ORIGINAL PAPER

Leveraging Technology to Support Teachers' Fidelity of Universal Classroom Management Interventions: Lessons Learned and Future Applications

Stephanie D. Smith[1] · Fayth Walbridge[1] · Tiffany Harris[1] · Mairin C. Cotter[1] · Rachel Kaplan[1] · Brittany Garza[1] · Zachary Wilde[1] · Arianna Delgadillo[1] · Richard Mohn[1] · Brad Dufrene[1]

Abstract

The Good Behavior Game (GBG), a universal classroom management intervention, has shown clear benefits in promoting the behavioral, social-emotional, and academic development of students. However, the quality with which this intervention is delivered tends to diminish over time, which decreases the likelihood of these positive outcomes. By leveraging the benefits of technology, we built a sophisticated online platform to support teachers' fidelity of the GBG in collaboration with expert consultants and education partners. This paper details initial steps to develop and refine GBG Technology (GBG Tech). Three teacher consultants and two experts in technology-enhanced and classroom management interventions provided ongoing feedback as GBG Tech was initially developed through a rapid prototyping approach by a team of high-tech engineers. Twenty-four teachers participated in focus groups to inform subsequent revisions of the technology, and seven teachers tested the feasibility of GBG Tech in their classrooms for 6 weeks. As anticipated, teachers found GBG Tech to be acceptable, understandable, and feasible to use. Moreover, teachers reached fidelity quickly ($M = 2.43$ weeks), sustained fidelity for 6 weeks, and delivered the GBG at the recommended dosage. The results of this study informed a full version of GBG Tech that is ready for large-scale testing and a set of design principles intended to guide the development of other technology-delivered interventions aimed at sustaining fidelity in authentic classroom settings.

Keywords Technology development · Classroom management interventions · Fidelity · Good Behavior Game

Leveraging Technology to Support Teachers' Fidelity of Universal Classroom Management Interventions: Lessons Learned and Future Applications

Disruptive and off-task behaviors in the classroom negatively impact student behaviors and academic outcomes, as learning is often interrupted and a suboptimal precedent is set for how students interact with their teacher and peers (Gaastra et al., 2016; Lewis, 2001). If implemented with high fidelity, Tier 1 universal classroom management interventions are highly effective in reducing disruptive behaviors (e.g., aggression, non-compliance, hyperactivity), promoting prosocial behaviors (e.g., helping others, sharing, cooperating), and establishing environments conducive to learning (Greenberg & Abenavoli, 2017). However, given the numerous demands placed on teachers, prior work has shown that most teachers are unable to sustain the level of fidelity required to obtain these positive student outcomes without support (e.g., coaching; Han & Weiss, 2005). As such, our research team has leveraged the benefits of technology to develop a sophisticated online platform to address implementation drift by supporting teachers in the delivery of the most thoroughly studied classroom-based intervention, the

✉ Stephanie D. Smith
   SD.Smith@usm.edu

1   Department of Psychology, School of Psychology, University of Southern Mississippi, 118 College Drive, #5025, Hattiesburg, MS 39406, USA

Good Behavior Game (GBG). The purpose of this paper is to describe the development of GBG Technology (GBG Tech) using an iterative, multiple method design approach in partnership with education partners (i.e., teachers, students); provide initial evidence of its usability, feasibility, acceptability, and sustained fidelity; and to offer guidance on how to develop future technology-delivered interventions with the goal of sustaining fidelity in classroom settings.

Description of and Evidence for the GBG

The Good Behavior Game (GBG) is an interdependent group contingency where students are placed on teams and work together to follow the rules of the classroom (Barrish et al., 1969). Students are immediately rewarded for their efforts if they engage in more rule-following behaviors than rule violating behaviors during a game period. Games are played while students complete a curriculum-based activity (e.g., reading or math assignment) so as not to interfere with classroom instruction. Numerous studies have evaluated the influence of the GBG on proximal student outcomes. The GBG has been shown to decrease aggressive, inattentive, and other disruptive behaviors; increase prosocial behaviors; and improve student achievement after 1 to 2 years of its implementation in the early elementary school years (Dion et al., 2011; Dolan et al., 1993; Ialongo et al., 1999; Leflot et al., 2010; Witvliet et al., 2009). A recent meta-analysis of randomized controlled trials evaluating the impact of the GBG on proximal student outcomes revealed small to medium-sized treatment effects on aggression/conduct problems, inattention, shy/withdrawn behavior, and reading comprehension (Smith et al., 2021). As a preventive intervention, there is also strong evidence to suggest that the GBG decreases the risk of conduct problems and tobacco use in late childhood and early adolescence (Furr-Holden et al., 2004; Huizink et al., 2008; Ialongo et al., 2001; van Lier et al., 2009). Additionally, there is some support suggesting the GBG prevents the emergence of substance abuse, internalizing symptoms, problematic peer relations, poor parenting practices, academic underachievement, special education placement, and school suspensions in late childhood and early adolescence (Bradshaw et al., 2009; Furr-Holden et al., 2004; Ialongo et al., 2001; Vuijk et al., 2007; van Lier et al., 2005) as well as criminal behavior, substance misuse/dependence, suicidal behaviors, risky sexual behavior, health service use, and the non-pursuit of a college degree in emerging and early adulthood (Bradshaw et al., 2009; Kellam et al., 2008; Wilcox et al., 2008). However, the GBG, like other evidence-based practices, is susceptible to implementation drift (Domitrovich et al., 2015; Ialongo et al., 1999) and poorer fidelity of the GBG has been shown to diminish its positive influence on student outcomes (Becker et al., 2013).

Factors Influencing Fidelity

Several factors have been identified as impeding or promoting the sustained fidelity of school-based interventions (see Century et al., 2010; Combs et al., 2022; Han & Weiss, 2005). Although these factors span multiple levels of system supports inclusive of macro-level factors (e.g., federal, state, or district policies), school-level factors (e.g., school characteristics or climate), and teacher-level factors (e.g., teacher beliefs or attitudes; Domitrovich et al., 2008), our focus is on malleable factors specific to teachers and interventions that are likely to have a technology-based solution. Such *teacher-level factors* include self-efficacy, professional burn-out, treatment acceptability, time constraints, proficiency in using interactive techniques, and adapting or not fully implementing treatment protocols (Dane & Schneider, 1998; Han & Weiss, 2005; Payne et al., 2006; Ringwalt et al., 2004). Sustained treatment fidelity is also more likely to be achieved if the following *intervention-level factors* are present: intervention is packaged simply with clear instructions, is easy to administer, and essential and non-essential components are clearly identifiable in the treatment manual (Dusenbury, et al., 2003; Payne, et al., 2006). Han and Weiss (2005) have suggested that enhancing initial teacher training and offering continual feedback concerning the delivery of the intervention through classroom consultants or coaches is one solution to address the sustained fidelity problem. However, it is unclear how much training is necessary for teachers to internalize what they learn or if these gains are maintained once added supports are removed. Another approach to overcome fidelity issues and lessen the burden on teachers is to harness the advancements made in technology and make use of resources typically available in the classroom setting (e.g., Vogl et al., 2012). Using technology as a delivery system of evidence-based practices has several advantages over traditional means of delivery in that: (1) the content and delivery of the program may be controlled; (2) the interactive nature of delivery may be consistently maintained; and (3) a greater level of support may be offered to those supervising or delivering the treatment. Moreover, it has the potential to be more user-friendly, cost-effective, and go to scale more quickly without compromising fidelity.

Technology-Based Supports

There are some notable and promising efforts to support teachers' implementation of behavior management practices in the classroom through technology. Such technology has been designed to enhance professional development (e.g., simulation training; Pas et al., 2016; Shernoff et al., 2022) and provide implementation supports (e.g., online behavior tracking platforms; Owens et al., 2022).

Results have been encouraging as teachers' use of evidence-based practices in the classroom have been shown to significantly increase with subsequent improvements in students' behaviors when teachers receive technology-assisted training or implementation supports (Shernoff et al., 2021; Owens et al., 2022). However, the use of simulation technology by teachers to practice behavior management strategies has been reported to be much lower than what is recommended (Shernoff et al., 2021, 2022), and if not continued to be used outside of school hours by teachers, it is unclear how that will impact the fidelity with which these strategies are implemented over time. Relatedly, a study examining the implementation of an online Daily Report Card platform revealed that several aspects of the intervention were used by teachers according to evidence-based guidelines (e.g., screening behaviors, setting goals) whereas other aspects were underutilized (e.g., shaping behaviors). Notably, teachers implemented the intervention by entering data on only 63% of the school days and 60% of teachers opted to use paper methods rather than the online platform to track behaviors. Although most elements of this intervention were built into the technology, there were no embedded features designed to prevent treatment adaptations or implementation errors (Owens et al., 2022).

Regarding technology-based supports for the GBG, an entirely online version of GBG training has been developed and an evaluation of web-mediated coaching for the GBG is currently underway. The GBG online training has been found to be comparable to in-person training with respect to fidelity; however, both conditions received ongoing coaching and feedback to sustain fidelity over time (Becker, et al., 2014; Poduska & Kurki, 2014). Web-mediated coaching may prove to be comparable to in-person coaching in achieving high-quality implementation; however, questions remain as to whether an adequate level of fidelity is maintained once web-mediated coaching is no longer available, what dosage of web-mediated coaching is required for teacher internalization, or if web-mediated coaching is feasible or less time consuming relative to in-person coaching given teachers must consistently provide coaches with recordings of their GBG delivery.

There are other web-based tracking and reporting systems (e.g., ClassDojo) teachers may use to award or take away points for student behaviors. For example, ClassDojo (www.classdojo.com) is a tool to assist teachers in tracking behaviors (i.e., behavior management chart) in the context of a behavior management intervention (e.g., token economy). However, the developers of ClassDojo offer no evidence-based behavior management framework to guide teachers on how to use this technology, so it may be used by teachers in a countless number of ways. More recently, ClassDojo was evaluated to support teachers' delivery of the GBG by using it to award points and track team progress. Using a single-case design with a sample of 3 teachers, the GBG with ClassDojo was found to significantly reduce disruptive behavior, increase on-task behavior, and was considered moderately to highly acceptable by teachers (Lynne, et al., 2017). At this time, we are unaware of any evaluations of technology that delivers the entirety of the GBG in real-time and has built-in features specifically designed to prevent adaptations, implementation errors, or promote GBG implementation at the recommended dosage.

Technology as a Delivery System for the GBG

The GBG is an ideal candidate for a technology-based solution to address fidelity challenges because it has well-defined and distinct treatment components, is time-limited with respect to its delivery, and is easily integrated into the curriculum, as implementation occurs during classroom activities. When considering what initial features should be required of a technology-delivered version of the GBG (GBG Tech), our research team drew from the relevant literature highlighting teacher-level (i.e., self-efficacy, professional burn-out, treatment acceptability, treatment quality/adherence) and intervention-level (i.e., ease of administration, discernability of treatment components; clarity of instructions) factors that are known to influence fidelity. We also reflected upon implementation oversights (e.g., infrequent game play, non-delivery of rewards, focus on disruptive behaviors) our research team and expert consultants observed when rating the fidelity of teachers' implementation of the GBG in the context of other efficacy trials (e.g., Integrated Brain, Body, and Social Intervention; Smith et al., 2020; GBG Baltimore Studies; Ialongo et al., 1999). The implementation factors (comprising both teacher and intervention-level factors) that were deemed both malleable and had a technology-based solution were prioritized for inclusion in the online platform.

Present Study

To our knowledge, no attempts have been made to develop technology that not only incorporates all treatment components of a universal classroom management intervention to support its delivery by teachers but also has built-in features that prevent adaptations and minimize implementation errors to promote sustained fidelity, so the beneficial treatment effects of the intervention are achieved and maintained in the classroom. The susceptibility of interventions to implementation drift in authentic educational settings is well documented and it requires a solution that targets factors known to influence the fidelity and sustainability of teacher implementation. Drawing from the research literature, we identified intervention-level and teacher-level factors that helped inform features to

include in the initial specifications of the technology. Further, our research team followed a collaborative research model (Diamond & Powell, 2011; Stein et al., 2002) by taking an iterative, multiple method approach to initially develop and refine GBG Tech while involving teachers at every stage of the development process. The purpose of this paper is to detail the research activities completed in Years 1 and 2 of a 3-year project that was funded as a development and innovation grant (Goal 2) awarded by the Institute of Education Sciences (IES). In these first two years, we had the following research aims: 1) develop technology designed to minimize implementation drift of the GBG when launched in authentic educational settings (Year 1 Development Phase); 2) refine and feasibility test the technology (Year 2 Refinement Phase); and 3) devise a set of principles outlining how to apply technology to other evidence-based interventions to promote their fidelity and sustainability in school settings. The Year 3 pilot study comparing GBG Tech to standard GBG delivery to determine its promise as an approach to maintain high-quality teacher implementation and subsequently improve student outcomes will be reported elsewhere. See Fig. 1 for a graphical depiction of each phase of this 3-year project.

The primary research aim in the Year 1 Development Phase was to develop a beta version of GBG Tech where all core components of the GBG were built into the technology and all features theorized to prevent implementation drift were fully operational. A rapid prototyping approach was adopted to allow for continual feedback from teacher consultants and experts in school-based and technology-enhanced interventions, so changes could be made quickly before actual coding began by a team of high-tech engineers. The Year 2 Refinement Phase was devoted to further development and refinement of this technology by obtaining teachers' feedback via focus groups and conducting feasibility testing by teachers in the classroom. The primary research aim for the Year 2 focus groups was to learn how to improve GBG Tech's acceptability, usability, feasibility, and fidelity through the solicitation of teachers' impressions about its core features, user interface, and design, which would subsequently inform changes to the technology. Our research aims specific to Year 2 feasibility testing were as follows: (1) the length of time it took teachers to implement GBG Tech with fidelity, (2) whether it was feasible for teachers to use GBG Tech at least once a day in their classrooms, and (3) if teachers continued to find GBG Tech feasible, usable, and acceptable after using the technology in their classrooms for 6 weeks. Importantly, each stage of this process informed design principles to guide the development of future technology-delivered interventions aimed at sustaining fidelity.

Method

Year 1 Development Phase

Teacher and Expert Consultants

In the Year 1 Development Phase, we contacted district superintendents and principals with whom our school and clinical psychology doctoral programs had long-standing relationships including ongoing contracts that allowed for the provision of services by our graduate students. After receiving approval from school administrators, three teachers were recruited through email announcements that were sent by their school principals and originated from our research team notifying them of part-time, paid consultant positions. School districts included in this recruitment effort were in the southeastern region of the USA and within a 30-mile radius of the university with which our research team is affiliated. These teachers were female (67% White, 33% Black) from regular education, core curriculum classrooms and taught students in grades 1 through 4, which is consistent with the characteristics of teachers participating in GBG efficacy studies (Ialongo et al., 1999; Leflot et al., 2010). They also had experience delivering the GBG or behavior management strategies with technology (e.g., ClassDojo) in the classroom. As they were designated as consultants and their role was to provide verbal and immediate feedback on prototypes built by the technology development firm, they were not considered research participants. Their students, who were also not considered research participants, voted on certain aspects of the technology and this
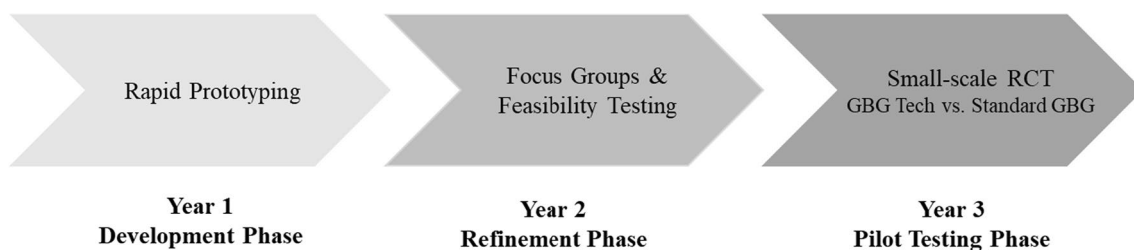

Fig. 1 Three phased approach to GBG Tech development

information was relayed to the research team via encrypted email or virtual meetings held with the teacher consultants. No personal identifying information about the students was shared with the research team, and only aggregate data were provided (i.e., number of votes counted by their teachers).

The two expert consultants of school-based and technology-enhanced interventions were researchers at prominent institutions of higher education located in the northeast and northwest regions of the USA. They were former mentors and colleagues of the Principal Investigator (PI). The expert consultant of school-based interventions has directed or been a key investigator of large-scale clinical trials of the GBG in school settings and has worked extensively in the prevention of mental health problems among children and adolescents for over 25 years. The expert consultant of technology-enhanced interventions has overseen the development of multiple applications designed to improve the quantity and quality of communication skills in children with Autism Spectrum Disorders.

Procedures

The foundation of GBG Tech was based on detailed specifications of the web application written by the PI, which described the flow and design of each page of the application; how user data would be captured; and what features were needed to address common implementation oversights. These specifications were written with guidance from the relevant research literature and from observations made by the PI when coaching teachers in their delivery of the GBG in the classroom. To prevent major modifications and adaptations of the treatment protocol, preserve the interactive nature of its delivery, and minimize the demands placed on teachers, it was determined that all treatment components of the GBG would be built into the technology, which was expected to target implementation factors considering many aspects of the treatment would be done automatically by GBG Tech (e.g., displaying classroom rules, tracking team performance at the individual and team level, determining team winners, delivering immediate rewards following each game). Given the greater level of support offered to teachers delivering the GBG with technology, a positive impact on perceived professional burn-out was anticipated. Implementation safeguards such as email notifications and corrective messaging were also considered necessary to sustain fidelity (i.e., treatment quality/adherence) and would involve reminding teachers to play the game daily, deliver rewards to team winners, shift their attention to prosocial behaviors (rather than focus on disruptive behaviors), and redistribute team membership if one team was continually winning the game. To ensure students receive a treatment dosage that is necessary for change, a built-in titration schedule was also deemed essential where the length of each game would be pre-determined and dependent upon the number of minutes previously played in the classroom. Finally, student performance data collected by the technology would be downloadable and visually displayed, so teachers could witness the effectiveness of the intervention easily and quickly, thus positively influencing their self-efficacy, which is consistent with the "self-sustaining feedback loop" outlined in the conceptual model by Han

Table 1 Design principles, GBG Tech features, and targeted implementation factors

Design principles	GBG Tech features	Targeted implementation factors
Built-in Treatment Components	Displays classroom rules Determines initial team membership (data-driven) Timer clearly specifies start/end of game Shows team membership Displays team progress Notifies teams of specific behaviors Immediately & appropriately addresses behaviors Reviews team performance Rewards winning team	Implementation Quality Professional Burn-Out
Implementation Safeguards: Email Notifications Corrective Messaging Blocked Game Flow	Prompts daily use Identifies weekly team winners & suggests rewards Proposes team redistributions based on wins/losses Suggests when to shift attention to prosocial behaviors (based on teachers' entry of observed behaviors)	Implementation Quality Adherence/Dosage
Automatic Titration Schedule	Sets game length (based on previous game play & dosage required for change)	Adherence/Dosage
Outcomes Tracked Over Time	Records & saves specific prosocial & disruptive behaviors of students	Implementation Quality Acceptability
Progress Reports	Behaviors graphically displayed at individual & team level	Self-Efficacy Program Evaluation

and Weiss (2005). Documented student progress was also expected to augment the acceptability of the intervention and motivate teachers to continue using GBG Tech in their classrooms. In fact, both professional burn-out and acceptability have been specifically linked to the adherence of the GBG dosage protocol (Domitrovich et al., 2015). Finally, the performance data could be used for program evaluation if shared with teachers and other school personnel as an additional metric to assess students' response to school-based interventions already in place, thus increasing acceptability across education partners. Table 1 highlights the overarching design principles guiding technology development, GBG Tech features supporting teacher implementation, and their targeted implementation factors.

To develop the initial version of the technology, we followed a rapid prototyping approach where every feature of GBG Tech was designed by a team of high-tech engineers with continual feedback from experts in the field of school-based prevention (including the GBG) and technology-enhanced interventions (e.g., mobile applications, social robots). Each cycle of deliverables from the engineering team was available every 3 to 5 days followed by 1 or 2 days of review by our research team. The engineering team used a modular software prototyping tool to build a model of the prototype where only the major elements of the software were presented in a schematic way. These clickable prototypes were developed quickly before actual coding began, so our research team could interact and understand the implications of the software's design, architecture, navigation, categorization, and interaction, which allowed for the testing and feedback phase to begin almost immediately. Modifications to the clickable prototypes were discussed at each weekly meeting with the engineering team and the engineering team did not begin coding until they received approval for the prototype design and logic (i.e., rules of conditional behaviors) from the research team. The teacher and expert consultants joined our meetings approximately once a month after a feature that was outlined in the technology specifications (e.g., play a game, view progress reports, assign students to teams) had been completed, so they could witness the entirety of its functionality and provide feedback. Their suggestions for changes were discussed with the team and were incorporated into the prototype if it improved the user experience and maintained the functionality of the feature.

Teachers also helped facilitate feedback from their students who were asked to vote on the GBG Tech theme to identify team mascots (e.g., silly birds, robots, dinosaurs), and behavioral rewards (e.g., Simon Says, Follow the Leader, Human Knot) that students earn after teams win a game, by raising their hand for the preferred options. Students indicated their preferences for rewards by raising their hand after teachers read a description of each reward. Teachers rated the reward as "love it" if 80% or more of students raised their hands, "like it" if between 60 and 70% of students raised their hands, "just okay" if 50% of students raised their hands, "dislike it" if between 20 and 40% of students raised their hands, and "hate it" if less than 10% of students raised their hands.

The Year 1 development phase (i.e., user experience/visual design, HTML development, engineering/architecture) of GBG Tech took approximately 8 months with another 2 months of initial testing (i.e., quality assurance, bug fixes) by the engineering and research team.

Year 2 Refinement Phase

Research Participants: Focus Groups

In the Year 2 Refinement Phase, we used the same procedures to recruit teachers to participate in focus groups where email announcements were sent to teachers by their principals or posts were made on social media (i.e., Facebook) about our study to private teacher groups in our state by our research coordinator who was previously employed as a teacher and a member of these groups. To participate, teachers had to have experience delivering the GBG in their classroom or using technology to manage classroom behaviors. They also had to teach grades 1 through 4 and be from general education classrooms in public schools. A total of 24 elementary teachers participated in our eight focus groups and 3 teachers comprised each group. Although we aimed to have 6 to 8 teachers participate in each focus group, this phase of the study occurred when teachers were transitioning back to in-person instruction and some hybrid instruction was still ongoing because of the COVID-19 pandemic, so scheduling proved to be challenging, as teachers' availability was extremely limited even when conducting focus groups via Zoom. Of these 24 participating teachers, 87% were female with 57% of teachers identifying as Black, 43% of teachers identifying as White, and 4% of teachers identifying as Hispanic. Most teachers (38%) were between the ages of 31–40 years, 29% were between the ages of 41–50, 17% were between the ages of 20–30 years, and 13% were between the ages of 51–60. Teachers' experience working in the field of education ranged from 1 to 39 years ($M = 13.63$, $SD = 9.58$), and all teachers had obtained a degree in higher education (i.e., Bachelor's, Master's, Doctorate) with concentrations in Elementary Education, Mathematics, Special Education, Early Childhood Education, and English. Standard state certificates or advanced professional certificates were held by almost all (96%) of the participating teachers.

Research Participants: Feasibility Testing

To recruit teachers to participate in the initial feasibility testing of GBG Tech, superintendents from four school districts

located in the surrounding metro area of the university where this study took place had committed to partnering with our research team to recruit teachers for this project before and after securing grant funding. They reached out to principals of elementary schools in their district to set up meetings with our research team, so we could describe the purpose and procedures of the study. Principals then identified teachers who were willing to use the technology in their classrooms for approximately 6 weeks and gave us their contact information. To participate, teachers had to be unfamiliar with the GBG and technology-delivered behavior management methods. The seven teachers who participated in GBG Tech feasibility testing were from 3 elementary schools and taught grades K through 4 to students from general education classrooms. Teachers were female, primarily White (86%) with 14 years of teaching experience on average (range = 5–22 years). They had a Bachelor's or more advanced degree (Master's or Doctorate) and held standard state certificates. Teachers participating in each phase of this project did not overlap and were not recruited from the same school districts.

Procedures

Study procedures did not commence until approval was obtained from school administrators (i.e., superintendents, principals) and the University Internal Review Board (IRB) and informed consent was provided by teachers.

Focus Group Procedures

Focus groups were conducted over Zoom and co-led by two members of the research team following procedures outlined by Krueger and Casey (2009). Each focus group lasted approximately 90 min. Teachers viewed a video detailing all functionalities of the technology and were then asked a standardized set of questions soliciting feedback on the core features, user interface, and design of GBG Tech as well as their impressions about its usability, feasibility, acceptability, and fidelity. Graduate students on the research team were trained using a standardized protocol developed by our expert qualitative analytic consultant detailing how to run focus groups (e.g., when to explicitly follow focus group script, ask follow-up questions, go "off-script" to answer participants' questions). Students also observed at least one focus group conducted by the PI to reinforce this protocol. The recording and automatic transcription features of Zoom were turned on to facilitate the transcribing process of the recordings from focus groups. Teachers were given gift cards immediately following their participation in focus groups.

Feasibility Testing Procedures

Teachers who participated in feasibility testing of GBG Tech in their classrooms attended a half-day training session (~ 4 h) headed by the PI at their school. The training session made use of existing standardized GBG training materials from other GBG efficacy trials (e.g., Smith et al., 2020), which included an overview of the theoretical basis of the GBG, a review of the core elements of the GBG, and a discussion of how to implement the GBG with fidelity. This training was supplemented with a presentation highlighting the functionalities of GBG Tech; a demonstration of how to play the GBG using the technology; and an opportunity for teachers to interact with the technology to set up their virtual classrooms, play a game, and view student performance data captured by GBG Tech. Trained teachers were instructed to implement GBG Tech in their classrooms once a day for 6 weeks. Fidelity ratings were made once a week by the research team and were used to inform brief 10-min feedback sessions with teachers immediately following classroom observations to improve their delivery of the GBG. Prior to and immediately following 6 weeks of GBG implementation in their classrooms, teachers were asked to complete measures assessing their views of the usability, acceptability, and feasibility of GBG Tech as well as their impressions about the technology, which included questions covering similar content as the focus groups. Teachers were compensated for their time with gift cards after attending study trainings and completing study measures.

Measures: Feasibility Testing

In the Year 2 Refinement Phase, several quantitative measures were administered to teachers or used by the research team to assess the fidelity, usability, acceptability, and feasibility of GBG Tech.

The GBG Implementation Rubric (Schaffer et al., 2006) was used by the research team to assess the fidelity of GBG Tech. It is comprised of 7 dimensions of GBG implementation that are reflective of its core components and include: (1) preparing students for the game, (2) choice of class activity, (3) timing the game, (4) quality of teams, (5) response to behaviors, (6) delivery of rewards, and (7) reviewing team performance. Each dimension is rated on a Likert-type scale from 0 to 4 with higher scores indicating better quality implementation. A mean rating of "3.5" on the implementation rubric corresponds to "fully trained" status in the GBG research literature (Rebok et al., 1996; Storr et al., 2002). The average interrater reliability of this measure is quite high (ICC = 0.93) and correlations between the overall implementation quality score and teacher ratings of the GBG's impact on classroom behavior, ease of use, and fit with schedule and teaching philosophy have ranged from

0.81 to 0.87 (Domitrovich et al., 2015). We adapted this measure so that it assesses the fidelity of both forms of GBG treatment delivery (i.e., GBG Tech and standard GBG). Specifically, some of the qualitative descriptors used to rate each dimension of GBG implementation were not applicable to GBG Tech (e.g., "reward is drawn randomly from a group of prizes", "teacher sets timer in full view of students"). Therefore, we revised the descriptors, so it was relevant to both forms of delivery; however, the rated dimensions of GBG implementation remained unaltered. Two members of the research team independently rated GBG implementation for thirty percent of the classroom observations. If discrepancies were found across raters, these ratings were discussed by the research team until agreement was reached and a final consensus score was entered into the database.

The Teacher Perceptions of the Intervention Attributes (TPIA; Domitrovich & Ialongo, 2008) is a 5-item measure using a 5-point Likert scale (ranging from "not at all" to "a lot") to assess how well the intervention fits with teaching style ($\alpha = 0.84$; two items of acceptability) and class schedules ($\alpha = 0.84$; two items of feasibility). The last item on this measure assesses teachers' level of motivation for continued implementation of the intervention (i.e., sustainability).

The Usage Rating Profile-Intervention (URP-I; Briesch et al., 2013) was used to measure teachers' perceptions of treatment acceptability, feasibility, and understanding. The 29 items of the URP-I load onto six factors including Acceptability, Understanding, Feasibility, Family-School Collaboration, System Climate, and System Support with internal reliability estimates ranging from 0.72 to 0.95. These items are rated according to a 6-point Likert scale ranging from "strongly disagree" to "strongly agree" with higher scores reflecting stronger agreement. For this study's purposes, we made use of the Acceptability ($\alpha = 0.80$), Understanding ($\alpha = 0.78$) and Feasibility scales ($\alpha = 0.88$), as the other scales were less relevant in providing feedback on the technology.

The GBG Tech Feedback Survey asked teachers for their impressions on the core features of GBG Tech and if GBG Tech met our main goals of development, which was to support teachers in their delivery of the GBG with sustained fidelity. Finally, the number of games played by teachers and other game play data (e.g., game length, date of game play, points/tallies received, rewards delivered) were automatically captured by GBG Tech; these data were accessible only by the research team via the GBG Tech researcher portal.

Results

Year 1 Development Phase

The GBG Tech specifications, which were reviewed by the engineering team well before the kick-off meeting with the research team and consultants, helped determine which feature would be developed first and provided structure to our weekly meetings. The primary modifications made to GBG Tech that were suggested by our teacher consultants involved changing the location of clickable buttons that initiated the main features of GBG Tech (e.g., play a game, take attendance, shuffle teams) to make them more intuitive to use and decreasing the number of clicks needed to award points and tallies to students. Our expert consultants provided feedback on how to optimize GBG Tech's performance, ensure the security of captured data, and verify the inclusion of all behavior management principles of the GBG. Teachers also helped brainstorm immediate behavior rewards and team mascots to be incorporated into the technology and selected 30 rewards and 3 mascot themes that were then voted on by students. Of the 74 participating students, 47% voted for the silly birds, 30% voted for dinosaurs and 23% voted for robots. Given these results, the silly birds were adopted as the mascot theme for GBG Tech. Additionally, six rewards fell below the "just okay" range and were replaced by rewards suggested by students and teachers across classrooms. By the end of the Year 1 Development Phase, an initial version of GBG Tech was fully operational; all GBG treatment components were built into the technology, and the features designed to limit treatment adaptations and implementation errors worked as intended. Importantly, teacher consultants agreed that the initial version of GBG Tech was user-friendly or intuitive to use, facilitated GBG implementation, met their behavior management needs, and could be practically used in the classroom as discussed in their monthly meetings with the research and engineering teams. As anticipated, the feedback we received from teachers and students greatly improved the product and highlighted the need to solicit feedback from both teachers and students throughout the refinement and pilot testing phases. Similarly, we continued to seek guidance from our expert consultants on how to optimize the performance and security of GBG Tech and navigate the training and coaching of teachers during feasibility and pilot testing.

Year 2 Refinement Phase

Analytic Approach: Focus Groups

Transcripts derived from the Zoom recordings of focus groups were independently coded by two members of the

research team and discrepancies across coders were discussed until 100% consensus was reached on how to code the transcript excerpt. The PI then completed a final review of all coded data to ensure consistency across coders. A pre-determined coding scheme was applied to the data, which were derived from the standardized questions asked during focus groups that pulled for content themes corresponding to the IES definitions of usability (i.e., "how easily the user understands or learns how to use the intervention effectively and efficiently"), feasibility (i.e., "extent to which the user can deliver the intervention within the constraints of an authentic educational setting"), acceptability (i.e., "user's attitudes toward the intervention and its perceived effectiveness"), and fidelity (i.e., "how well the user delivers the intervention as intended") (Hill et al., 2023; Proctor et al., 2011). This finalized codebook can be made available upon request. Ultimately, we used a hybrid approach to content analysis where the codes were initially generated deductively but were revised if new codes emerged when the coding scheme was systematically applied to the data (Krueger & Casey, 2009). We chose this approach considering it determines the existence and frequency of concepts in selected text, so we were able to code suggested changes according to themes that would improve specific aspects of the technology (e.g., usability, acceptability) and the frequency with which these themes were mentioned by teachers guided our prioritization of changes (Hsieh & Shannon, 2005). To increase the confidence in the trustworthiness of our qualitative findings, we followed the procedures outlined by Elo and colleagues (2014). Specifically, we selected a data collection method that best met our research aims; had our interview questions pre-screened by our teacher consultants and reviewed by our qualitative data analysis expert; used purposive sampling and carefully outlined criteria that was used to select our participants; thoroughly defined our pre-determined codes and how they were created; selected a suitable unit of analysis; had two members of the research team independently code data, discuss and resolve discrepancies; had the PI complete a final review of the coded data; ensured data were well saturated by observing replication across categories; and presented representative quotations from the transcribed text in table format.

A hierarchical decision-making process was used to determine what changes should be adopted and when they should be implemented. Specifically, changes were prioritized according to the following criteria: (1) change was directly tied to intervention-level and teacher-level factors that would lead to high implementation quality and continued use of the technology; (2) change was both time and cost-effective (i.e., has the greatest impact with the least amount of labor); (3) the frequency with which the change was mentioned across teachers and focus groups; and (4) change was not already planned for in future iterations of GBG Tech.

Results: Focus Groups

Seventy-one percent of qualitative data were coded according to our pre-determined themes inclusive of usability, feasibility, acceptability, and fidelity. Two additional themes emerged when applying the coding scheme to the data. Specifically, teachers gave feedback concerning the visual design of the technology (i.e., design elements such as colors, shapes, layout, and typefaces; the behavior of dynamic elements such as buttons, boxes, and menus) and how much it would engage students (i.e., elements of the technology that would increase student buy-in or interest). Teachers' comments were further categorized based on whether it was a positive aspect about the technology or whether it was an aspect that could benefit from change. These suggested changes reflected newly proposed changes or changes that were already planned by the research team in future iterations of GBG Tech.

Regarding positive aspects of the technology that were relevant to its usability, teachers indicated that the technology was easy to use, and it would take the burden off teachers considering many aspects of the game was done automatically (e.g., reviewing classroom rules, giving points/tallies, delivering rewards, determining teams). Teachers' feedback related to feasibility revealed that they appreciated there was some choice offered by the technology allowing for its easy integration into the classroom, as teachers could choose when to play the game and they could pause or stop the game if they needed to switch to a different instructional activity or if unforeseen interruptions occurred. Regarding acceptability, teachers were interested in trying it out in their classrooms, as the GBG is backed by research and GBG Tech resembles other reward systems teachers are using in their classrooms. The benefits of the team-based approach of the GBG were also noted. Specifically, team members could model positive behavior and students working together for a common goal could increase cohesion and collaboration, thus building a positive classroom environment. They particularly liked how the technology tracked student and team progress, highlighted the most-improved students and teams, and updated the points and tallies of teams in real-time to increase student accountability. Feedback related to student engagement revealed that teachers liked the digital and game-like nature of the technology and its ability to immediately reward students for rule-following behaviors because it would keep students engaged. Moreover, they felt that allowing students to choose their team mascot and image associated with their name would increase student buy-in. Regarding GBG Tech's visual design, teachers felt the color scheme and chosen mascots would attract the

attention of students. Finally, most teachers agreed that GBG Tech incorporated all components of the intervention, which would allow for sustained fidelity of the GBG.

There were several newly proposed changes suggested by teachers that were coded as relevant to the usability, feasibility, acceptability, student engagement, visual design, and fidelity of the technology. With respect to usability, teachers indicated that the ability to automatically upload their class rosters and create student avatars rather than uploading student photos would greatly reduce the amount of time spent setting up their virtual classrooms. To improve feasibility, they noted that GBG Tech would be easier to use in the classroom if the GBG scoreboard did not interfere with

displaying instructional materials on interactive boards, the virtual classrooms were accessible to substitute teachers, and student performance data could be shared across classrooms. In terms of acceptability, they asked for greater flexibility with some of the features such as the ability to adjust classroom rules, switch immediate rewards, and change the theme of mascots for older students. To improve the visual design, there were requests to simplify the performance data output by incorporating graphs that would accompany these data. Finally, it was suggested that videos of teachers using the technology in their classrooms were provided as a resource in GBG Tech, so teachers could witness how the GBG is delivered with high fidelity.

Table 2 Codes of suggested changes from transcript excerpts

Code	Suggested changes	Transcript excerpt
Usability	Automatic upload of class roster Student avatars instead of student photos Student portal—students take attendance Parent portal—messaging & sharing of data	I don't think I would upload a photo of the kids. Just because to me that would be something that would take a lot of time to try to get a photo. Maybe if they could choose an avatar or something like that, or if I could choose a different avatar for them rather than a picture. (F1P1) Is there a mass upload way? So, instead of having to manually add one by one, put all the student's information in an Excel sheet that you can import? (F2P2) Is there a way that students could log in once they get to class, and it's not on the teacher to check attendance? (F8P1) Is there a feature where you could share this information with parents, so parents could have access to it? (F7P2) A parent portal for the parents to see the data would be really helpful because they can see how their child is doing behaviorally (F3P3)
Feasibility	Display GBG scoreboard without interrupting ongoing classroom instruction Substitute teacher account Data follow students (integration across classrooms)	I think the biggest issue would just be trying to figure out what works as far as what I want to display for the students. How do I work around displaying work that I'm already showing the students versus pulling up the screen to record their points or their tallies? (F1P1) Kids act out when we have subs, so is there a way to make a substitute account or give them access to the main account? (F2P3) If the behavior data could follow students from class to class that would be great. (F4P3)
Acceptability	Multiple virtual classrooms Change theme of mascots for older students Modify classroom rules	I think adding the multiple roster feature for teachers with multiple classes each day would be helpful. (F5P1) Once they get to like fourth, fifth grade, they start to grow out of this. I think if they can tweak it per grade or age level so changing their picture or mascot on the scoreboard or showing them reaching a goal. (F7P1)
Student Engagement	Option to change immediate rewards Ability to change or add to mascots as a reward Student portal—personalize space within app	Do teachers have the option of eliminating certain rewards or adding to them? Are you able to input your own rewards? (F1P1) Maybe if they had avatars, if they were able to earn accessories for their avatar, like sunglasses or hats or different colors of hair or something like that, that might interest them. (F1P2) The mascots could change every week, so they could be a dog and then change the next week, so it would continue to be fun. (F6P2)
Visual Design	Simplify data output—graphical display	I might have a little bit of trouble on those reports, but I could probably work my way through it. (F5P2) The reports should show exactly what is going on. Like comparing individual students to a group or team of students. (F6P3)
Fidelity	Videos showing use of technology by teachers	I would like to, even though, you showed a video overview of it, I would like to see this stuff in action. I would like to see a video of somebody actually using it. (F7P2)

F Focus Group; *P* Participant

Some of the proposed changes suggested by teachers that were relevant to usability and student engagement were already planned for in future iterations of GBG Tech. Specifically, teachers suggested that students have the option to personalize their space within the application, take their own attendance, and change or add to their mascot as a reward. Teachers also suggested adding a feature that allowed them to message and share data with parents, as this would increase parental involvement and knowledge of how their students are behaving in the classroom. Table 2 provides several examples of how suggested changes by teachers were coded and the transcript excerpts from which these changes were retrieved.

Regarding changes adopted prior to the Year 3 Pilot Testing Phase, the ability for teachers to copy and paste their student roster directly into GBG Tech and create student avatars in addition to uploading pictures of their students was added as a new feature to the virtual classroom setup given the amount of time it was estimated to save teachers and the frequency with which this change was mentioned by teachers in the focus groups. An additional feature that was adopted included allowing teachers to set up multiple virtual classrooms within an account considering this change was outlined in the original specifications for the technology and teachers at the elementary school level teach different groups of students and should have the option to use the intervention with all their students. We also added the feature to switch immediate rewards if students or teachers were not satisfied by the random selection offered by GBG Tech considering this change could produce a greater level of impact (i.e., increase motivating nature of reward) while expending a low level of effort (i.e., time required for coding). Further, studies have shown that student-selected rewards tend to be more reinforcing and have a greater likelihood of leading to behavioral change than teacher-selected rewards (Cosden et al., 1995; Robichaux & Gresham, 2014). Finally, videos of teachers using GBG Tech in the classroom were added to the resource page of the application, as this was expected to increase teachers' understanding of how to use the technology while providing a model of how to implement GBG Tech with high fidelity, which is standard practice when training and coaching teachers to deliver new interventions (e.g., Reinke et al., 2014). Some changes were reserved for later versions of the technology (i.e., student portal, parent portal) because they were already planned for but not within the scope or budget allowance of the current funded project. We also did not incorporate changes that would negatively impact the fidelity of the GBG intervention (e.g., adjusting classroom rules).

Analytic Approach: Feasibility Testing

Descriptive statistics were run on quantitative data from measures of fidelity, usability, acceptability, and feasibility and on data collected from GBG Tech to: (1) obtain an estimate of the training time needed for teachers to reach a high level of fidelity, (2) determine how often teachers played a game using GBG Tech, and (3) obtain teachers' impressions of GBG Tech before and after using it in their classrooms. Paired t-tests were run to capture significant changes in teachers' impressions over time and Cohen's d was calculated as a measure of effect size. Given our small sample size, we also tested changes in teachers' impressions about GBG Tech using Bayesian methods. The Bayes Factor is a ratio that describes the odds of rejecting a null, so a value of a 3 means that the alternative hypothesis is 3 times more likely than the null. The results of both analytical methods are reported, as t-tests are commonly used and more straightforward to interpret than the Bayes Factor.

Results: Feasibility Testing

On average, teachers took 2.43 weeks (SD = 1.90) to reach a high level of fidelity (i.e., mean rating of at least "3.5"), which was calculated by averaging across teachers the week when they first reached fidelity. Importantly, the fidelity ratings continued to improve each week with a final mean fidelity score of 3.77 (SD = 0.37) suggesting teachers were able to maintain fidelity above this established threshold (see Table 3 for weekly mean fidelity scores by teacher). Teachers played the GBG with technology on average 5.40 times per week (SD = 1.76, range = 3 to 7.5 games per week), which

Table 3 Weekly mean fidelity scores by teacher

Teacher	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
A	2.88	2.63	3.00	3.13	2.88	**3.63**
B	3.00	3.13	3.38	**3.63**	**4.00**	–
C	**3.63**	**3.63**	**3.50**	**3.50**	**3.75**	**3.75**
D	**3.88**	**3.63**	**4.00**	**4.00**	–	–
E	**3.75**	–	3.25	3.00	–	–
F	3.13	**3.63**	**4.00**	**4.00**	–	–
G	2.75	**3.50**	3.25	3.25	3.13	**4.00**

Bold indicates that fidelity criterion (3.50 or above) was reached

meets the suggested dosage requirements for the GBG (Kellam et al., 2011).

Regarding teachers' impressions of GBG Tech before and after using it in their classrooms, the mean ratings for the TPIA (M range = 3.83–4.64, SD range = 0.48–1.07) fell within the "agree" to "very much agree" range for feasibility (i.e., ease of fitting intervention into their routine) and acceptability (i.e., how pleased they are with the intervention) and the "somewhat agree" to "agree" range for sustainability (i.e., motivation to continue using the intervention) across Time 1 and Time 2. The mean ratings for the URP-I (M range = 4.60–5.29, SD range = 0.41–0.52) fell within the "agree" to "strongly agree" range at Time 1 and within the "slightly agree" to "agree" range at Time 2 for feasibility, acceptability and understanding (i.e., knowledge of how to implement the intervention). A significant decrease in mean ratings was found for TPIA acceptability, $t(6) = 3.58$, $p = 0.01$, $BF_{10} = 5.95$, $d = 0.48$, and URP-I feasibility, $t(6) = 2.83$, $p = 0.04$, $BF_{10} = 2.40$, $d = 1.00$, from Time 1 to Time 2 but these ratings still fell within the "agree" range. No other significant change in mean ratings were found for the remaining scales on the TPIA or URP-I. Refer to Table 4 for means, standard deviations, and mean differences across timepoints for these outcome measures.

On the GBG Tech Feedback Survey, 100% of teachers indicated that the GBG Tech could be learned in a reasonable amount of time, and it met our main goals for development such as incorporating all treatment components of the GBG into the technology and ensuring the technology supports teachers' delivery of the GBG in the classroom. Further, five of the seven (72%) teachers involved in feasibility testing indicated that they would continue using GBG Tech every day with their students. Qualitatively, teachers noted that certain features of the technology stopped working for brief periods of time, which interfered with their ability to play the game, and the GBG scoreboard could not be resized without losing important team information (i.e., number of points/tallies

disaggregated by team), so it blocked some instructional materials on their interactive boards. These qualitative results coincided with quantitative findings, as there was a significant decrease found for acceptability and feasibility over time. Given these results and to improve feasibility, the scalability of the GBG Tech scoreboard was addressed, so team information could still be displayed when it was resized to occupy only 50% or 25% of the interactive board. Further, we learned that major modifications to GBG Tech should not occur when teachers are actively using the technology in their classrooms considering such modifications might result in regressions in the code (i.e., existing functionalities become disabled) so to maintain acceptability this practice was avoided during the Phase 3 Pilot Testing Phase.

Discussion

The primary goal of this project was to iteratively develop technology in collaboration with relevant education partners, so it is acceptable and feasible to use and sustains the fidelity of GBG implementation in classroom settings. During the development and refinement phase, teachers provided insightful feedback relevant to the usability (e.g., automatic upload of class roster), feasibility (e.g., scalability of GBG scoreboard), acceptability (e.g., multiple virtual classrooms), and fidelity (e.g., videos of teachers using the technology) of GBG Tech that guided important revisions to the technology. Interestingly, teachers often mentioned they preferred having a choice regarding when they deliver the intervention, what classroom rules are displayed, and which rewards are distributed to students; a preference that aligns with findings from other technology development studies (e.g., Owens et al., 2022). As suggested by Owens and colleagues (2022), the availability of more customizable options may enhance teachers' engagement and thus acceptability; however, when deciding what aspects of the technology to customize, it is

Table 4 Mean differences across timepoints for outcome measures

Scale	T1 Mean (SD)	T2 Mean (SD)	Difference	*p* value	BF_{10}	Cohen's *d*
Teacher Perceptions of Intervention Attributes (TPIA; Likert Scale 1–5)						
Feasibility	4.43 (0.61)	4.14 (0.75)	− 0.29	0.23	0.68	0.57
Acceptability	4.64 (0.48)	4.00 (0.50)	− 0.64	0.01	5.95	0.48
Sustainability	4.57 (0.54)	3.83 (1.07)	− 0.74	0.14	0.95	1.11
User Rating Profile—Intervention (URP-I; Likert Scale 1–6)						
Feasibility	5.11 (0.49)	4.56 (0.47)	− 0.55	0.04	2.40	1.00
Acceptability	5.01 (0.46)	4.60 (0.41)	− 0.41	0.07	1.57	0.84
Understanding	5.29 (0.52)	5.10 (0.45)	− 0.19	0.37	0.51	0.36

T1 = Time 1 and corresponds to when teachers received initial GBG training. T2 = Time 2 and corresponds to the week following the last GBG observation. BF_{10} = Bayes Factor. The Bayes Factor is a ratio that describes the odds of rejecting a null, so a value of a 3 means that the alternative hypothesis is 3 times more likely than the null

recommended that features are modifiable only if fidelity is not compromised. Students also contributed to the development process by selecting team mascots and behavioral rewards that were engaging and motivating to them. Several positive aspects of the technology were noted by teachers including its automatic features supportive of teacher delivery; game-like and team-based approach to engage students and cultivate collaboration among students; tracking of student performance data to inform future intervention efforts; engaging theme and motivating rewards to increase student buy-in, and equivalence to other behavior management techniques used in the classroom that align with their school culture.

During feasibility testing, it was revealed that teachers reached a high level of fidelity in 2.5 weeks when it typically takes 2 months for teachers to internalize evidence-based practices (Ramsey et al., 2021). Further, most teachers sustained that level of fidelity for 6 weeks and played the game at the recommended dosage (i.e., daily/5 times per week; Ialongo et al., 1999). This finding is quite extraordinary as large-scale clinical trials of the GBG often report dosage levels below these recommendations (e.g., Humphrey et al., 2022) and teachers are not using other technology-based supports to the degree that is expected (e.g., simulation training; Shernoff et al., 2021, 2022). Quantitative data derived from teacher-rated measures (i.e., TPIA, URP-I) provided initial evidence of GBG Tech's feasibility, acceptability, understanding, and sustainability, which suggests the contextual fit between GBG Tech, and its users is adequate. In fact, these ratings of social validity are comparable to ratings found for the GBG when enhanced with technology (i.e., ClassDojo; Lynne et al., 2017) and other technology-based supports (e.g., Behavior Report Card Online Platform; Owens et al.,). Although there was a slight drop in acceptability and feasibility after feasibility testing, these ratings were not unexpected considering they reflected the qualitative feedback we received from teachers. One of the primary issues impacting acceptability was that there were regressions in the code (i.e., existing features or functionalities became disabled) when new features were added. Regarding feasibility, teachers indicated that they were unable to use their interactive boards for other curriculum-based activities when using GBG Tech to play a game because minimizing the GBG scoreboard would make certain information no longer visible to their students. Indeed, it is not uncommon to see high expectations of technology about its potential benefits before actual use and find statistically significant reductions in enthusiasm after its use, which appears to be related to usability (Larson et al., 2020; Meiselwitz & Sadera, 2008). Fortunately, the feasibility testing of GBG Tech allowed for the identification and resolution of these bugs and feasibility issues prior to the larger-scale testing that was planned in the Year 3 Pilot Testing Phase.

Overall, these initial findings suggest that GBG Tech may be a promising alternative to standard GBG, especially if teachers reach fidelity quickly, this level of fidelity is sustained, and they continue to rate GBG Tech as acceptable, feasible and understandable after using the product in their classrooms for one or more years. In future studies, we intend to examine the sustained fidelity of GBG Tech in comparison to standard GBG implementation when added supports (i.e., coaching) are removed in a large-scale efficacy trial. If these goals are achieved, GBG Tech may train teachers more quickly compared to standard GBG training and ongoing coaching may not be needed to sustain fidelity in the classroom, thus circumventing the need for these additional resources.

Limitations and Future Directions

In the context of these promising findings, there are some limitations that should be discussed. First, the sample size of teachers, particularly for feasibility testing, was small although it is in line with recommendations for the number of participants needed for obtaining feedback and initial feasibility testing of technology (Turner et al., 2006). Second, there were only two months of the school year left for feasibility testing after scheduling initial GBG Tech trainings with teachers in the Spring, so there was limited time available to get a good estimate of sustained fidelity. Third, our research design in the Year 2 Refinement Phase did not include a comparison group, so we could not directly test differences in acceptability, feasibility, usability, and fidelity across versions of the GBG (i.e., standard vs. GBG Tech), although our mean ratings of acceptability for GBG Tech appear to be comparable to a positive variation of the GBG, a technology-enhanced version of GBG using ClassDojo, and higher than ratings given to standard GBG (Lynne et al., 2017; Wahl et al., 2016). Further, we did not collect student outcomes during the Year 2 Refinement Phase, so it is unclear how sustained fidelity of GBG Tech would influence student functioning in the behavioral, social, or academic domain. However, our Year 3 Pilot Testing Phase using a randomized controlled trial (RCT) that compared GBG Tech to standard GBG addresses many of these limitations and the analysis of these data is currently underway. Fourth, our sampling methods in Years 1 and 2 increased the risk of self-selection bias, as teachers participating in focus groups and feasibility testing may have had greater interest and motivation to use GBG Tech, thus translating into better impressions about the technology. Fifth, teachers who participated in focus groups did not have an opportunity to test out the technology and only watched a video of its

functionalities, which could have limited the depth or complexity of the feedback given. Finally, although the teachers participating in focus groups and feasibility testing matched many of the characteristics of teachers participating in prior efficacy studies of the GBG, some of the school districts comprising our recruitment area were more rural relative to other regions in the USA, so generalizability may be limited.

Regarding future directions, we intend to build a student portal in future iterations of GBG Tech that will allow students to take daily attendance, track their team progress, personalize the image or avatar linked to their account, and access or redeem their weekly rewards. It is anticipated that the student portal will increase the rewarding nature of the GBG and reduce the number of GBG-related tasks assigned to teachers. We also plan to build a parent portal that will allow parents to check how their children are behaving in the classroom, obtain suggestions of how to address these behaviors in the home setting (e.g., how to acknowledge or deliver rewards for prosocial behaviors), and regularly communicate with teachers by means of instant messaging. Some prior RCTs of the GBG have included a family–school partnership (FSP) component of treatment where families received messages from teachers that included updates on their students' progress in school as well as behavior management and learning activities to use in the home (e.g., Ialongo et al., 1999; Wilcox et al., 2022). The FSP was shown to positively influence conduct problems and achievement scores albeit the effect size was smaller than that of the GBG (Ialongo et al., 1999). As a result, it has been suggested that a multiplicative treatment effect may occur when a FSP component is combined with the GBG. Overall, the parent and student portal align with the planned functionalities outlined in the original specifications of GBG Tech and the changes suggested by teachers who participated in focus groups and feasibility testing.

Design Principles to Guide Technology Development

An initial set of design principles was outlined after the specifications of the technology were written and they have been added to and refined as we progressed with the development of GBG Tech. They were primarily informed by the lessons we learned through the process of building this technology and are less data-driven. They are intended to be used as guidelines to assist with the development of future technology-delivered interventions that are aimed at sustaining fidelity in classroom settings. These 10 design principles are described below along with suggestions of how to apply them to be amenable to technology-based solutions. Table 5 provides a summary of these 10 design principles.

Choose an intervention with well-defined and evidence-based treatment components

Ideally, the intervention should have an accompanying manual that identifies what components are necessary to promote treatment gains. If the active components of treatment are not clearly identified but the program is comprised of practices that are known to be evidence-based, it is vital these practices are built into the technology. A fidelity checklist may also provide guidance as to what treatment components should be included as part of the technology-delivered intervention. Importantly, the intervention should have undergone rigorous testing by means of well-designed experimental studies that show evidence of improving targeted outcomes.

Intervention should have a widely used and thoroughly tested fidelity checklist

A fidelity checklist allows for a quantifiable way of assessing how well an intervention is implemented over time. To increase its sensitivity to change, a fidelity checklist should not only rate whether a treatment component is present versus absent but also rate the quality with which that treatment component was delivered. It is also necessary to adapt the fidelity checklist so that it can assess the fidelity of both the standard and technology-based version of the intervention and equivalent enough, so it is possible to quantitatively compare differences in fidelity across delivery modalities. We were able to adapt the GBG Implementation Rubric, so the treatment components that were rated stayed the same, but the descriptors used to evaluate the quality of implementation were relevant to both delivery methods.

Know intervention and teacher-level factors that pose challenges to fidelity

There are several well-known intervention-level (e.g., easily administered, clear instructions) and teacher-level (e.g., self-efficacy, professional burn-out) factors that pose a challenge to implementing an intervention with high fidelity. However, this research area is ever evolving so staying abreast of research advancements is vital to inform what is needed to be addressed with a technology-based solution. As technology-delivered interventions are tested and their use is observed in classroom settings, it is likely new fidelity-interfering factors will emerge, which should prompt the consideration of novel solutions with input from relevant education partners. Indeed, the very act of changing an intervention's delivery method may result in new implementation challenges (e.g., standard delivery is more intuitive for teachers who have less experience using technology in their classrooms) and as we pivot and begin to use more technology-delivered

Table 5 Checklist of ten design principles

Design principle	Steps of implementation
1. Choose an intervention with well-defined and evidence-based treatment components	Intervention must have undergone rigorous testing and shown evidence of improving targeted outcomes Identify treatment components that facilitate change by referring to the intervention manual or fidelity checklist
2. Intervention should have a widely used and thoroughly tested fidelity checklist	Fidelity checklist should not only rate presence/absence of a treatment component but also the quality with which it is delivered Fidelity checklist should be adapted to assess fidelity of both versions of the intervention but be equivalent enough, so differences across delivery methods can be compared
3. Know intervention and teacher-level factors that pose challenges to fidelity	Stay abreast of research advancements to know what factors may undermine fidelity to inform a technology-based solution Testing technology-delivered interventions in the classroom may help identify new fidelity-interfering factors in consultation with education partners
4. Observe the delivery of the intervention in authentic settings and document implementation errors	Systematically observe the delivery of the intervention by teachers in their classrooms using psychometrically sound fidelity checklists Document implementation errors that occur most frequently or undermine the intervention's active treatment components
5. Incorporate features that prevent implementation errors and take the burden off users	Fidelity promoting features should be informed by the research literature or by observations made when coaching teachers on intervention delivery Attempt to automate most aspects of the intervention that is delivered by the user
6. Develop detailed specifications of the technology	Written specifications should detail all requirements of the technology including treatment components and fidelity promoting features Specify what aspects of the technology should be developed in the present iteration versus future iterations of the technology
7. Technology is expensive, so build the technology over stages	Prioritize building the active components of treatment and fidelity promoting features Reserve funds in the budget for revisions and be mindful that each developmental activity costs money
8. Allow for continual feedback from education partners and continual updates to the technology	Involve education partners as early as possible in the development process and give both formal and informal opportunities for feedback Solicit feedback from recipients of the intervention as it will increase buy-in and motivation to participate and ultimately their responsiveness to treatment
9. Technology can never be tested too much	Allow enough time (~4 months) to address regressions or additional bugs resulting from each round of updates Observe teachers use the technology in authentic settings to improve usability and identify additional bugs
10. Consider technological capabilities available to the user and how to integrate the newly developed technology into their routines	Surveys and classroom observations will provide useful information about how and when teachers use technology to support their teaching practices Use these data to inform what approach should be taken when training teachers on the newly developed technology and when its use is most feasible for them

interventions, future studies are needed to determine how this delivery method may influence fidelity.

Observe the delivery of the intervention in authentic settings and document implementation errors

To gain a thorough understanding of how interventions could fail or get derailed in authentic settings, it is important to systematically observe the delivery of the intervention by teachers in their classrooms using psychometrically sound fidelity checklists. Implementation errors that occur most frequently or undermine the intervention's active treatment components, thus diminishing expected treatment gains, should be documented in written form. Regarding our own observations when rating fidelity in the context of GBG efficacy trials, we found that students would often lose their motivation and discontinue their participation in a game once they received the total number of tallies for

rule violating behaviors that were permitted by their teacher within a game period. This observation informed how GBG technology would determine team winners (i.e., more points earned for rule-following behaviors than tallies for rule violating behaviors; positive variation of the GBG), and we developed notification windows that prompted teachers to be more mindful of prosocial behaviors and award points if tallies continually outweighed points for a specific team or if more than 5 tallies were consecutively given to the same team.

Incorporate features that prevent implementation errors and take the burden off users

The features that are built into a technology-delivered intervention should target fidelity-interfering factors highlighted by the extant literature and implementation errors frequently observed when teachers deliver the intervention in their classrooms. These features may be informed by what feedback teachers typically receive during consultation sessions with coaches aimed at improving the quality of intervention delivery. To lessen the burden on teachers, it is important to consider what aspects of the intervention could be automatically done by the technology that was the responsibility of teachers in the standard version of intervention.

Develop detailed specifications of the technology

To obtain accurate quotes from technology development firms, the written specifications must be as detailed as possible including the requirements of the technology (i.e., what must be present) and purpose of the technology. The treatment components and features designed to promote fidelity should also be clearly identified and noted as requirements. It is recommended that visual representations of each page of the application accompany the written specifications displaying how the different treatment components work together, how one page of the application flows to the next, and the functionality of each feature. The security of the data captured by the technology should also be considered, including how the data will be accessed, stored, and protected. Finally, it is important to specify what aspects of the technology should be developed in the present iteration versus future iterations of the technology, which can be guided by what aspects of the technology must be developed to fully deliver the intervention and feedback from expert consultants and relevant education partners.

Technology is expensive, so build the technology over stages

Technology built in stages supports a developmental process that is iterative and collaborative, as it allows opportunities for feedback and testing by users. This approach is also cost-effective, as it takes more money to go back and fix something that does not work for the intended user or recipient than building it right the first time. When determining what should be built now versus later, it is recommended that building the active components of treatment and features that address common implementation errors are prioritized whereas features that enhance student responsiveness (e.g., buy-in, motivation) or improve the user experience of teachers (e.g., time effectiveness, flexibility) are saved for later iterations. Be mindful that each development activity including user experience design, visual design, HTML development, and engineering, costs money; however, there should also be room in the budget for revisions, quality assurance/user acceptance testing, and bug fixes. Finally, there are less obvious expenses to consider such as hosting costs and the purchasing of domain names.

Allow for continual feedback from education partners and continual updates to the technology

Involve education partners as early as possible in the development process (e.g., writing technology specifications) and choose experts who are already well versed with the intervention and have experience delivering the intervention in settings where the technology will be used. There are both formal and informal opportunities for feedback with formal feedback occurring when specific deliverables or stages of development have been completed, whereas informal feedback may be ongoing. Formal forms of feedback may be obtained through questionnaires, focus groups, and direct observation by the research team and informal forms of feedback may include asking teachers how the technology is working for them during consultation visits or fidelity checks. In addition to soliciting feedback from users of the technology, it is advisable to give recipients of the intervention opportunities to provide feedback, as this will increase their buy-in and motivation to participate in the intervention, and thus their responsiveness. Given the time and cost of implementing feedback, it is not possible to enact all suggested changes and some changes are not in the best interest of the technology (e.g., not aligned with its purpose). As revealed by the results obtained from focus groups, teachers often asked for greater flexibility or customization within the application (e.g., modifying the length of game, rules of the classroom, immediate rewards) and we did not enact feedback if it changed the active components of the treatment or negatively influenced its fidelity.

Technology can never be tested too much

Although we allowed for 2 months of testing by the research team and engineering team and another 6 weeks of initial

feasibility testing by participating teachers, this was not sufficient time to address the regressions or additional bugs that resulted from updates informed by our focus groups. In retrospect, it would have been more advantageous to add another 2 to 3 months to our timeline to allow our engineering team to make updates before initial feasibility testing and another 2 to 3 months for updates after receiving feedback and observing teachers use the technology in their classrooms before our pilot testing phase comparing standard GBG to GBG Tech. Importantly, our ability to witness teachers using the technology in unexpected ways was helpful in improving the usability of the product and identifying additional bugs.

Consider technological capabilities available to the user and how to integrate the newly developed technology into their routines

Surveys should be administered to teachers to obtain a better understanding of their experience and comfort using technology, as well as what technology-related resources and level of technical support are currently available to them. If possible, classroom observations will also provide useful information about how and when teachers use technology to support their teaching practices. These data will help inform what approach should be taken when training teachers on the newly developed technology (e.g., direct instruction, hands-on learning) and when during their day using the newly developed technology is most feasible for them.

Conclusion

The GBG is an evidence-based universal intervention that supports students' behavioral, social-emotional, and academic development. Implementation drift is often a barrier to optimizing student outcomes, so our research team leveraged the benefits of technology to build a sophisticated online platform to support teachers' fidelity of the GBG. Following a collaborative research model, teachers were involved at every stage of GBG Tech development (i.e., consultants, focus group participants, evaluators of feasibility) and decisions of what feedback to implement was primarily guided by the time and cost-effectiveness of the change and if it promoted fidelity, among other considerations. Initial feasibility testing revealed that GBG Tech is considered acceptable, understandable, and feasible to teachers. Moreover, it substantially decreased the training time needed for teachers to reach a high level of fidelity and teachers played the game at the recommended dosage. Importantly, the results of this study informed a full version of GBG Tech that is ready for large-scale testing and this

development process informed a set of principles to guide the development of other technology-delivered interventions aimed at sustaining fidelity in authentic classroom settings.

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 Behavioral Analysis, 2*, 119–124.

Bradshaw, C. P., Zmuda, J. H., Kellam, S. G., & Ialongo, N. S. (2009). Longitudinal impact of two universal preventive interventions in first grade on educational outcomes in high school. *Journal of Educational Psychology, 101*, 926–937. https://doi.org/10.1037/a0016586

Becker, K. D., Bohnenkamp, J., Domitrovich, C., Keperling, J. P., & Ialongo, N. S. (2014). Online training for teachers delivering evidence-based preventive interventions. *School Mental Health, 6*(4), 225–236.

Becker, K. D., Darney, D., Domitrovich, C., Bradshaw, C., & Ialongo, N. S. (2013). Coaching teachers to enhance the implementation quality of evidence-based programs. *PsycEXTRA Dataset*. https://doi.org/10.1037/e601972013-001

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.

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation, 31*, 199–218. https://doi.org/10.1177/1098214010366173

Combs, K. M., Buckley, P. R., Lain, M. A., Drewelow, K. M., Urano, G., & Kearns, S. (2022). Influence of classroom-level factors on implementation fidelity during scale-up of evidence-based interventions. *Prevention Science, 23*(6), 969–981.

Cosden, M., Gannon, C., & Haring, T. G. (1995). Teacher-control versus student-control over choice of task and reinforcement for students with severe behavior problems. *Journal of Behavioral Education, 5*, 11–27. https://doi.org/10.1007/BF02110212

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control. *Clinical Psychology Review, 18*, 23–45. https://doi.org/10.1016/S0272-7358(97)00043-3

Diamond, K. E., & Powell, D. R. (2011). An iterative approach to the development of a professional development intervention for Head Start teachers. *Journal of Early Intervention, 33*(1), 75–93.

Dion, E., Roux, C., Landry, D., Fuchs, D., Webby, J., & Dupéré, V. (2011). Improving attention and preventing reading difficulties among low-income first-graders: A randomized study. *Prevention Science, 12*(1), 70–79. https://doi.org/10.1007/s11121-010-0182-5

Dolan, L. J., Kellam, S. G., Brown, C. H., Werthamer-Larsson, L., Rebok, G. W., Mayer, L. S., Laudolff, J., & Turkkan, J. S. (1993). The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement.

Journal of Applied Developmental Psychology, 14(3), 317–345. https://doi.org/10.1016/0193-3973(93)90013-L

Domitrovich, C. E., Bradshaw, C. P., Poduska, J. M., Hoagwood, K., Buckley, J. A., Olin, S., & Ialongo, N. S. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: A conceptual framework. *Advances in School Mental Health Promotion, 1*(3), 6–28. https://doi.org/10.1080/1754730X.2008.9715730

Domitrovich, C. E., Pas, E. T., Bradshaw, C. P., Becker, K. D., Keperling, J. P., Embry, D. D., & Ialongo, N. (2015). Individual and school organizational factors that influence implementation of the PAX good behavior game intervention. *Prevention Science, 16*(8), 1064–1074.

Domitrovich, C., & Ialongo, N. (2008). *The Teacher Perceptions of the Intervention Attributes (TPIA) scale*. Unpublished technical report.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research: Theory and Practice, 18*, 237–256. https://doi.org/10.1093/her/18.2.237

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE Open*. https://doi.org/10.1177/2158244014522633

Furr-Holden, C. D. M., Ialongo, N. S., Anthony, J. C., Petras, H., & Kellam, S. G. (2004). Developmentally inspired drug prevention: Middle school outcomes in a school-based randomized prevention trial. *Drug and Alcohol Dependence, 73*, 149–158.

Gaastra, G. F., Groen, Y., Tucha, L., & Tucha, O. (2016). The effects of classroom interventions on off-task and disruptive classroom behavior in children with symptoms of attention-deficit/hyperactivity disorder: A meta-analytic review. *PLoS ONE, 11*(2), 1–19. https://doi.org/10.1371/journal.pone.0148841

Greenberg, M. T., & Abenavoli, R. (2017). Universal interventions: Fully exploring their impacts and potential to produce population-level impacts. *Journal of Research on Educational Effectiveness, 10*, 40–67.

Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology, 33*(6), 665–679. https://doi.org/10.1007/s10802-005-7646-2

Hill, C. J., Scher, L., Haimson, J., & Granito, K. (2023). Conducting implementation research in impact studies of education interventions: A guide for researchers. (NCEE 2023–005). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288.

Huizink, A. C., Van Lier, P. A. C., & Crijnen, A. A. M. (2008). Attention deficit hyperactivity disorder symptoms mediate early-onset smoking. *European Addiction Research, 15*, 1–9.

Humphrey, N., Hennessey, A., Troncoso, P., Panayiotou, M., Black, L., Petersen, K., et al. (2022). The Good Behaviour Game intervention to improve behavioural and other outcomes for children aged 7–8 years: a cluster RCT. *Public Health Research, 10*(7), 1.

Ialongo, N. S., Werthamer, L., Kellam, S. G., Brown, C. H., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. *American Journal of Community Psychology, 27*(5), 599–641.

Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two first-grade preventive interventions on conduct

problems and disorder in early adolescence. *Journal of Emotional and Behavioral Disorders, 9*, 146–160.

Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., Petras, H., Ford, C., Windham, A., & Wilcox, H. C. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence, 95*(Suppl 1), S5–S28. https://doi.org/10.1016/j.drugalcdep.2008.01.004

Kellam, S. G., Mackenzie, A. C., Brown, C. H., Poduska, J. M., Wang, W., Petras, H., & Wilcox, H. C. (2011). The good behavior game and the future of prevention and treatment. *Addiction Science & Clinical Practice, 6*(1), 73–84.

Krueger, R. A., & Casey, M. A. (2009). *Focus groups: A practical guide for applied research* (4th ed.). Sage Publications.

Larson, K. E., Hirsch, S. E., McGraw, J. P., & Bradshaw, C. P. (2020). Preparing preservice teachers to manage behavior problems in the classroom: The feasibility and acceptability of using a mixed-reality simulator. *Journal of Special Education Technology, 35*(2), 63–75.

Leflot, G., Van Lier, P. A. C., Onghena, P., & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention study with the good behavior game. *Journal of Abnormal Child Psychology, 38*(6), 869–882.

Lewis, R. (2001). Classroom discipline and student responsibility: The students' view. *Teaching and Teacher Education, 17*(3), 307–319. https://doi.org/10.1016/S0742-051X(00)00059-7

Lynne, S., Radley, K. C., Dart, E. H., Tingstrom, D. H., Barry, C. T., & Lum, J. D. K. (2017). Use of a technology-enhanced version of the good behavior game in an elementary school setting. *Psychology in Schools, 54*, 1049–1063. https://doi.org/10.1002/pits.22043

Meiselwitz, G., & Sadera, W. (2008). Investigating the connection between usability and learning outcomes in online learning environments. *Journal of Online Learning and Teaching, 4*(2), 235–242.

Owens, J. S., Lee, M., Eackles, K., Medina, D., Evans, S. W., & Reid, J. (2022). Exploring teacher adoption, adaptation, and implementation of a daily report card intervention when using the daily report card online platform. *Frontiers in Education*. https://doi.org/10.3389/feduc.2022.981916

Owens, J. S., McLennan, J. D., Hustus, C. L., Haines-Saah, R., Mitchell, S., Mixon, C. S., et al. (2019). Leveraging technology to facilitate teachers' use of a targeted classroom intervention: Evaluation of the daily report card.online (DRC.O) system. *School Mental Health, 11*, 665–677. https://doi.org/10.1007/s12710-019-09320-6

Pas, E. T., Johnson, S. R., Larson, K. E., Brandenburg, L., Church, R., & Bradshaw, C. P. (2016). Reducing behavior problems among students with autism spectrum disorder: Coaching teachers in a mixed-reality setting. *Journal of Autism and Developmental Disorders, 46*(12), 3640–3652.

Payne, A. A., Gottfredson, D. C., & Gottfredson, G. D. (2006). School predictors of the intensity of implementation of school-based prevention programs: Results from a national study. *Prevention Science, 7*, 225–237. https://doi.org/10.1007/s11121-006-0029-2

Poduska, J. M., & Kurki, A. (2014). Guided by theory, informed by practice: Training and support for the good behavior game, a classroom-based behavior management strategy. *Journal of Emotional and Behavioral Disorders, 22*(2), 83–94.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health, 38*(2), 65–76. https://doi.org/10.1007/s10488-010-0319-7

Ramsey, K., Smith, S. D., Hansen, L., Mohn, F. S., Walbridge, F., Barajas, K., Ellis, B., & Dufrene, B. A. (2021). Internalization

of behavior management skills among teachers in a specialized school serving students with Neurodevelopmental Disorders. *Behavior Modification*. https://doi.org/10.1177/0145445521 1010708

Rebok, G. W., Hawkins, W. E., Krener, P., Mayer, L. S., & Kellam, S. G. (1996). Effect of concentration problems on the malleability of children's aggressive and shy behaviors. *Journal of American Academy of Child and Adolescent Psychiatry, 35*, 193–203.

Reinke, W. M., Stormont, M., Herman, K. C., Wang, Z., Newcomer, L., & King, K. (2014). Use of coaching and behavior support planning for students with disruptive behavior within a universal classroom management program. *Journal of Emotional and Behavioral Disorders, 22*(2), 74–82. https://doi.org/10.1177/1063426613519820

Ringwalt, C., Vincus, A., Ennett, S., Johnson, R., & Rohrbach, L. A. (2004). Reasons for teachers' adaptation of substance use prevention curricula in schools with non-white student populations. *Prevention Science, 5*, 61–67. https://doi.org/10.1023/B:PREV.0000013983.87069.a0

Robichaux, N. M., & Gresham, F. M. (2014). Differential effects of the mystery motivator intervention using student-selected and mystery rewards. *School Psychology Review, 43*(3), 286–298. https://doi.org/10.1080/02796015.2014.12087428

Schaffer, K., Rouiller, S., Embry, D., & Ialongo, N. (2006). *The PAX good behavior game implementation rubric*. Unpublished technical report.

Shernoff, E. S., Frazier, S. L., Lisetti, C., Delmarre, A., Bibi, Z., & Gabbard, J. (2021). Supporting the implementation of evidence-based behavior management practices through simulation: A mixed method study. *Journal of Educational and Psychological Consultation, 31*(4), 463–497.

Shernoff, E. S., Lekwa, A. L., Delmarre, A., Gabbard, J., Stokes-Tyler, D., Lisetti, C., & Frazier, S. L. (2022). Bridging simulation technology with positive behavioral supports to promote student engagement and behavior. *Journal of School Psychology, 95*, 121–138. https://doi.org/10.1016/j.jsp.2022.10.002

Smith, S., Barajas, K., Ellis, B., Moore, C., McCauley, S., & Reichow, B. (2021). A meta-analytic review of randomized controlled trials of the good behavior game. *Behavior Modification, 45*(4), 641–666. https://doi.org/10.1177/0145445519878670

Smith, S. D., Vitulano, L. A., Katsovich, L., Li, S., Moore, C., Li, F., Grantz, H., Zheng, X., Eicher, V., Aktan Guloksuz, S., Zheng, Y., Dong, J., Sukhodolsky, D. G., & Leckman, J. F. (2020). A Randomized controlled trial of an integrated brain, body, and social intervention for children with ADHD. *Journal of Attention Disorders, 24*(5), 780–794. https://doi.org/10.1177/1087054716647490

Storr, C. L., Ialongo, N. S., Kellam, S. G., & Anthony, J. C. (2002). A randomized controlled trial of two primary school intervention strategies to prevent early onset tobacco smoking. *Drug and Alcohol Dependence, 66*, 51–60. https://doi.org/10.1016/s0376-8716(01)00184-3

Stein, B. D., Kataoka, S., Jaycox, L. H., Wong, M., Fink, A., Escudero, P., & Zaragoza, C. (2002). Theoretical basis and program design of a school-based mental health intervention for traumatized immigrant children: A collaborative research partnership. *The Journal of Behavioral Health Services & Research, 29*(3), 318–326.

Turner, C. W., Lewis, J. R., & Nielsen, J. (2006). Determining usability test sample size. In W Karwowski (Ed.). International Encyclopedia of ergonomics and human factors (Vol. 3, No. 2, pp. 3084–3088). CRC Press.

Van Lier, P. A., Huizink, A., & Crijnen, A. (2009). Impact of a preventive intervention targeting childhood disruptive behavior problems on tobacco and alcohol initiation from age 10 to 13 years. *Drug and Alcohol Dependence, 100*, 228–233. https://doi.org/10.1016/j.drugalcdep.2008.10.004

Van Lier, P. A., Vuijk, P., & Crijnen, A. A. (2005). Understanding mechanisms of change in the development of antisocial behavior: The impact of a universal intervention. *Journal of Abnormal Child Psychology, 33*, 521–535.

Vogl, L. E., Teesson, M., Newton, N. C., & Andrews, G. (2012). Developing a school-based drug prevention program to overcome barriers to effective program implementation: The CLIMATE schools: Alcohol module. *Open Journal of Preventive Medicine, 2*(3), 410–422.

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*(4), 493–517. https://doi.org/10.1177/0145445516644220

Wilcox, H. C., Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., & Anthony, J. C. (2008). The impact of two universal randomized first- and second-grade classroom interventions on young adult suicide ideation and attempts. *Drug and Alcohol Dependence, 95*(Suppl 1), S60–S73. https://doi.org/10.1016/j.drugalcdep.2008.01.005

Wilcox, H. C., Petras, H., Brown, H. C., & Kellam, S. G. (2022). Testing the impact of the whole-day Good Behavior Game on aggressive behavior: Results of a classroom-based randomized effectiveness trial. *Prevention Science*. https://doi.org/10.1007/s11121-022-01334-y

Witvliet, M., Van Lier, P. A. C., Cuijpers, P., & Koot, H. M. (2009). Testing links between childhood positive peer relations and externalizing outcomes through a randomized controlled intervention study. *Journal of Consulting and Clinical Psychology, 77*(5), 905–915.