutional Review Board (IRB) and district approval, three teachers consented to participate, and we sought passive parental consent and student assent. Per our IRB protocol, we kept a password-protected list of students without consent or assent to ensure we did

not collect data from them. Although they did not participate in the GBG, they participated in classroom instruction while the GBG was in effect. All instruction was provided online using the Google Meet platform. Students met with their teacher for synchronous instruction four days a week and completed asynchronous instruction one day a week. We only collected data during synchronous instruction.

Teacher A was a 32-year-old White woman with a master's degree who had taught for 9 years, eight of which were spent at the current school. Her fourth-grade classroom (Classroom A) consisted of 23 students, of which 73.9% ($n = 17$) assented to participate.

Teacher B was a 34-year-old White woman with a bachelor's degree who had taught for 3 years. This was her first year teaching at the current school, where she served as a substitute teacher for a third-grade classroom whose typical teacher was on a maternity leave. This study began during Teacher B's second week working in that classroom (Classroom B), which consisted of 19 students. A total of 16 students (84.2%) assented to participate.

Teacher C was a 54-year-old White woman with a master's degree who had taught for 17 years, all of which were spent at the current school. Her fifth-grade classroom (Classroom C) consisted of 20 students, of which 85% ($n = 17$) assented to participate.

Dependent Measure: Direct Observation of Students' Academic Engagement

Graduate researchers directly observed student participants' academic engagement in their online classrooms. Academic engagement was defined by actively engaged behaviors (i.e., answering or asking a question in the chat or aloud, raising one's hand physically or by displaying a "raise hand" icon, or complying with class-wide demands or instructions) or passively engaged behaviors (i.e., physically orienting one's body toward the screen and sitting down). Given that all teachers expressed concern about students' engagement when their cameras were off (see the "Procedures" section), students whose cameras were off were coded as not engaged, unless they were displaying evidence of active engagement as described above. Other non-examples of engagement included manipulating objects unrelated to the task (e.g., playing with toys or pets), having one's body or face turned 90 to 180 degrees away from the screen, and not complying with class demands. This definition was developed upon interviewing the teachers to determine their expectations for engagement (see the "Procedures" section).

Similar to other GBG studies (e.g., Fallon et al., 2020), we used momentary time sampling to measure class-wide engagement. Observers collected engagement data for 10-min sessions split into 15-s intervals. At the start of each interval, observers identified a singular target student and noted whether the student demonstrated engagement at the

start of the interval. In rare instances, if observers were unable to see the target student at the start of the interval (e.g., due to technical difficulties), this was noted, and the interval was excluded from analysis. Students were observed on a rotating basis throughout the 10-min interval, starting at the top left of their screen, with observers cycling back through the students as time allowed.

Interval-by-interval inter-observer agreement (IOA) was conducted for 53% of sessions. It was calculated as the total number of agreements divided by the total number of intervals (Reed & Azulay, 2011), although any intervals during which either the primary or secondary observer could not see the target student (e.g., due to technical difficulties) were removed from calculations. On average, 8% of intervals were removed per session (range = 0%–53% of intervals removed). Across observers and sessions, and during intervals where both coders could see the target student, mean IOA was 99% (range = 98%–100%).

Descriptive Measures

Teacher Report of Classroom Management Self-Efficacy. Before the COVID-19 pandemic, most of the teachers' prior teaching experience was conducted in the context of in-person instruction. Thus, the *Classroom Management Self-Efficacy Instrument* (CMSEI; adapted from Slater & Main, 2020) was used to describe teachers' self-efficacy in classroom management when teaching in person and when teaching remotely. All original 14 items on the CMSEI were included. In addition, items were duplicated and adapted such that each of the 14 questions were asked first in the context of in-person instruction, and then again in the context of remote instruction (e.g., "When teaching in person, I am able to use a variety of behavior management models and techniques"; "When teaching remotely, I am able to use a variety of behavior management models and techniques"). Teachers were asked to respond to each of the 28 items on the adapted scale using a 4-point Likert-type scale ranging from *strongly disagree* (1) to *strongly agree* (4), with higher scores representing higher self-efficacy. Slater and Main's (2020) validation of the CMSEI with a third-year cohort of pre-service teachers demonstrated high reliability and accuracy of measurement (Person Separation Index = 0.89; Cronbach's $\alpha = .90$), as well as good construct, face, and content validity (M item fit residual = 0.09, $SD = 1.47$; M person fit residual = -0.41 , $SD = 1.35$).

Teacher Report of Changes in Students' Academic Engagement. The *Teacher Report of Students' Motivation and Engagement in Remote Instruction* scale (SMERI; adapted from Aguilera-Hermida, 2020) was used to describe teacher-reported changes in students' academic engagement during remote learning compared to their prior experiences with in-person instruction. This measure prompted the teachers to

report their experiences with students generally, not the particular students in this study, and like the CMSEI, it was used solely for descriptive purposes. This scale was originally sampled with 270 undergraduate- and graduate-level college students to measure their self-reported perceptions regarding their use and acceptance of emergency remote instruction (Aguilera-Hermida, 2020). The scale was adapted in the current study such that it only included the six items related to student engagement, and instructions were modified from its original self-report format to be used as a teacher-report measure. Teachers were provided with the following instructions: "Compared to in-person learning, describe changes in students' school performance during remote learning using the scale below," in which they were asked to answer six items (i.e., grades, knowledge, concentration, level of engagement, class attendance, and interest and enthusiasm) using a 5-point Likert-type scale ranging from *much worse* (1) to *much better* (5). The original scale produced high internal consistency across items related to engagement (Cronbach's $\alpha = .92$; Aguilera-Hermida, 2020), all of which were included in the current study.

Paired-Stimulus Preference Assessment. To determine daily and weekly rewards, we modified the paired-stimulus preference assessment (Fisher et al., 1992) for each classroom and populated it with reward options suggested by the teachers (see the "Procedures" section). For each preference assessment, which was administered via Qualtrics, two potential rewards were presented at a time below the prompt, "Which do you like more?," and students clicked on their preference. All potential rewards identified by the teacher for each classroom were included and paired together once.

Direct Observation of Teachers' BSP. We measured praise for treatment integrity purposes to ensure that points were delivered with BSP. General praise was defined as positive feedback delivered to a student or group of students that did not state a specific academic or social behavior (e.g., "Good job!"), and BSP was defined as feedback delivered to a student or group of students that identified a specific academic or social behavior (e.g., "Great job facing your screen during instruction!"); adapted from Lastrapes et al., 2018; Sallese & Vannest, 2020). During each 20-min GBG lesson, we tallied the number of times a point was awarded with GP, BSP, or no praise. At the end of each observation session, we calculated the percentage of points awarded with GP or BSP.

Teachers' Perceived Usability of the GBG. The *Usage Rating Profile–Intervention Revised* (URP-IR) is a self-report questionnaire that assesses six factors (i.e., Acceptability, Understanding, Home–School Collaboration, Feasibility, System Climate, and System Support), which capture individual, intervention, and system influences on the quality of use and maintenance of interventions (Briesch et al., 2013). The

URP-IR uses a 6-point Likert-type scale ranging from *strongly disagree* to *strongly agree* across 29 items. By reverse-scoring items in the Home–School Collaboration and System Support factors, high scores on all factors indicated more favorable perceptions of the GBG. In prior research with elementary teachers (Briesch et al., 2013), the URP-IR demonstrated acceptable levels of internal consistency across each subscale (Cronbach’s α range = .72–.95).

Students’ Perceived Acceptability of the GBG. The *Kids Intervention Profile* (KIP) is an eight-item self-report scale that has been adapted for use with the GBG (Grasley-Boy & Gage, 2022). The response options for each item include boxes of increasing size, used on a 5-point Likert-type scale, with higher scores representing higher acceptability. Ratings are summed across all items, and total scores greater than 24 indicate adequate acceptability of the intervention (Eckert et al., 2017). In previous studies with elementary students (Eckert et al., 2017), the internal consistency (Cronbach’s α = .79) and test–retest reliability across a 3-week period (r = .70) were adequate. A factor analysis indicated that this scale consists of two factors: Overall Intervention Acceptability and Skill Improvement. For this study, each question was modified from the original KIP to ask questions related to the GBG (e.g., “How much do you like playing the Good Behavior Game in your online class?”). The internal consistency of the adapted measure with our sample was lower (Cronbach’s α = .66) than in previous literature (Eckert et al., 2017), but still in the acceptable range.

Procedures

Problem Identification and Preference Assessment. The first author met with interested teachers ($n = 3$) via Google Meet to inform them of the study and obtain consent through Qualtrics. All participating teachers completed the CMSEI and the SMERI via Qualtrics. The following week, a graduate researcher conducted a problem identification interview (PII; Witt & Elliott, 1983) with each teacher independently via Google Meet or Zoom to gain information about their scope of concern regarding class-wide academic engagement, their expectations and goals, the operationalization of academic engagement, and potential daily and weekly rewards that could be provided remotely. A second researcher observed two of the three PII sessions (67%) and collected integrity data using a checklist to ensure the PII’s steps were adhered to. Procedural integrity was calculated by tallying the number of steps administered accurately, dividing by the total number of steps, and multiplying by 100. Across both teachers, the PII was conducted with 100% procedural integrity.

Upon completing the PII, we used the information provided by the teachers to develop a single definition of academic engagement that applied to all classrooms (see the “Dependent Measure: Direct Observation of Students’

Academic Engagement” section). In addition, we used the reward suggestions provided by each teacher to create a separate paired-stimulus preference assessment for each classroom (see the “Descriptive Measures” section). To inform daily and weekly reward selection, students completed their respective classroom preference assessment using Qualtrics under the instruction and supervision of their teacher. Response rates were 94.1%, 81.3%, and 82.4% for Classrooms A, B, and C, respectively. Teachers A and B consistently provided the single most desirable reward to winning teams (i.e., ending class early and free time with classmates in breakout rooms), whereas Teacher C provided winning teams with a choice from the five most desirable options from the preference assessment (i.e., spending time with the teacher, replacing their worst grade of the day with 100%, ending class early, free time with classmates in breakout rooms, playing online games). Across classrooms, weekly rewards were the same as daily rewards but were available for longer periods of time.

Baseline Phase. Graduate researchers conducted baseline observations of students’ academic engagement during the two lessons that teachers reported as having the lowest rates of class-wide academic engagement during the PII (i.e., English Language Arts [ELA] and math across all classrooms). During the observations, researchers entered the online classroom with their cameras and microphones off. Teachers were instructed to teach as they typically did.

GBG Implementation Training. We met with the teachers individually via Zoom for approximately 45 min to train them to implement the GBG in their online classrooms. Using a PowerPoint and following a detailed procedural checklist (see Note 1), we assisted each teacher in (a) developing three teams of approximately five to six students, (b) identifying explicit behavioral expectations and rules for academic engagement to teach and praise, (c) awarding points with BSP for academic engagement, (d) developing a daily and weekly reward system based on their classroom’s preference assessment results, and (e) developing a plan for the teachers to introduce the GBG to students. We also reviewed a copy of the GBG procedural checklist with teachers and discussed our plan to monitor treatment integrity and provide them with emailed feedback.

Next, we provided teachers with a task-analyzed lesson plan for introducing the GBG and modeled how to implement it before asking the teachers to practice it. The lesson plan prompted teachers to (a) introduce the GBG by name, (b) inform students of the goals of the GBG, (c) inform students when they would play the GBG, (d) describe the rules, (e) explicitly teach behavioral expectations for academic engagement, (f) show students where the team names and points would be displayed, and (g) describe the daily and weekly reward system. We corrected any incorrect steps and

had the teachers re-implement the steps until accurate. Finally, we engaged in aspects of implementation planning (e.g., Sanetti et al., 2018) by asking teachers to identify barriers that could interfere with GBG implementation and assisting with problem-solving. Researchers conducted all three teachers' training sessions with 100% procedural integrity.

Intervention Phase. Following baseline data collection and the GBG implementation training sessions, we told the teachers when to introduce the GBG in their classrooms using the task-analyzed lesson plan. Based on the teachers' schedules, they then implemented the GBG 1 to 3 times per day for approximately 20 min in their ELA and math lessons. During the intervention sessions, which spanned 1 to 4 days, teachers followed a procedural checklist to (a) announce the GBG was starting, (b) review the academically engaged behaviors that would earn points (e.g., cameras on, asking or answering questions, completing work), (c) inform students of the daily reward, (d) remind students of the weekly reward, (e) divide students into teams and display team names, (f) identify the point criterion (10–15 points) to earn the daily reward (see Note 2), (g) award points for academic engagement, (h) pair points with BSP, (i) identify whether each team met the point criterion, (j) provide the team(s) that reached the point criterion the daily reward, (k) withhold the daily reward from the team(s) that did not meet the point criterion, and (l) provide the team with the highest number of points the weekly reward. Consistent with baseline procedures, we observed academic engagement for 10-min sessions upon entering the online classrooms with our cameras and microphones off. Thus, we obtained two sessions of data for each lesson in which the GBG was implemented. This approach ensured we would have enough data points for a scientifically sound analysis in the event of unpredictable pandemic-related policy changes truncating the intervention phase, and it allowed for the display of both within- and between-session student behavior data (Fahmie & Hanley, 2008). Following each GBG implementation, we provided teachers with emailed feedback that included the percentage of GBG steps implemented correctly, the percentage of points awarded with BSP, a praise statement indicating positive aspects of implementation, and an area for improvement. Upon each classroom's completion of the intervention phase, the teachers completed the URP-IR, and students completed the KIP to report the GBG's usability and acceptability, respectively.

Attendance

We measured the number of students in attendance during each session across the baseline and GBG phases to determine whether classroom composition threatened the study's internal validity. Across all classrooms, attendance rates were similar between the baseline and GBG phases. For Classroom A, there was a mean attendance rate of 18 students per session

(range = 16–19) during baseline and 19 students per session (range = 16–22) during the GBG phase. Classroom B had a mean attendance rate of 13.6 students per session (range = 10–16) during baseline and 14.6 students per session (range = 14–16) during the GBG phase. Classroom C's mean session attendance was 16.5 students for both the baseline (range = 14–19) and GBG (range = 12–20) phases.

Treatment Fidelity

We assessed two aspects of treatment fidelity (Moncher & Prinz, 1991): (a) treatment integrity (i.e., the clinical component that examines whether the treatment was implemented as intended) and (b) treatment differentiation (i.e., the methodological component that examines the extent to which treatment and no-treatment conditions differed from one another). To accomplish this, we measured treatment integrity during 57% of baseline sessions and 100% of intervention sessions. Treatment integrity was calculated by tallying the number of GBG steps adhered to during each session, dividing by the total number of steps, and multiplying by 100. We also measured the percentage of GBG lesson plan steps adhered to when teachers initially introduced the GBG. We planned to provide teachers with additional training if integrity fell below 80% during any session. No retraining was required.

Table 1 summarizes the treatment fidelity and praise rates. Across all classrooms, mean treatment integrity was 0% during baseline, indicating no GBG components were implemented prior to the phase change. Teachers A and B introduced the GBG to their classrooms using the lesson plan with 100% integrity; Teacher C did so with 89% integrity. Specifically, Teacher C did not model academically engaged behaviors, and she prompted students to practice behavioral non-examples. During intervention, mean treatment integrity ranged from 91.4% for Classroom B to 100% for Classroom A. Overall, the fidelity data suggest that treatment was implemented with integrity and was adequately differentiated between phases.

Experimental Design and Data Analysis Plan

Using a multiple-baseline design (MBD), we collected data in each classroom until a predictable pattern of baseline responding was obtained. Then, teachers introduced the GBG in a time-lagged fashion, allowing for replication of our results across interventionists, settings, and participants. We collected at least five data points per phase, meeting the standards for a rigorous MBD as outlined by the What Works Clearinghouse (Kratochwill et al., 2010).

We then visually analyzed the data. As recommended by Kratochwill et al. (2010), our visual analysis consisted of four steps. First, we inspected the data for evidence of a stable baseline pattern of responding. Next, we examined the data within each phase for level, trend, and variability. Third, we compared patterns of responding between baseline and

Table 1. Treatment Fidelity, Awarded Points, and Praise Rates Across Classrooms.

Classroom	GBG treatment integrity ^a							Points awarded across GBG implementations			Percentage of points delivered with praise across GBG implementations					
	Baseline			GBG intro.	Intervention			Total points			Behavior-specific praise			General praise		
	M	Min	Max		M	Min	Max	M	Min	Max	M	Min	Max	M	Min	Max
Classroom A	0%	0%	0%	100%	100%	100%	100%	53	45	64	85%	76%	92%	11%	7%	14%
Classroom B	0%	0%	0%	100%	91%	83%	100%	23	22	24	97%	91%	100%	3%	0%	9%
Classroom C	0%	0%	0%	89%	95%	82%	100%	27	3	41	100%	100%	100%	0%	0%	0%

Note. GBG = Good Behavior Game.

^aTreatment integrity for the baseline and intervention phases was computed as the percentage of GBG steps adhered to for each session (baseline phase) or each GBG implementation (intervention phase). Contrasts between the baseline and intervention phases indicate treatment differentiation. Treatment integrity of the teachers' introduction of the GBG to students was computed as the percentage of steps adhered to on the lesson plan.

intervention phases to determine whether there was an impact of the GBG on students' academic engagement. Finally, we considered the study as a whole to determine whether we had sufficient evidence to suggest a functional relation between our independent and dependent variables. Following our visual analysis, we calculated effect sizes for each classroom according to recommendations by Tarlow (2017) by first computing Tau to determine whether a monotonic baseline trend was present and then controlling for that trend if needed.

Results

Below, we first present the results of the descriptive measures the teachers completed at baseline (see Table 2). Next, we describe the results of the MBD (see Table 3). Figure 1 displays (a) the percentage of intervals with academic engagement during each 10-min observation session and (b) treatment integrity for each GBG implementation. Finally, we describe the results of the usability and acceptability measures administered post-intervention (see Table 2).

Teacher Report of Classroom Management Skills

In general, teachers indicated a high degree of self-efficacy in their classroom management skills for both in-person and remote settings on the CMSEI (see Table 2). Teacher A reported greater self-efficacy when teaching in-person versus remotely. However, Teachers B and C reported similarly high levels of self-efficacy in both in-person and remote instruction.

Teacher Report of Students' Academic Engagement

On the SMERI (see Table 2), Teachers A and B reported a decrement in students' overall academic engagement during remote learning. Teacher C reported an increase.

Effect of the GBG on Students' Academic Engagement

Visual Analysis. During the baseline phase, Classroom A demonstrated low ($M = 38.41\%$), moderately variable ($SD = 9.35$) levels of academic engagement, with a decreasing trend and steep decline in engagement during the final baseline session. When the GBG was introduced, levels of engagement immediately increased and generally remained higher than baseline ($M = 56.93\%$), with a slight increasing trend. Although engagement was relatively stable throughout most of the intervention, there was a steep decrease in engagement during the middle two treatment sessions. During those sessions, despite 100% of GBG steps being adhered to, the observer noted that the tone of the praise statements was less genuine. During the GBG phase, Classroom A demonstrated a 48.22% increase in academic engagement from the mean of baseline sessions.

Classroom B demonstrated moderate levels ($M = 59.09\%$) and variability ($SD = 9.41$) of academic engagement with a decreasing trend across the baseline phase, particularly in ELA. Levels of engagement immediately and substantially increased upon introducing the GBG, remaining relatively high ($M = 77.39\%$) with slightly less variability ($SD = 8.37$). There was an increasing trend in engagement for the first four GBG sessions when treatment integrity was high. Levels of engagement then decreased during the final two sessions, which coincided with lower levels of treatment integrity and the use of social studies content for reading and writing, a different activity than their other ELA lessons. Following those sessions, we were unable to continue as the district required resumption of in-person instruction. Overall, there was a 31% increase in academic engagement relative to the mean of baseline sessions.

Throughout the duration of the study, Classroom C demonstrated high, moderately variable levels of academic engagement. As a result, changes in level, variability, and trend were most evident when the data were plotted by subject area. In

Table 2. Means and Standard Deviations for Teacher and Student Report Measures.

Measure	Total sample		Classroom A		Classroom B		Classroom C	
	<i>n</i> (%) ^a	<i>M</i> (<i>SD</i>)	<i>n</i> (%) ^a	<i>M</i> (<i>SD</i>)	<i>n</i> (%) ^a	<i>M</i> (<i>SD</i>)	<i>n</i> (%) ^a	<i>M</i> (<i>SD</i>)
CMSEI	3 (100)		1 (100)		1 (100)		1 (100)	
In-person		3.56 (0.92)		3.54 (0.41)		3.57 (1.17)		3.57 (1.17)
Remote		3.28 (0.85)		2.67 (0.65)		3.43 (0.85)		3.64 (0.74)
SMERI	3 (100)	2.72 (1.02)	1 (100)	2.00 (0.00)	1 (100)	2.17 (0.41)	1 (100)	4.00 (0.63)
Grades		2.67 (1.15)		2		2		4
Knowledge		2.67 (1.15)		2		2		4
Concentration		2.33 (0.58)		2		2		3
Engagement		3.00 (1.73)		2		2		5
Attendance		3.00 (1.00)		2		3		4
Enthusiasm		2.67 (1.54)		2		2		4
URP-IR	3 (100)	5.00 (1.00)	1 (100)	4.55 (0.99)	1 (100)	4.93 (0.96)	1 (100)	5.52 (0.83)
Acceptability		5.26 (0.53)		5.00 (0.00)		5.00 (0.50)		5.78 (0.44)
Understanding		5.78 (0.44)		5.67 (0.58)		5.67 (0.58)		6.00 (0.00)
Home-school		3.33 (1.32)		3.33 (1.53)		2.67 (0.58)		4.0 (1.73)
Feasibility		4.67 (1.03)		3.67 (1.03)		5.17 (0.75)		5.17 (0.41)
System climate		5.47 (0.52)		5.00 (0.00)		5.40 (0.55)		6.00 (0.00)
System support		5.00 (0.71)		4.33 (0.58)		5.00 (0.00)		5.67 (0.58)
KIP ^b	37 (74.0)	25.11 (4.82)	14 (82.4)	25.00 (6.11)	10 (62.5)	24.40 (2.76)	13 (76.5)	25.77 (4.76)
Acceptability ^c		19.97 (4.22)		19.57 (4.94)		19.70 (2.75)		20.62 (4.54)
Skill improvement ^c		5.14 (1.51)		5.43 (1.95)		4.70 (1.06)		5.15 (1.28)

Note. CMSEI = Classroom Management Self-Efficacy Instrument (1–4 scale); SMERI = Teacher Report of Students' Motivation and Engagement in Remote Instruction (1–5 scale); URP-IR = Usage Rating Profile–Intervention Revised (1–6 scale); KIP = Kids Intervention Profile (total scores reported).

^aPercent of total or classroom sample that responded. ^bTotal KIP scores greater than 24 demonstrate adequate acceptability. ^cAcceptability factor includes items 1 to 6 on the KIP. Skill Improvement factor includes items 7 and 8.

Table 3. Descriptive Statistics and Effect Sizes for the Percentage of Intervals With Academic Engagement Across Phases.

Classroom	Baseline		Good Behavior Game		% Change	Effect size estimate		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		Tau	<i>SE</i> _{Tau}	<i>p</i>
Classroom A								
Overall	38.41	9.35	56.93	12.07	48.22	0.52	0.30	.02
ELA	39.66	11.62	63.47	7.47	60.04			
Math	35.90	3.63	52.58	13.11	46.46			
Classroom B								
Overall	59.09	9.41	77.39	8.37	30.96	0.61	0.28	.006
ELA	55.31	10.14	77.39	8.37	39.92			
Math	64.76	4.76	—	—	—			
Classroom C								
Overall	75.95	7.29	84.77	7.78	11.61	0.44	0.25	.01
ELA	72.31	6.34	83.35	8.47	15.27			
Math	77.78	7.41	86.67	7.01	11.43			

Note. ELA = English Language Arts.

ELA, academic engagement was moderately high (*M* = 72.31%) with relatively low variability (*SD* = 6.34) during baseline. Despite a slightly increasing trend, academic engagement decreased during the final baseline session in ELA before immediately—albeit slightly—increasing above

baseline levels during the first GBG session in ELA. Overall, the level of academic engagement (*M* = 83.35%) increased by 15.27% during the GBG phase relative to the mean of baseline during ELA. The increased variability in engagement (*SD* = 8.47) during the GBG correlated with treatment

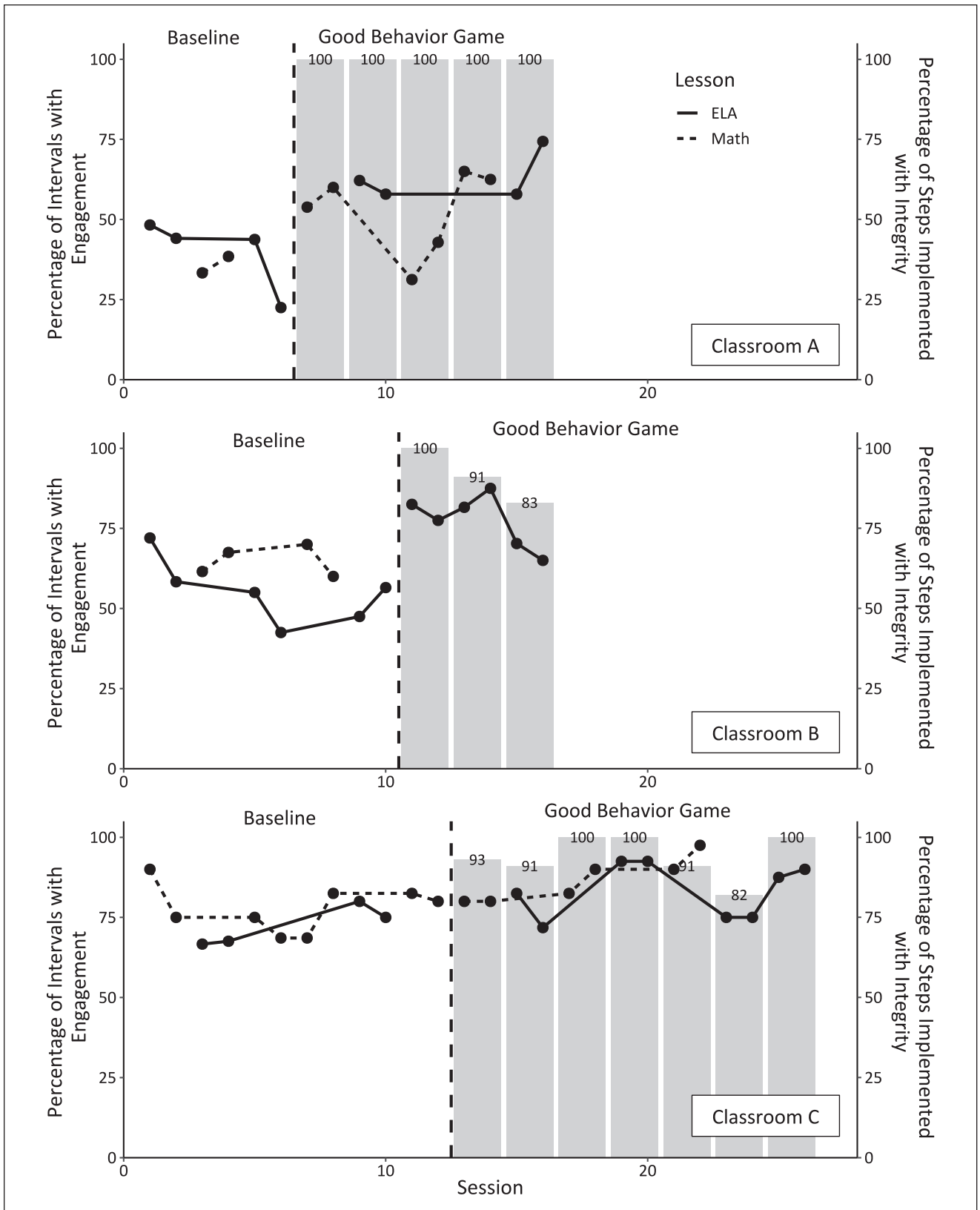


Figure 1. Academic Engagement and Treatment Integrity Across Phases.

Note. Academic engagement is plotted as a line graph. Treatment integrity is represented as a bar graph. ELA = English Language Arts.

integrity, with the highest levels of engagement occurring during sessions when integrity was highest.

In math, during the baseline phase, Classroom C demonstrated high ($M = 77.78\%$), moderately variable ($SD = 7.41$) levels of engagement with a slightly decreasing trend. Compared to the mean of baseline sessions, levels of engagement ($M = 86.67\%$) gradually increased by 11.43% during the GBG phase and variability slightly decreased ($SD = 7.01$). Unlike during ELA, academic engagement did not neatly mirror treatment integrity in math.

A vertical analysis of the MBD panels indicates that as the GBG was introduced in each classroom, academic engagement remained relatively stable for those classrooms in the baseline phase. Taken together, the within-classroom and vertical analyses indicate there were two clear demonstrations of a GBG effect; however, the smaller and delayed effects observed in Classroom C limit evidence of experimental control.

Effect Size. Using a web-based calculator (Tarlow, 2016), we supplemented our visual analysis by calculating an effect size for each classroom using guidelines offered by Tarlow (2017). Given that the GBG was only implemented during ELA in Classroom B, we aggregated the data across instructional subjects for each effect size calculation. None of the classrooms evidenced statistically significant monotonic baseline trends (all p values $>.05$); therefore, we computed Kendall's Tau rank-order correlation coefficients without correcting for baseline trends (Tarlow, 2017). The results suggested there were moderate to strong, statistically significant improvements in academic engagement from baseline to the GBG phase in each classroom (see Table 3).

Teachers' Perceived Usability of the GBG for Remote Instruction

Teachers' responses on the URP-IR indicated high perceptions of usability of the GBG in a remote instruction setting (see Table 2). Teachers reported that they found the GBG acceptable and feasible to implement in the remote setting, that they understood the GBG, and that the system climate of their school was favorable for the GBG. The teachers did not view the GBG as needing high home-school collaboration or requiring additional system support.

Students' Perceived Acceptability of the GBG in Online Classrooms

Overall, as reported on the KIP, students across all remote classrooms considered the intervention to be acceptable based upon Eckert et al.'s (2017) acceptability threshold of a total score greater than 24 (see Table 2). Students in

Classroom C considered the intervention most acceptable, with students in Classroom A reporting the greatest skill improvement.

Discussion

The purpose of this study was to examine the effects of the GBG, when implemented in an online classroom setting during the COVID-19 pandemic, on elementary students' academic engagement. The secondary purpose was to determine whether participating teachers and students found the intervention usable and acceptable, respectively. In addition, we collected descriptive data on treatment fidelity and teachers' perceptions of self-efficacy in classroom management and changes in their students' engagement with the switch to remote learning.

Effects of GBG on Students' Academic Engagement

In this study, the GBG was associated with moderate to strong, statistically significant improvements in students' academic engagement in their online classrooms (Tau = 0.44–0.61). The Tau coefficients observed in this study were smaller than the average TauU in (a) a meta-analysis across GBG studies with various target behaviors (0.82; Bowman-Perrott et al., 2016) and (b) a recent study that also targeted academic engagement (0.80; Fallon et al., 2020). These differences in effect sizes may be due to the limitations associated with the TauU metric (e.g., inflated magnitude estimates; Tarlow, 2017) that we attempted to control in this study. The size of the effects in this study are notable given that targeting desirable behaviors, such as academic engagement, during the GBG typically produces a smaller effect size (0.59) than targeting undesirable behaviors (0.81; Bowman-Perrott et al., 2016). It is possible that the additional procedural variations we incorporated, such as the preference assessment to inform reward selection and the increased frequency of reinforcement (i.e., both daily and weekly rewards), may have contributed to the effect sizes we observed (Bowman-Perrott et al., 2016).

Although engagement increased in all classrooms during the GBG phase, the changes in level and trend were far less pronounced in Classroom C, contributing to more limited experimental control. Several possible reasons may account for this difference. First, Classroom C had relatively high baseline levels of engagement, which, in past GBG research, was associated with similarly small improvements (Collier-Meek et al., 2017; Murphy et al., 2020). Second, because we only measured praise when delivered with points during the GBG phase, it is possible that Teacher C had high baseline praise rates, particularly given her higher self-efficacy in remote classroom management per the CMSEI. Third,

different instructional strategies used during math (e.g., increased opportunities to respond) may have competed with the GBG to maintain academic engagement in that subject area. Although GBG studies often do not include data on teachers' rates of praise or opportunities to respond (e.g., Fallon et al., 2020; Joslyn et al., 2014), future researchers may consider monitoring these variables across study phases to gain a clearer understanding of treatment differentiation.

By overlaying the treatment integrity data on the time-series graph of academic engagement (see Figure 1), we were able to observe the relationship between integrity and student behavior. In Classrooms B and C, engagement levels were highest during sessions with higher treatment integrity. More specifically, in Classroom B, the data show a decreasing trend in both engagement and treatment integrity across intervention sessions. In Classroom C, the lowest levels of engagement were observed during intervention sessions with the lowest treatment integrity, particularly during ELA lessons. Importantly, Classroom C was the only classroom that did not demonstrate an immediate increase in academic engagement upon GBG implementation; this aligned with the treatment integrity data, as Teacher C was the only teacher who (a) did not introduce the GBG to students with 100% integrity (see Table 1) and (b) did not implement the first GBG session with 100% integrity. That association suggests future experimental research should examine whether adherence to all steps of the GBG is particularly important when initially introducing the game to students.

The lack of variability in Classroom A's treatment integrity makes it difficult to determine the association between integrity and engagement in that classroom. Interestingly, though, the dip in engagement during the third GBG implementation (sessions 11 and 12; see Figure 1) corresponded with lower praise quality per observers' anecdotal reports, despite adhering to all GBG steps. This highlights the importance of considering different aspects of treatment integrity, such as implementation quality—a construct our field is in the early stages of operationalizing for direct measurement (e.g., Sanetti et al., 2018).

Teachers' Perceived Usability of the GBG for Remote Instruction

The URP-IR results provide preliminary evidence that teachers may find the GBG usable to implement remotely. Compared to previous research examining the usability of the GBG in person (Collier-Meek et al., 2017; Fallon et al., 2018, 2020), our sample of teachers provided slightly higher URP-IR ratings for intervention acceptability and understanding, and similar ratings of feasibility. These differences warrant further exploration to determine whether teachers find the GBG more usable in remote formats, or whether our results are an artifact of our sample, implementation planning and support efforts, or procedural variations.

Students' Perceived Acceptability of the GBG in Online Classrooms

Similar to prior research that suggests students generally find the GBG acceptable when participating in person (e.g., Grasley-Boy & Gage, 2022), the KIP results suggest our sample of students found the GBG to be an acceptable intervention to participate in as they engaged in online learning. Although these results are promising, the response rates for the KIP were fairly low (62.5%–82.4%; see Table 2), possibly because teachers may have asked students to complete the KIP during asynchronous instruction or when there were competing activities. We encourage additional research to further explore acceptability of the GBG in remote settings.

Limitations and Directions for Future Research

Limitations of this study must be considered. First, without clear evidence of experimental control from the third leg of the MBD, interpretations of the GBG's effect should be limited. Second, conducting direct observations remotely highlighted the challenges of researching interventions in online settings. Specifically, when students were off-screen or had their cameras off, it is possible that some were still actively engaged. However, given that the classroom teachers reported concerns during the PII about students' engagement when they were off-screen or had their cameras off, we coded those instances as disengaged. In addition to interviewing teachers, researchers may find it beneficial to conduct pilot observations across subject areas to inform their behavioral definitions. Second, due to time constraints, the number of treatment sessions was limited. Given that intervention length and duration may affect GBG outcomes (Kosiec et al., 1986), researchers should consider increasing the GBG's implementation. Third, future research should examine whether the effects of a remotely implemented GBG maintain and generalize to in-person settings. Finally, similar to other single-case research, the external validity of our findings is limited until replicated across samples, settings, and activities.

Implications for Practice

The results of this study provide initial evidence that teachers can feasibly implement the GBG remotely in their online classrooms to improve academic engagement, although clearer demonstrations of experimental control are needed. To implement the GBG remotely, adaptations need to be made to the measurement of engagement (i.e., taking the online context into account), the provision of rewards (i.e., privileges that can be accessed remotely), and the support of treatment integrity (e.g., emailed feedback). The relationship between treatment integrity and academic engagement in this study highlights the possible importance of

effective teacher training in the GBG and adherence to the intervention's steps during implementation.

Conclusion

This was the first experimental study to implement the GBG in a remote classroom setting. The results demonstrated class-wide academic engagement clearly and immediately increased in two of three elementary online classrooms when the GBG was introduced, and teachers and students found the intervention to be usable and acceptable, respectively. These results highlight the versatility of the GBG and indicate that it may be a promising intervention option to increase academic engagement when virtual schooling is a necessity, although replications with stronger demonstrations of experimental control are needed. Future research should seek to extend these results by examining virtual implementation of the GBG over longer time frames and for samples of varying demographics and grade levels.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Emily C. Helminen  <https://orcid.org/0000-0002-3884-9603>

Notes

1. All teacher training materials and intervention materials developed for this study can be accessed from <https://bit.ly/3Mfv4dr> (Hier et al., 2023). Others are welcome to use or adapt these materials.
2. We encouraged teachers to adjust the point criterion to keep it challenging yet attainable; if teams were easily meeting the criterion, teachers could increase it the following session.

References

- Aguilera-Hermida, A. P. (2020). College students' use and acceptance of emergency online learning due to COVID-19. *International Journal of Educational Research Open*, 1, Article 100011. <https://doi.org/10.1016/j.ijedro.2020.100011>
- Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education*, 70(2), 87–107. <https://doi.org/10.2307/2673158>
- 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(2), 119–124. <https://doi.org/10.1901/jaba.1969.2-119>
- Beaman, R., & Wheldall, K. (2000). Teachers' use of approval and disapproval in the classroom. *Educational Psychology*, 20(4), 431–446. <https://doi.org/10.1080/713663753>
- 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(3), 180–190. <https://doi.org/10.1177/1098300715592355>
- Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the usage rating profile-intervention. *Journal of School Psychology*, 51(1), 81–96. <https://doi.org/10.1016/j.jsp.2012.08.006>
- Collier-Meek, M. A., Fallon, L. M., & DeFouw, E. R. (2017). Toward feasible implementation support: E-mailed prompts to promote teachers' treatment integrity. *School Psychology Review*, 46(4), 379–394. <https://doi.org/10.17105/SPR-2017-0028.V46-4>
- Dhawan, S. (2020). Online learning: A panacea in the time of Covid-19 crisis. *Journal of Educational Technology Systems*, 49(1), 5–22. <https://doi.org/10.1177/0047239520934018>
- Eckert, T. L., Hier, B. O., Hamsho, N. F., & Malandrino, R. D. (2017). Assessing children's perceptions of academic interventions: The Kids Intervention Profile. *School Psychology Quarterly*, 32(2), 268–281. <https://doi.org/10.1037/spq0000200>
- Fahmie, T. A., & Hanley, G. P. (2008). Progressing toward data intimacy: A review of within-session data analysis. *Journal of Applied Behavior Analysis*, 41(3), 319–331. <https://doi.org/10.1901/jaba.2008.41-319>
- Fallon, L. M., Collier-Meek, M. A., Kurtz, K. D., & DeFouw, E. R. (2018). Emailed implementation supports to promote treatment integrity: Comparing the effectiveness and acceptability of prompts and performance feedback. *Journal of School Psychology*, 68, 113–128. <https://doi.org/10.1016/j.jsp.2018.03.001>
- Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M. H., & Johnson, A. H. (2015). Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children*, 81(2), 227–246. <https://doi.org/10.1177/0014402914551738>
- Fallon, L. M., Marcotte, A. M., & Ferron, J. M. (2020). Measuring academic output during the Good Behavior Game: A single case design study. *Journal of Positive Behavior Interventions*, 22(4), 246–258. <https://doi.org/10.1177/2F1098300719872778>
- Fishbein, J. E., & Wasik, B. H. (1981). Effect of the Good Behavior Game on disruptive library behavior. *Journal of Applied Behavior Analysis*, 14(1), 89–93. <https://doi.org/10.1901/jaba.1981.14-89>
- Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis*, 25, 491–498. <https://doi.org/10.1901/jaba.1992.25-491>
- Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, 95(1), 148–162. <https://doi.org/10.1037/0022-0663.95.1.148>
- Grasley-Boy, N. M., & Gage, N. A. (2022). The Lunchroom Behavior Game: A single-case design conceptual replication.

- Journal of Behavioral Education*, 31, 595–613. <https://doi.org/10.1007/s10864-020-09415-0>
- Hier, B. O., MacKenzie, C. K., Ash, T. L., Maguire, S. C., Nelson, K. A., Helminen, E. C., Watts, E. A., Matsuba, E. S. M., Masters, E. C., Finelli, C. C., Circe, J. J., Hitchings, T. J., Goldstein, A. R., & Sullivan, W. E. (2023). *Good Behavior Game in remote classrooms during COVID-19: Teacher training and intervention materials*. <https://doi.org/10.17605/OSF.IO/UR75Y>
- Jia, C., Hew, K. F., Bai, S., & Huang, W. (2022). Adaptation of a conventional flipped course to an online flipped format during the COVID-19 pandemic: Student learning performance and engagement. *Journal of Research on Technology in Education*, 54, 1–21. <https://doi.org/10.1080/15391523.2020.1847220>
- Joslyn, P. R., Austin, J. L., Donaldson, J. M., & Vollmer, T. R. (2020). A practitioner's guide to the Good Behavior Game. *Behavior Analysis: Research and Practice*, 20(4), 219–235. <https://doi.org/10.1037/bar0000199>
- 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(3), 1673–1682. [https://doi.org/10.1016/S2007-4719\(14\)70973-1](https://doi.org/10.1016/S2007-4719(14)70973-1)
- Klein, P. D., Casola, M., Dombroski, J. D., Giese, C., Sha, K. W.-Y., & Thompson, S. C. (2022). Response to intervention in virtual classrooms with beginning writers. *Reading & Writing Quarterly: Overcoming Learning Difficulties*. Advance online publication. <https://doi.org/10.1080/10573569.2022.2131662>
- Kosiec, L. E., Czernicki, M. R., & McLaughlin, T. F. (1986). The Good Behavior Game: A replication with consumer satisfaction in two regular elementary school classrooms. *Techniques*, 2(1), 15–23.
- 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*. What Works Clearinghouse. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_scd.pdf
- Lastrapes, R. E., Fritz, J. N., & Casper-Teague, L. (2018). Effects of the teacher versus students game on teacher praise and student behavior. *Journal of Behavioral Education*, 27(4), 419–434. <https://doi.org/10.1007/s10864-018-9306-y>
- McPherson, M. S., & Bacow, L. S. (2015). Online higher education: Beyond the hype cycle. *Journal of Economic Perspectives*, 29(4), 135–154. <https://doi.org/10.1257/jep.29.4.135>
- 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(2), 191–207. <https://doi.org/10.17105/spr-14-0063.1>
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Child Psychology Review*, 11(3), 247–266. [https://doi.org/10.1016/0272-7358\(91\)90103-2](https://doi.org/10.1016/0272-7358(91)90103-2)
- Murphy, J. M., Hawkins, R. O., & Nabors, L. (2020). Combining social skills instruction and the Good Behavior Game to support students with emotional and behavioral disorders. *Contemporary School Psychology*, 24(2), 228–238. <https://doi.org/10.1007/s40688-019-00226-3>
- Neugebauer, S. R., Chafouleas, S. M., Coyne, M. D., McCoach, D. B., & Briesch, A. M. (2016). Exploring an ecological model of perceived usability within a multi-tiered vocabulary intervention. *Assessment for Effective Intervention*, 41(3), 155–171. <https://doi.org/10.1177/1534508415619732>
- New York State Education Department. (2019). *New York State education at a glance*. <https://data.nysed.gov>
- Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., Koenig, J. L., Resetar, J. L., & Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, 34(1), 87–106. <https://doi.org/10.1080/02796015.2005.12086277>
- Novianti, R., & Garzia, M. (2020). Parental engagement in children's online learning during COVID-19 pandemic. *Journal of Teaching and Learning in Elementary Education*, 3(2), 117–131. <https://doi.org/10.33578/jtlee.v3i2.7845>
- Rathel, J. M., Drasgow, E., Brown, W. H., & Marshall, K. J. (2014). Increasing induction-level teachers' positive-to-negative communication ratio and use of behavior-specific praise through e-mailed performance feedback and its effect on students' task engagement. *Journal of Positive Behavior Interventions*, 16, 219–233. <https://doi.org/10.1177/1098300713492856>
- Reed, D. D., & Azulay, R. L. (2011). A Microsoft Excel® 2010 based tool for calculating interobserver agreement. *Behavior Analysis in Practice*, 4(2), 45–52. <https://doi.org/10.1007/BF03391783>
- Royer, D. J., Lane, K. L., Dunlap, K. D., & Ennis, R. P. (2019). A systematic review of teacher-delivered behavior-specific praise on K-12 student performance. *Remedial and Special Education*, 40(2), 112–128. <https://doi.org/10.1177/0741932517751054>
- Sallese, M. R., & Vannest, K. J. (2020). The effects of a multicomponent self-monitoring intervention on the rates of pre-service teacher behavior-specific praise in a masked single-case experimental design. *Journal of Positive Behavior Interventions*, 22(4), 207–219. <https://doi.org/10.1177/1098300720908005>
- Sanetti, L. M. H., Williamson, K. M., Long, A. C. J., & Kratochwill, T. R. (2018). Increasing in-service teacher implementation of classroom management practices through consultation, implementation planning, and participant modeling. *Journal of Positive Behavior Interventions*, 20(1), 43–59. <https://doi.org/10.1177/1098300717722357>
- Slater, E. V., & Main, S. (2020). A measure of classroom management: Validation of a pre-service teacher self-efficacy scale. *Journal of Education for Teaching*, 46(5), 616–630. <https://doi.org/10.1080/02607476.2020.1770579>
- Spilt, J. L., Leflot, G., Onghena, P., & Colpin, H. (2016). Use of praise and reprimands as critical ingredients of teacher behavior management: Effects on children's development in the context of a teacher-mediated classroom intervention. *Prevention Science*, 17(6), 732–742. <https://doi.org/10.1007/s1121-016-0667-y>
- Tarlow, K. R. (2016). *Baseline corrected Tau calculator*. <http://www.ktarlow.com/stats/tau>
- Tarlow, K. R. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline Corrected Tau. *Behavior Modification*, 41(4), 427–467. <https://doi.org/10.1177/0145445516676750>
- Witt, J. C., & Elliott, S. N. (1983). Assessment in behavioral consultation: The initial interview. *School Psychology Review*, 12(1), 42–49. <https://doi.org/10.1080/02796015.1983.12085007>
- Zaccoletti, S., Camacho, A., Correia, N., Aguiar, C., Mason, L., Alves, R. A., & Daniel, J. R. (2020). Parents' perceptions of student academic motivation during the COVID-19 lockdown: A cross-country comparison. *Frontiers in Psychology*, 11, Article 592670. <https://doi.org/10.3389/fpsyg.2020.592670>