



## OPEN ACCESS

## EDITED BY

Sarah Powell,  
University of Texas at Austin,  
United States

## REVIEWED BY

Stefan Blumenthal,  
University of Rostock, Germany  
Jana Jungjohann,  
University of Regensburg, Germany

## \*CORRESPONDENCE

Tatjana Leidig  
tleidig@uni-koeln.de

## SPECIALTY SECTION

This article was submitted to  
Educational Psychology,  
a section of the journal  
Frontiers in Education

RECEIVED 10 April 2022

ACCEPTED 28 June 2022

PUBLISHED 29 July 2022

## CITATION

Leidig T, Casale G, Wilbert J,  
Hennemann T, Volpe RJ, Briesch A and  
Grosche M (2022) Individual,  
generalized, and moderated effects  
of the good behavior game on at-risk  
primary school students: A multilevel  
multiple baseline study using  
behavioral progress monitoring.  
*Front. Educ.* 7:917138.  
doi: 10.3389/feduc.2022.917138

## COPYRIGHT

© 2022 Leidig, Casale, Wilbert,  
Hennemann, Volpe, Briesch and  
Grosche. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Individual, generalized, and moderated effects of the good behavior game on at-risk primary school students: A multilevel multiple baseline study using behavioral progress monitoring

Tatjana Leidig <sup>1\*</sup>, Gino Casale <sup>2</sup>, Jürgen Wilbert <sup>3</sup>,  
Thomas Hennemann <sup>1</sup>, Robert J. Volpe <sup>4</sup>,  
Amy Briesch <sup>4</sup> and Michael Grosche <sup>2</sup>

<sup>1</sup>Department of Special Education and Rehabilitation, Faculty of Human Sciences, University of Cologne, Cologne, Germany, <sup>2</sup>Institute of Educational Research, University of Wuppertal, Wuppertal, Germany, <sup>3</sup>Institute of Inclusive Education, Faculty of Human Sciences, University of Potsdam, Potsdam, Germany, <sup>4</sup>Department of Applied Psychology, Northeastern University, Boston, MA, United States

The current study examined the impact of the Good Behavior Game (GBG) on the academic engagement (AE) and disruptive behavior (DB) of at-risk students' in a German inclusive primary school sample using behavioral progress monitoring. A multiple baseline design across participants was employed to evaluate the effects of the GBG on 35 primary school students in seven classrooms from grade 1 to 3 ( $M_{age} = 8.01$  years,  $SD_{age} = 0.81$  years). The implementation of the GBG was randomly staggered by 2 weeks across classrooms. Teacher-completed Direct Behavior Rating (DBR) was applied to measure AE and DB. We used piecewise regression and a multilevel extension to estimate the individual case-specific treatment effects as well as the generalized effects across cases. Piecewise regressions for each case showed significant immediate treatment effects for the majority of participants (82.86%) for one or both outcome measures. The multilevel approach revealed that the GBG improved at-risk students' classroom behaviors generally with a significant immediate treatment effect across cases (for AE,  $B = 0.74$ ,  $p < 0.001$ ; for DB,  $B = -1.29$ ,  $p < 0.001$ ). The moderation between intervention effectiveness and teacher ratings of students' risks for externalizing psychosocial problems was significant for DB ( $B = -0.07$ ,  $p = 0.047$ ) but not for AE. Findings are consistent with previous studies indicating that the GBG is an appropriate classroom-based intervention for at-risk students and expand the literature regarding differential effects for affected students. In addition, the study supports

the relevance of behavioral progress monitoring and data-based decision-making in inclusive schools in order to evaluate the effectiveness of the GBG and, if necessary, to modify the intervention for individual students or the whole group.

#### KEYWORDS

classroom behavior, good behavior game, multilevel analysis, piecewise regression, single case design

## Introduction

The national prevalence rates of mental health problems in Germany indicate that approximately 17 to 20% of schoolchildren demonstrate psychosocial problems (i.e., externalizing problems such as aggressive behavior or hyperactivity, and internalizing problems such as depressiveness or anxiety) with various degrees of severity (Barkmann and Schulte-Markwort, 2012; Klipker et al., 2018). These problems can negatively affect classroom behaviors and the social and academic performance of individual students (Kauffman and Landrum, 2012) and their peers (Barth et al., 2004). As the majority of students with behavioral problems in Germany are educated in inclusive schools without special education services (Volpe et al., 2018), teachers need strategies to successfully deal with their behavior problems, promote social-emotional development, and continuously evaluate the effectiveness of the implemented strategies.

Research reviews have demonstrated that group contingencies such as the Good Behavior Game (GBG; Barrish et al., 1969) are effective interventions for managing externalizing behavior problems in general education classrooms (Maggin et al., 2017; Fabiano and Pyle, 2019). The GBG is an easy-to-use interdependent group contingency intervention with extensive empirical support that utilizes students' mutual dependence to reduce problem behaviors and to improve prosocial and academic behaviors (Flower et al., 2014). The main features of this game are easily comprehensible (Flower et al., 2014): (1) selecting goals and rules, (2) recording rule violations, (3) explaining the rules of the game and determining the rewards, (4) dividing the students in two or more teams to play against each other, and (5) playing the GBG for a specified amount of time. In the classic variant, the team receives a mark ("foul") when a team member breaks a rule. At the end of the game, the team with the fewest fouls wins. In multi-tiered-systems of support (MTSS; Batsche, 2014), the GBG is typically integrated in regular classroom instruction in tier 1 as a part of a proactive classroom management approach (Simonsen and Myers, 2015). When the GBG is played in conjunction with behavioral progress monitoring, this combination provides an appropriate way to facilitate data-based decision-making in supporting students at risks for emotional and behavioral disorders (EBD) within

a MTSS: If an individual student does not respond to the GBG, this information can be used to decide whether further interventions should be added in tier 2 (Donaldson et al., 2017). In Germany, evidence-based practice within MTSS is still in its infancy (e.g., Voß et al., 2016; Hanisch et al., 2019); this also applies to globally known interventions such as the GBG and data-based decision-making based on behavioral progress monitoring. Given the potential benefits of the GBG for students with or at risk for EBD, additional information is needed on its impact and suitability in a German population.

## Evidence base of the good behavior game

For group design studies in general, the GBG meta-analysis by Flower et al. (2014) indicated moderate effects (Cohen's  $d = 0.50$ ) on problem behaviors (e.g., aggression, off-task behavior, and talking out), whereas the meta-analysis of randomized-controlled trials of the GBG by Smith et al. (2021) resulted in only small but significant effect sizes for conduct problems (i.e., aggression or oppositional behavior) (Hedges'  $g = 0.10$ ,  $p = 0.026$ ) and moderate but not significant effect sizes for inattention (i.e., concentration problems and off-task behavior) (Hedges'  $g = 0.49$ ,  $p = 0.123$ ). For conduct problems, the comparatively smaller effects are likely due to the rigorous standards of randomized-controlled trials, implying, for instance, less biased estimates of study effects; for inattention, this may reflect the significant heterogeneity of the findings and the overall small number of included studies (Smith et al., 2021). Meta-analyses of single-case research of the GBG (Bowman-Perrott et al., 2016; Flower et al., 2014) and class-wide interventions for supporting student behavior (Chaffee et al., 2017) found a significant and immediate treatment effect for reducing challenging behaviors (e.g., disruptive behavior, aggression, off-task behavior, talking out, and out-of-seat) (Flower et al., 2014:  $\beta = -0.2038$ ,  $p < 0.01$ ) and medium to high effects across both general and special education settings (Bowman-Perrott et al., 2016:  $\text{Tau}U = 0.82$ ; Chaffee et al., 2017:  $\text{Tau}U = 1.00$ ) with larger effects on disruptive and off-task behavior (e.g., out-of-seat, talking out, interrupting, pushing and fighting) ( $\text{Tau}U = 0.81$ ) than on-task behavior (e.g., working quietly, following teacher's instructions, and getting

materials without talking) ( $\text{Tau}U = 0.59$ ), and higher effects for students with or at risk for EBD ( $\text{Tau}U = 0.98$ ) than for students without any difficulties ( $\text{Tau}U = 0.76$ ) (Bowman-Perrott et al., 2016). Although the meta-analyses revealed the effectiveness of the GBG across settings, it must be noted that the majority of research examined the impact of the GBG in general education settings with typically developing students (Moore et al., 2022). Even though the studies included often lack concrete information on the implementation of mainstreaming or inclusive education in the sample, against the background of the development of the school systems since the 1990s, it can be assumed that many studies of the last 30 years were conducted in mainstreaming or inclusive settings.

When interpreting the meta-analytic findings, some methodological aspects have to be considered. In some studies using a group design, students were nested within classrooms. This approach is helpful to infer the effectiveness for a population of students based on inferential statistics. However, it does not automatically allow to distinguish which students the GBG was effective for and which students did not benefit at all (or even increased their problem behavior). In fact, in group research designs, the behavior of individual students differing from the mean of all students is considered a measurement error, and the groups' effect size does not tell us anything about case-specific treatment effects (Lobo et al., 2017). In principle, this also applies to subgroup analyses, even if they examine the effects more specifically. Furthermore, as Smith et al. (2021) noted, the available group design studies and single case studies measure similar but finally different outcomes: While the results of group design studies allow conclusions regarding general, cross-situational changes in student behavior (trait), no conclusions can be drawn about the effects of GBG on targeted behavior in the classroom situations in which GBG is played (state). The reverse is equally true. Against this background, the comparability of results of group studies and single-case studies is constrained. In addition, regarding the meta-analyses of single-case research, it is important to consider that most of the included studies investigated the impact of the GBG on a group or a class, not on individual students (Donaldson et al., 2017).

### Evidence base for effects on at-risk students in single-case research using progress monitoring

There is limited information about the impact of the GBG on students with or at risk for EBD in single-case research using progress monitoring. As Bowman-Perrott et al. (2016) critically noted, most studies included in their meta-analysis focused on typically developing students with a normal range of disruptive behaviors. From a methodological perspective, the information provided by single-case research is limited due to the composition and nature of the samples and the methods of data analysis used. With only a few exceptions, the vast majority of single-case studies investigated a class or a group as a single-case and analyzed data at the classroom level (Donaldson et al., 2017). This approach is helpful to test the positive impact of

the GBG on a group as a whole, but it does not allow analysis at the level of individual students (Donaldson et al., 2017; Foley et al., 2019). Furthermore, the sample sizes are relatively small (i.e., three to 12 cases), and the studies varied widely with regard to sample characteristics (e.g., setting, classroom size, class composition). As a result, neither the findings of individual studies nor the aforementioned meta-analyses can simply be generalized to specific populations of students, e.g., at-risk students (Bowman-Perrott et al., 2016; Donaldson et al., 2017).

In only three of the four studies reporting individual student data included in the meta-analysis by Bowman-Perrott et al. (2016), the target students were explicitly identified as the most challenging students in their class (i.e., students displaying more disruptive behaviors than peers) (Medland and Stachnik, 1972; Tanol et al., 2010; Hunt, 2012). We found six further studies investigating the impact of the GBG on students with or at risk for EBD at the level of individual students (Donaldson et al., 2017; Groves and Austin, 2017; Pennington and McComas, 2017; Wiskow et al., 2018; Foley et al., 2019; Moore et al., 2022). In seven of the aforementioned nine studies, all of the target students showed improvements, although to differing degrees. However, Donaldson et al. (2017) and Hunt (2012) found individual non-responders. Furthermore, Donaldson et al. (2017) and Moore et al. (2022) reported decreasing positive effects of the GBG over time for some children who frequently exhibited disruptive behavior. As Donaldson et al. (2017) concluded, only teachers who play the GBG with progress monitoring at the level of individual students can avoid unnecessary implementation of individualized interventions, and identify students who need additional support beyond the class-wide intervention.

In the few studies that have examined the effects of the GBG at the level of individual students, one can analyze the differential impact on individual students, but one cannot conclude whether the GBG is effective for a specific group of students, e.g., with or at risk for EBD. No study at the individual level or at the group level used an inferential statistic approach to examine whether the effects are statistically significant, which is important to generalize the results (Shadish et al., 2013). Even in single-case research, researchers are not only interested in specific individual treatment effects to support evidence-based decisions, but also whether the effects can be generalized to other cases. To our knowledge, no study evaluating the efficacy of the GBG has used inferential statistics to generalize the average treatment effect across cases within the same study. Accordingly, neither the results of studies at classroom level nor those at the level of individual students are representative for students with or at risk for EBD, so that the evidence for this group remains comparatively weak (Bowman-Perrott et al., 2016; Donaldson et al., 2017). In addition to case-specific inferential statistics, multilevel models enable investigating the overall effectiveness across participants considering the average

treatment effect, variations across cases and possible influential factors (Moeyart et al., 2014).

### Psychosocial problems as moderator for the effects of the good behavior game

Additional research on the students' individual characteristics is needed to assess the possible influential factors (Bowman-Perrott et al., 2016; Maggin et al., 2017). The results of group design studies suggest that students' psychosocial problems (i.e., externalizing problems such as aggressive behavior or hyperactivity, and internalizing problems such as depressiveness or anxiety) are associated with the effectiveness of the GBG: They indicate differential effects of the GBG on externalizing and internalizing problems depending on the students' individual risk levels and types (e.g., van Lier et al., 2005; Kellam et al., 2008; Spilt et al., 2013) with partly contradictory results for students with or at risk for severe externalizing behavior problems (e.g., aggressive, violent, and criminal behavior) and for students with combinations of risks (e.g., combination of social and behavior risks). Analyzing single-case research, Bowman-Perrott et al. (2016) identified the EBD risk status as a potential moderator for the effect on externalizing classroom behaviors (i.e., larger effects for students with EBD or at risk for EBD), however without further investigation of the risk level and subtype of behavior problems. Therefore, the role of students' psychosocial problems as a potential moderator requires further investigation.

### The current study

In sum, the existing research adds different pieces to the puzzle whether GBG is an evidence-based intervention, partly also for students with or at risk for EBD (Bowman-Perrott et al., 2016). There is some support for the effectiveness of the GBG either for individual students or for groups of students, e.g., with or at risk for EBD (Joslyn et al., 2019; Smith et al., 2021). However, our study is the first study about the GBG that uses a large sample of individual at-risk students nested in classrooms and tested the hypothesis of individual, general, and differential effectiveness within the same study using inferential statistics on individual and group level at the same time. Typically, the number of cases in single-case research varies between 1 and 13 (Shadish and Sullivan, 2011). Our large sample of 35 students and our methodological approach give us the opportunity to examine case-specific treatment effects as well as effects across cases under stable conditions. Combining group statistics and single-case statistics makes all Council of Exceptional Children's standards for classifying evidence-based special education interventions (Cook et al., 2015) applicable for our study (see [Supplementary Material](#)).

Therefore, the current study is designed to extend previous research by investigating the impact of the GBG on at-risk students' classroom behaviors in inclusive settings in Germany

using behavioral progress monitoring and analyzing the data using regression analyses to estimate case-specific treatment effects as well as effects across students. In particular, we investigate the interaction between the impact of the GBG and students' psychosocial problems due to the potential influence of students' individual risks. We hypothesized moderate to large effects on at-risk students' academic engagement and disruptive behavior for the majority of students and across all cases with an immediate treatment effect, and only a small additional slope effect. Furthermore, we anticipate that students' psychosocial problems moderate the intervention effects. We hypothesize that the impact of the GBG would interact with the magnitude of students' behavioral problems. Specifically, we assume that the higher students' externalizing problems are, the more effective the GBG will be.

## Method

### Participants and setting

In Germany, inclusion in schools is primarily understood as learning together of students with and without special educational needs in general school settings (KMK, 2011). This study was conducted in an inclusive primary school in a midsize town in western Germany (North Rhine-Westphalia) with 12 first-through-fourth-grade classrooms. Due to the legal requirements of North Rhine-Westphalia, support in the areas of learning, speech and behavior can be provided in all grades regardless of the formally identified special educational needs, so that there is no administrative ascriptive diagnosis for some of the students with special educational needs. Furthermore, as a rule, no formal diagnosis of special educational needs in learning, behavior, and speech is provided for first and second graders. In the inclusive primary school in our study, approximately 50 children were with or at risk for learning disabilities and/or EBD. In line with official school records, approximately 70% of all students had a migration background.

After the introduction of the project during the teachers' conference, seven general education teachers decided to participate in the study with their classes. Each of these classroom teachers nominated five students based on their professional experience and judgment as the most challenging students in their classroom ( $n = 35$ ; five first graders, 15 second graders and 15 third graders; five girls and 30 boys). Based on teacher ratings on the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), 22 (63%) of the nominated students showed externalizing risks, whereas only two (6%) demonstrated internalizing risks, two students (6%) showed risks in both areas, and nine students (26%) exposed no psychosocial risks. Further demographic information and the SDQ data are summarized in [Table 1](#).

After consultation with the teachers to identify the classroom situation believed to be the most problematic

TABLE 1 Student demographic information and data of the behavioral screening instrument (SDQ).

Classroom (Grade)	Student	Gender	Age	Migration background	SDQ	
					Externalizing score <sup>a</sup>	Internalizing score <sup>b</sup>
A (1 <sup>st</sup> )	S1	Boy	6	0	9 <sup>+</sup>	1
	S2	Boy	6	0	11 <sup>++</sup>	0
	S3	Girl	7	0	6	0
	S4	Boy	7	1	5	3
	S5	Boy	6	1	7 <sup>+</sup>	1
B (2 <sup>nd</sup> )	S6	Boy	7	1	8 <sup>+</sup>	1
	S7	Boy	8	1	4	3
	S8	Boy	7	1	13 <sup>+++</sup>	3
	S9	Boy	7	1	15 <sup>+++</sup>	6 <sup>+</sup>
	S10	Boy	7	0	18 <sup>+++</sup>	4
C (2 <sup>nd</sup> )	S11	Girl	7	1	10 <sup>++</sup>	3
	S12	Boy	7	0	13 <sup>+++</sup>	1
	S13	Boy	9	0	13 <sup>+++</sup>	3
	S14	Boy	7	1	10 <sup>++</sup>	4
	S15	Boy	8	1	8 <sup>+</sup>	1
D (2 <sup>nd</sup> )	S16	Boy	7	1	4	0
	S17	Boy	7	0	6	3
	S18	Boy	7	0	9 <sup>+</sup>	4
	S19	Boy	7	1	4	7 <sup>+</sup>
	S20	Girl	7	1	6	10 <sup>+++</sup>
E (3 <sup>rd</sup> )	S21	Boy	9	1	3	1
	S22	Boy	8	1	6	0
	S23	Boy	9	0	8 <sup>+</sup>	5
	S24	Boy	9	0	6	1
	S25	Boy	8	0	12 <sup>+++</sup>	3
F (3 <sup>rd</sup> )	S26	Girl	9	0	13 <sup>+++</sup>	2
	S27	Boy	8	1	11 <sup>++</sup>	2
	S28	Girl	8	1	7 <sup>+</sup>	3
	S29	Boy	8	1	10 <sup>++</sup>	1
	S30	Boy	8	0	10 <sup>++</sup>	9 <sup>++</sup>
G (3 <sup>rd</sup> )	S31	Boy	8	0	6	2
	S32	Boy	8	1	9 <sup>+</sup>	0
	S33	Boy	8	0	9 <sup>+</sup>	3
	S34	Boy	9	1	11 <sup>++</sup>	1
	S35	Boy	8	0	10 <sup>++</sup>	0

For the migration background, 0 = no and 1 = yes. For SDQ externalizing and internalizing scores, + = slightly raised, ++ = high and +++ = very high.

<sup>a</sup>Categorization by Goodman et al. (2010): 7-9 = slightly raised, 10-11 = high, 12-20 = very high. <sup>b</sup>Categorization by Goodman et al. (2010): 6-7 = slightly raised, 8-9 = high, 10-20 = very high.

concerning externalizing classroom behaviors, we defined the target instructional period for playing the GBG for each classroom (i.e., individual work in math for classroom B, D, and G, and individual work in German language for classrooms A, C, E, and F).

## Measures

### Direct behavior rating

Direct Behavior Rating (DBR; Christ et al., 2009) was used to measure *academic engagement* (AE) and *disruptive behavior* (DB). DBR has already been adapted, evaluated, and

implemented for German classrooms using the operational definitions of AE and DB provided by Chafouleas (2011) (Casale et al., 2015, 2017). AE included students' active or passive participation in ongoing academic activities such as engaging appropriately in classroom activities, concentrated working, completing tasks on time and raising their hands. DB was defined as behaviors disrupting others or affecting students' own or other students' learning, such as speaking without permission, leaving one's seat, noisemaking, having undesirable private discussions and fooling around. For both dependent variables, we used a single-item scale (SIS) with these broadly defined items representing a common behavior class (Christ et al., 2009). Previous

research focusing on the German DBR scales supports their generalizability and dependability across different raters (i.e., general classroom teachers and special education teachers), items, and occasions (Casale et al., 2015, 2017). These studies showed that the DBR-SIS provides dependable scores ( $\Phi > 0.70$ ) for an individual student's behavior after four measurement occasions. Furthermore, measurement invariance testing across high-frequency occasions with short intervals showed the sensitivity of DBR (Gebhardt et al., 2019). In addition, previous research from the United States supports the reliability, validity and sensitivity of DBR to changes to the SIS regarding the targeted behaviors (e.g., Briesch et al., 2010; Chafouleas et al., 2012). The teachers observed the behavior of the five nominated students in the target classroom situation during baseline and intervention. Immediately after each observation, the teachers rated both AE and DB for each of the five students on a paper-pencil questionnaire using a 6-point Likert-type scale (i.e., 0 = *never*, 5 = *always*).

### Strengths and difficulties questionnaire for teachers

Participants' internalizing and externalizing problems were assessed with the German version of the worldwide used behavioral screening questionnaire SDQ (Goodman, 1997) for teachers. It consists of 25 items equally divided across five subscales (hyperactivity, conduct problems, emotional problems, peer problems, and prosocial behavior). Besides the 5-factor model, a 3-factor model containing the factors of externalizing problems, internalizing problems, and prosocial behavior is used to interpret results (Goodman et al., 2010). Evaluations of the German version indicate an acceptable fit both of the 5-factor model (Bettge et al., 2002) and the 3-factor model (DeVries et al., 2017), and a good internal consistency for the teacher version of the SDQ (Cronbach's  $\alpha$  between 0.77 and 0.86; Saile, 2007). After each item was scored by teachers on a 3-point Likert-type scale (0 = *not*

*true*, 1 = *somewhat true*, 3 = *certainly true*), we calculated the internalizing and externalizing subscales by summing the conduct and hyperactivity scales for the externalizing subscale and summing the emotional and peer problems scales for the internalizing subscale (Goodman et al., 2010).

## Design and procedures

A multiple baseline across-classroom design with a 2-week staggered randomized-phase start was used. Each classroom was randomly assigned to a single *a priori* designated intervention start point (Kratochwill and Levin, 2010). After the baseline phase, which was previously determined to be at least 8 but not more than 38 days, the intervention phase (70 to 100 days) started with a 2-week interval between the groups depending on the start point. Due to holidays and cancelled lessons because of school events, full-day conferences, part-time employment of teachers, and in-service training, the number of days with the opportunity to play the GBG fluctuated between 61 to 87 days. The GBG was played *de facto* 3 to 5 days a week. Phase lengths for each classroom are displayed in Table 2.

### Teacher training and mentoring

To support an effective implementation, we developed a program with the components shown in Figure 1 according to recommendations from the literature (Hagermoser Sanetti et al., 2014; Poduska and Kurki, 2014) and considering the specific school setting. In addition, the teachers received training on the use of DBR based on Chafouleas (2011), consisting of the theoretical background, the practical use of DBR, and feedback by a research assistant.

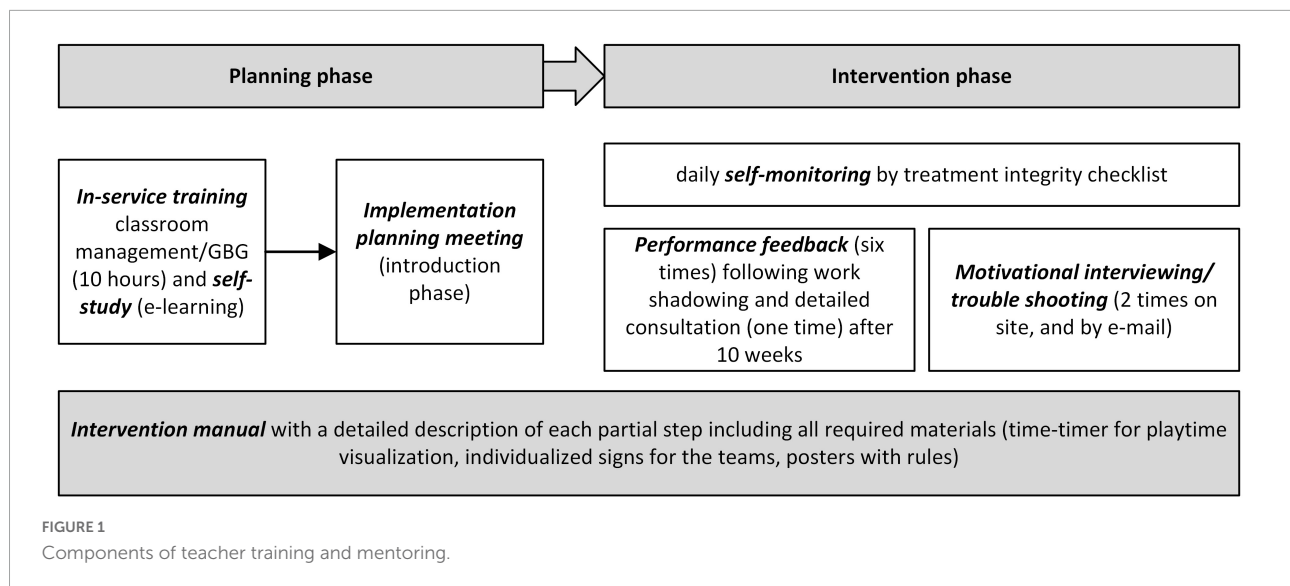
In total, the training before the implementation of the GBG included 14 hours for classroom management and GBG (i.e., 10 hours in-service training and 4 hours self-study) and 5 hours for DBR (in-service training). Self-study was implemented as an e-learning course with text, audio, video, and

TABLE 2 Design data, treatment integrity, and treatment usability for each classroom.

Classroom	A	B	Treatment Intensity <sup>a</sup>	Treatment Integrity <sup>b</sup>	Treatment Usability (M)			
					Acceptability	Understanding	Collaboration	Feasibility
A	8	100	97.70%	92.86%	5.29	4.75	5.33	5.17
B	8	100	79.76%	100%	5.43	5.25	6.00	5.17
C	18	90	87.01%	100%	4.29	4.75	4.67	4.67
D	38	70	81.97%	89.29%	4.86	4.75	5.33	5.00
E	18	90	70.89%	100%	4.71	5.25	5.00	4.67
F	28	80	83.10%	88.10%	4.71	4.50	5.67	4.67
G	28	80	64.79%	88.57%	5.00	5.00	5.00	4.83

A = baseline phase length in days; B = intervention phase length in days; M = mean values. For treatment usability, higher scores are indicative of more extreme responding in the direction of the scale assessed.

<sup>a</sup>Treatment intensity as number of days the GBG was played based on the adjusted opportunities to play, expressed as percentage. <sup>b</sup>Treatment integrity rated by observer.



interactive elements. The in-service training featured input with video examples, discussion, and small group practice including feedback from both trainers and peers. In addition, the teachers received a detailed printed intervention manual with all steps and materials. Immediately before the start of the intervention phase, details regarding the introduction phase of the GBG and specific individual questions were clarified in a two-hour implementation planning meeting. During the intervention phase, we provided six brief appointments (15 to 20 minutes) for performance feedback spread over the entire intervention period following work shadowing, in each case combined with talks about pending issues and difficulties, and one appointment (1 hour) for a detailed consultation after ten weeks based on the results of the behavioral progress monitoring. Motivational interviewing/trouble shooting was offered two times on site by the trainers (1 hour) and additionally by email.

## Baseline

Teachers completed the paper-pencil-based DBR-SIS for each student each day at the end of the target instructional period. The assessment units were 10 minutes long.

## Intervention

The classic version of the GBG was played (Flower et al., 2014) with dividing the students in two or more teams to play against each other, marking “fouls” for rule-breaking, and rewarding the team with the fewest fouls as winner of the game. The teachers divided their class into five to six teams using the existing assigned group seating arrangements, with five to six students on each team. The teams were maintained throughout the intervention phase. Contrary to previous practices reported by Donaldson et al. (2017), no student was placed on his or her own team because of behavior, e.g. disruptive behavior. If

problems arose on a team, the teacher spoke with the team to find common solutions. After the teams had given themselves names (e.g., “Lions” and “The cool kids”), each team received a sign with the team name, which was affixed to the team table. A poster at the front of the classroom listed three to four rules that had been previously jointly compiled (e.g., “I am quiet” – “I work intently” – “I sit in my seat”). Moreover, a timer was placed near the poster to be clearly visible to all teams, and the names of teams were written on the board to mark the fouls. Respecting the specific classroom situation and the relationship with the students, the goals, rules, rule violations, and rewards slightly deviated across classrooms (e.g., rewarding with goodies such as chocolate or gummy bears, small prizes such as stickers or pens, or activities such as group games or extra reading time). After 10 minutes playing in the defined period, the teachers counted the fouls, named the winning team and delivered the reward to the winning team.

## Treatment integrity

At six measurement points, trained observers collected treatment integrity data on the accuracy with which the teacher implemented the GBG using a seven-item treatment integrity checklist adapted from the Treatment Integrity Planning Protocol (TIPP; Hagermoser Sanetti and Kratochwill, 2009) with dichotomous ratings (*agreement – disagreement*). In addition, the teachers completed an analogous treatment integrity checklist with additional options for comments (e.g., reasons for not playing the game or information on problems) and noted the number of fouls overall and per team on each day. The total adherence across all classrooms throughout the six measurement points was 94.12% (range = 88.1% to 100%). For these days, the interobserver agreement (observer – teacher) was 100%. Teachers’ ratings over the whole treatment period

showed that the overall treatment integrity measured by daily self-assessment was 96.15% (range = 85.5% to 100%).

### Treatment usability

Treatment usability from the teachers' points of view was assessed with the German version of the Usage Rating Profile (URP; Briesch et al., 2017) consisting of 20 items loading on four factors. The teachers indicated the extent to which they agreed with each of the items using a 6-point Likert scale (i.e., 1 = *strong disagreement*, 6 = *strong agreement*). The results of the URP revealed agreement regarding the usage of the GBG intervention across all teachers for each factor:  $M = 4.9$  ( $SD = 0.38$ ) for acceptability,  $M = 4.9$  ( $SD = 0.28$ ) for understanding,  $M = 5.2$  ( $SD = 0.45$ ) for home-school collaboration, and  $M = 4.88$  ( $SD = 0.23$ ) for feasibility. Treatment integrity and usability are shown in Table 2.

### Data analysis

To examine the impact of the GBG on AE and DB on each single case and across cases, descriptive statistics and inferential statistics were used to analyze the data. Therefore, after calculating phase means and effect sizes for each single case, we conducted regression analyses for each single case and across cases to replicate the results as well as to obtain statistical significance and overall quantification (Manolov and Moeyaert, 2017). The multilevel approach to estimate the overall effects was also used to analyze the interaction between students' psychosocial problems and the intervention effects. First, we reported descriptive statistics and calculated the non-rescaled non-overlap of all pairs (NAP; see Alresheed et al., 2013) with medium effects indicated by values of 66% to 92%, and strong effects indicated by values of 93% to 100%. Subsequently, we analyzed the data using a piecewise regression approach (Huitema and McKean, 2000). This procedure enables the control of developmental trends in the data (trend effects) and the differentiation between continuous (slope effect) and immediate (level effects) intervention effects. We conducted piecewise regressions for each single case and a multilevel extension (see Van den Noortgate and Onghena, 2003; Moeyart et al., 2014) for all cases with measurements at level 1 nested in subjects at level 2. The multilevel analyses were set up as a random intercept and random slope model with all three parameters (trend, slope, and level effects) as fixed and random factors. To test the significance of the random effects, we applied a likelihood ratio test for each random slope factor comparing the full model against a model without the target factor. To analyze the moderating effects of students' internalizing and externalizing problems on the intervention effects, we inserted cross-level interactions into the model. All analyses were conducted using R (R Core Team, 2018) and the scan package (Wilbert and Lüke, 2018).

## Results

### Descriptive statistics

Table 3 summarizes descriptive statistics, including the NAP as a common effect size for both AE and DB. The non-rescaled NAP indicated a medium or strong effect for both dependent variables for all participants, varying between 77.8% and 99.8% for AE and between 78.3% and 100% for DB. In detail, there were strong effects for 12 (34.29%) and medium effects for 23 (65.71%) participants regarding AE, and strong effects for 18 (51.54%) and medium effects for 17 (48.46%) participants regarding DB.

### Inferential statistics

We analyzed the data case by case, conducting piecewise regression analyses to calculate the impact of the GBG on AE and DB for each case. Subsequently, we used a multilevel extension to calculate the impact of the GBG for all cases with measurements at level 1 nested in subjects at level 2.

### Piecewise regression for each single case

Fourteen cases showed both significant level increases in AE and decreases in DB ( $p < 0.05$ ). Three participants (S2, S3, and S20) showed significant level increases only for AE ( $p < 0.05$ ), whereas 12 participants demonstrated significant level decreases only for DB ( $p < 0.05$ ). The slope effect was significant for increases in AE as well as decreases in DB for three participants ( $p < 0.05$  for S10, S26, and S33). Four participants (S19, S21, S29, and S35) demonstrated neither a significant level nor a significant slope effect for one of the dependent variables ( $p > 0.05$ ). The results for each case are shown in Table 4.

### Multilevel analyses

First, we conducted multilevel analyses of all single cases. Overall, for AE and DB, we found significant level effects. On average, AE increased by 0.74 points ( $p < 0.001$ ) on a 6-point Likert-type scale, and DB decreased by 1.29 points ( $p < 0.001$ ). Furthermore, the significant slope effect for DB indicated a decrease of 0.01 points ( $p = 0.002$ ) per measurement occasion. The similar increasing slope effect for AE failed to reach statistical significance ( $p = 0.055$ ). Second, we calculated the random effects regarding the variability between cases. We found significant level effects for both of the dependent variables. For AE, the estimated standard deviation for the level effect was  $SD = 0.65$  ( $p < 0.001$ ), whereas for DB, it was  $SD = 0.81$  ( $p < 0.001$ ). The estimated standard deviation for the slope effect was  $SD = 0.00$  ( $p = 1.000$ ) for AE and  $SD = 0.15$  ( $p = 0.034$ ) for DB. The results for both fixed and random effects are shown in Table 5.



TABLE 3 Descriptive statistics for the 35 single-cases for both dependent variables.

Case	$n_A$	$n_B$	Academic Engagement					Disruptive Behavior				
			$mis_A$	$mis_B$	$M_A (SD)$	$M_B (SD)$	NAP (%)	$mis_A$	$mis_B$	$M_A (SD)$	$M_B (SD)$	NAP (%)
S1	8	100	1	24	2.0 (1.2)	3.5 (1.1)	88.6	1	24	1.1 (0.4)	0.2 (0.4)	93.9
S2	8	100	1	17	2.1 (1.2)	3.6 (1.1)	86.9	1	18	1.1 (0.7)	0.4 (0.8)	83.8
S3	8	100	1	20	3.6 (0.8)	4.1 (0.8)	77.8	1	20	1.1 (0.7)	0.2 (0.6)	90.2
S4	8	100	1	20	3.4 (0.5)	4.7 (0.7)	95.9	1	21	2.3 (1.1)	0.1 (0.3)	99.4
S5	8	100	1	19	3.1 (1.1)	4.3 (0.8)	87.8	1	20	1.1 (0.9)	0.1 (0.3)	88.7
S6	8	100	1	36	1.9 (0.9)	3.5 (0.8)	94.2	1	38	2.1 (1.1)	0.0 (0.2)	99.8
S7	8	100	2	41	2.3 (1.2)	2.5 (1.2)	79.6	2	41	1.8 (0.8)	0.2 (0.4)	98.6
S8	8	100	1	38	1.9 (1.3)	3.2 (0.9)	88.9	1	38	2.1 (1.1)	0.0 (0.1)	99.9
S9	8	100	1	34	1.4 (0.8)	3.6 (0.8)	97.8	1	35	3.7 (0.5)	0.0 (0.3)	100.0
S10	8	100	2	41	2.2 (1.5)	3.3 (0.9)	88.5	2	42	1.0 (0.6)	0.0 (0.1)	96.1
S11	18	90	5	33	1.9 (0.3)	2.6 (0.6)	90.0	5	34	2.9 (0.9)	0.2 (0.7)	98.7
S12	18	90	5	43	3.8 (0.6)	4.4 (0.7)	89.5	5	43	2.7 (0.9)	0.2 (0.5)	99.3
S13	18	90	7	37	2.4 (0.5)	2.7 (0.5)	88.3	7	37	1.8 (0.6)	0.2 (0.5)	98.7
S14	18	90	5	31	2.4 (0.7)	3.0 (0.5)	88.1	5	32	1.8 (0.7)	0.1 (0.3)	99.4
S15	18	90	5	34	2.9 (0.3)	3.7 (0.5)	93.7	5	35	1.8 (0.7)	0.1 (0.4)	98.9
S16	38	70	11	23	3.7 (0.5)	4.1 (0.6)	84.5	11	23	1.5 (0.8)	0.8 (0.7)	86.9
S17	38	70	11	20	3.7 (0.7)	4.3 (0.8)	86.2	11	22	1.0 (0.7)	0.4 (0.5)	87.0
S18	38	70	13	28	4.0 (0.5)	4.7 (0.5)	91.1	13	29	0.7 (0.6)	0.1 (0.4)	90.6
S19	38	70	13	26	3.1 (0.8)	4.1 (0.8)	91.7	13	27	0.6 (0.8)	0.3 (0.5)	85.3
S20	38	70	11	20	3.5 (0.7)	4.7 (0.6)	94.2	11	22	0.1 (0.3)	0.0 (0.0)	78.3
S21	18	90	7	38	3.7 (0.6)	3.5 (0.7)	79.9	7	40	0.9 (0.7)	0.7 (0.8)	85.4
S22	18	90	4	36	3.6 (0.7)	4.4 (0.6)	88.9	5	38	1.4 (0.9)	0.6 (0.6)	90.3
S23	18	90	4	37	3.3 (0.6)	3.7 (0.9)	85.0	4	38	1.3 (0.6)	0.5 (0.6)	90.5
S24	18	90	4	43	4.0 (0.8)	4.4 (0.6)	84.9	4	45	1.3 (0.6)	0.6 (0.6)	90.7
S25	18	90	3	55	3.3 (0.8)	4.2 (0.6)	93.6	3	57	2.0 (0.7)	0.7 (0.6)	97.3
S26	28	80	12	29	3.3 (0.8)	3.3 (0.6)	82.2	12	29	2.0 (1.1)	1.7 (0.8)	84.5
S27	28	80	12	26	3.5 (0.6)	3.4 (0.7)	79.3	12	26	2.1 (1.0)	1.8 (0.9)	83.6
S28	28	80	13	29	2.1 (0.5)	2.7 (0.8)	90.4	13	29	3.1 (0.7)	2.0 (0.9)	94.0
S29	28	80	14	32	2.4 (0.6)	3.0 (0.9)	91.2	14	32	2.8 (0.8)	1.6 (0.9)	94.8
S30	28	80	16	30	3.2 (0.6)	3.9 (0.7)	93.6	16	30	2.1 (0.9)	0.6 (0.7)	96.7
S31	28	80	12	36	2.7 (1.3)	4.8 (0.6)	97.7	12	36	1.2 (1.4)	0.1 (0.4)	92.4
S32	28	80	12	36	3.2 (1.2)	4.9 (0.3)	97.4	12	36	1.1 (0.7)	0.0 (0.2)	96.5
S33	28	80	12	35	1.9 (1.1)	4.9 (0.4)	99.8	12	36	1.7 (1.2)	0.1 (0.3)	96.2
S34	28	80	12	36	4.2 (0.6)	4.9 (0.4)	93.8	12	36	0.6 (0.6)	0.0 (0.2)	92.5
S35	28	80	12	36	4.1 (0.8)	4.8 (0.6)	93.1	12	36	0.4 (0.6)	0.1 (0.5)	88.6

$n$  = number of data points, which was the same for all of the students within the same group in each classroom;  $mis$  = missing data points;  $M$  = mean,  $SD$  = standard deviation;  $NAP$  = non-rescaled non-overlap of all pairs. Missing data points resulted from the school and teacher factors (e.g., public and movable holidays, cancelled lessons, part-time employment, and illness) as well as student illness, incomplete DBR, etc.

### Interaction with students' psychosocial problems

We inserted cross-level interactions between level effects and the different SDQ subscales into the model. For AE, we did not find significant interactions with the level of the intervention phase for either the SDQ total score ( $B = 0.03$ ,  $p = 0.292$ ) or one of the subscales ( $B_{int} = 0.03$ ,  $p = 0.505$ ;  $B_{ext} = 0.03$ ,  $p = 0.423$ ). The analysis for DB showed a significant interaction for the externalizing subscale ( $B = -0.07$ ,  $p = 0.047$ ) and no significant interactions for the other scales ( $B_{total} = -0.05$ ,  $p = 0.076$ ;

$B_{int} = -0.02$ ,  $p = 0.927$ ). Table 6 contains the results for both dependent variables.

## Discussion

The purpose of the current study was to evaluate the impact of the GBG on at-risk students' AE and DB in an inclusive primary school in Germany using behavioral progress monitoring. Extending previous studies, we used

**TABLE 4** Piecewise regression analyses for the 35 single cases for both dependent variables.

Case	Academic Engagement			Disruptive Behavior		
	Trend	Level	Slope	Trend	Level	Slope
S1	-0.22	2.12*	0.22	0.08	-1.31*	-0.08
S2	-0.23	1.85*	0.24	0.08	-0.71	-0.09
S3	-0.08	1.26*	0.07	-0.16	-0.65	0.16
S4	-0.09	1.58*	0.09	0.12	-2.68*	-0.12
S5	0.03	0.82	-0.03	-0.3*	-0.06	0.3*
S6	0.23	0.72	-0.23	-0.23*	-1.34*	0.23*
S7	-0.15	0.14	0.16	-0.06	-1.5*	0.06
S8	-0.06	0.84	0.07	-0.13*	-1.64*	0.13*
S9	-0.16	2.4*	0.17	-0.09*	-3.33*	0.09*
S10	-0.36*	1.93*	0.36*	0.11*	-1.33*	-0.11*
S11	0.03	1.16*	-0.04	-0.07	-2.3*	0.07
S12	0.06	-0.59	-0.04	-0.05	-2.08*	0.05
S13	0	0.36	0	-0.04	-1.14*	0.04
S14	0.02	0.62*	-0.02	0.07*	-2.06*	-0.07*
S15	0.03	0.34	-0.02	-0.02	-1.45*	0.02
S16	-0.01	0.56*	0.01	0.02	-1.21*	-0.01
S17	-0.01	0.54	0.02	0	-0.73*	0.01
S18	0	0.53*	0	-0.01	-0.55*	0.01
S19	0.03*	-0.18	-0.02	-0.02	0.04	0.02
S20	0.01	0.74*	-0.01	0	-0.16	0
S21	-0.05	0.06	0.05	0.05	-1.03	-0.04
S22	-0.06	1.28*	0.05	0.11*	-1.94*	-0.11*
S23	0	0.11	0.01	0.02	-0.83*	-0.02
S24	-0.05	0.77*	0.05	0.04	-1.17*	-0.03
S25	-0.01	0.62	0.01	0.03	-0.99*	-0.04
S26	-0.05*	0.39	0.05*	0.07*	-1.4*	-0.06*
S27	-0.01	-0.15	0.01	-0.06*	0.36	0.06*
S28	-0.01	0.51	0.01	0	-1.21*	0
S29	-0.02	0.08	0.04	-0.01	-0.27	-0.01
S30	-0.03	0.52	0.04	0.02	-1.44*	-0.02
S31	0.01	1.88*	-0.01	0.05*	-1.77*	-0.05*
S32	-0.01	1.7*	0.01	0.01	-1.02*	-0.01
S33	-0.04*	3.51*	0.04*	0.06*	-2.38*	-0.06*
S34	-0.01	0.7*	0.02	0	-0.48*	0
S35	0	0.55	0	-0.03	0.1	0.02

\* = significant ( $p < 0.05$ ).

piecewise regression for each of the 35 single cases and a multilevel extension to examine both level and slope effects. Furthermore, we examined the interaction with students' psychosocial problems as potential influencing factors moderating the effectiveness.

## Main findings

The individual-level data analyses revealed that the majority of at-risk students benefited from the GBG. Whereas the non-rescaled NAP indicated a medium or strong effect for both dependent variables for all participants, the inferential statistics did not reveal statistically significant improvements for all cases. Piecewise regressions for each single case enabled us to identify

significant immediate treatment effects for 14 participants for both outcomes (40.0%), for three participants for AE only (8.57%), and for 12 participants for DB only (34.29%). However, six students (17.14%) showed no significant level effects for AE or DB. The results at the individual level support and extend the limited prior research, indicating that the classroom-based GBG intervention is effective for improving at-risk students' classroom behaviors, but there are students who do not respond to the intervention (Hunt, 2012; Donaldson et al., 2017; Moore et al., 2022). Consistent with previous research reporting class-wide data (Flower et al., 2014; Bowman-Perrott et al., 2016), the multilevel approach revealed that the GBG improved AE and reduced DB for students with challenging behavior with a significant immediate treatment effect across cases. In line with the meta-analysis by Bowman-Perrott et al. (2016), the GBG was more effective in reducing DB than increasing AE. Similar to Flower et al. (2014), we found statistically significant slightly decreasing DB throughout the intervention phase. In contrast to the single case studies by Donaldson et al. (2017) and Moore et al. (2022), in which some participants with or at risk for EBD showed slightly increasing trends over the course of the intervention, this slope effect indicates a continuous and decreasing change in DB.

The findings from the present study extend prior meta-analysis results (Bowman-Perrott et al., 2016) by investigating students' psychosocial risks as potential moderators for the impact of the intervention. We found significant moderating effects for students' externalizing problems on the intervention effect for DB, meaning that the higher students' externalizing problems were, the more effective the GBG was. These findings correspond with the results of reviews on the subject of interventions for aggressive behavior (Waschbusch et al., 2019) as well as longitudinal studies evidencing the GBG as an effective

**TABLE 5** Fixed and random effects of the multilevel piecewise-regression models for academic engagement and disruptive behavior.

Parameter	Fixed effects				Random effects		
	B	SE	t	p	Estimated SD	L	p
<b>Academic Engagement</b>							
Intercept	3.02	0.14	21.83	< 0.001**	0.74	109.59	< 0.001**
Trend	0.00	0.00	-0.92	0.358	0.01	1.94	0.746
Level Phase B	0.74	0.13	5.86	< 0.001**	0.65	65.71	< 0.001**
Slope Phase B	0.01	0.00	1.92	0.055	0.00	0.02	1.000
<b>Disruptive Behavior</b>							
Intercept	1.50	0.13	11.63	< 0.001**	0.71	210.92	< 0.001**
Trend	0.01	0.00	3.19	0.001**	0.02	12.90	0.012**
Level Phase B	-1.29	0.15	-8.77	< 0.001**	0.81	138.84	< 0.001**
Slope Phase B	-0.01	0.00	-3.18	0.002**	0.15	10.43	0.034*

\* = significant ( $p < 0.05$ ); \*\* = significant ( $p < 0.001$ ).

TABLE 6 Interaction of psychosocial problems with the impact of the good behavior game.

Scale	B	SE	df	t	p	B	SE	df	t	p
	Academic Engagement					Disruptive Behavior				
<b>SDQ total</b>										
Intercept	2.94	0.12	2376	24.25	< 0.001**	1.61	0.12	2349	12.24	< 0.001**
Trend	0.00	0.00	2376	3.72	< 0.001**	0.00	0.00	2349	0.30	0.766
Level Phase B	0.67	0.17	2376	5.30	< 0.001**	-1.20	0.13	2349	-9.33	< 0.001**
SDQ total	-0.06	0.03	33	-2.22	0.034	0.04	0.03	33	1.40	0.172
Level Phase B*SDQ total	0.03	0.03	2376	1.06	0.292	-0.05	0.03	2349	-1.77	0.076
<b>SDQ int</b>										
Intercept	2.94	0.13	2376	22.99	< 0.001**	1.61	0.14	2349	11.76	< 0.001**
Trend	0.00	0.00	2376	3.70	< 0.001**	0.00	0.00	2349	0.34	0.738
Level Phase B	0.67	0.13	2376	5.32	< 0.001**	-1.20	0.05	2349	-9.15	< 0.001**
SDQ int	-0.05	0.05	33	-0.94	0.356	-0.01	0.05	33	-0.09	0.927
Level Phase B*SDQ int	0.03	0.05	2376	0.67	0.505	-0.02	0.05	2349	-0.39	0.697
<b>SDQ ext</b>										
Intercept	2.94	0.12	2376	23.88	< 0.001**	1.61	0.13	2349	12.60	< 0.001**
Trend	0.00	0.00	2376	3.70	< 0.001**	0.00	0.00	2349	0.30	0.766
Level Phase B	0.67	0.13	2376	5.37	< 0.001**	-1.20	0.13	2349	-9.41	< 0.001**
SDQ ext	-0.07	0.04	33	-1.98	0.056	0.07	0.04	33	1.93	0.062
Level Phase B*SDQ ext	0.03	0.03	2376	0.80	0.423	-0.07	0.04	2349	-1.99	0.047*

SDQ total = SDQ total score; SDQ int = SDQ subscale internalizing problems; SDQ ext = SDQ subscale externalizing problems; \* = significant ( $p < 0.05$ ); \*\* = significant ( $p < 0.001$ ).

intervention to reduce aggressive behavior for children with high risks in general (van Lier et al., 2005) and in boys with persistent high risks (Kellam et al., 2008). However, even though the majority of students with externalizing risks benefited from the intervention, three of the non-responders showed high risks assessed by their teachers. This finding leads to the assumption that for children with high externalizing risks, further individual factors moderate the effectiveness of group contingencies (Maggin et al., 2017). No effects were found for internalizing risks. Considering that the majority of the students nominated by teachers showed no internalizing problems, we cannot deduce effects for students with internalizing risks from our study.

Overall, our findings indicated that the impact of the classroom-based GBG program varied as a function of individual children. For both outcomes, and in particular for AE, there was large variability between individuals. There are several explanations for this finding. Although no clear pattern emerges in our data, it is possible that aspects of treatment integrity and usability may have affected the outcomes, particularly among the non-responders (Moore et al., 2022). From a methodological point of view, for some students in our sample, the high AE values at baseline led to minimal room for improvement, suggesting a possible ceiling effect (Ho and Yu, 2015). Likewise, there might have been floor effects for some of the students with low DB at baseline. Furthermore, it is important to consider the situation in which the GBG was played. It is possible that the target situation was not the most difficult part of the lesson for all of the nominated students; thus, their AE and, in part, their DB might not have been as problematic as usual. In addition,

struggling with learning strategies, as a common problem of students with challenging behavior (Kauffman and Landrum, 2012), could affect the effectiveness of the GBG regarding AE. The implementation of additional components such as self-monitoring strategies (Bruhn et al., 2015) could be necessary to increase AE for non-responding students (Smith et al., 2021).

Interestingly, we could not find any interaction of externalizing problems and intervention effectiveness for AE. In addition to the aforementioned possible ceiling effects, the similarity of the measured constructs must be considered. Although hyperactivity and conduct problems can negatively impact school functioning and academic performance (Mundy et al., 2017), the externalizing subscales of the SDQ as a standardized measure for assessing child mental health problems are more closely linked to disruptive classroom behaviors than to academic engagement as a typical school-related construct. Further associated factors, such as psychological and cognitive dimensions of engagement, including having sense of belonging or motivational beliefs (Wang and Eccles, 2013) or the level of academic enabling skills (Fabiano and Pyle, 2019), could moderate the intervention effects on AE.

## Limitations

The findings from this study should be interpreted by considering several potential limitations. First, not all of the nominated students in our sample had high ratings in the externalizing and/or internalizing subscales of the SDQ.

Therefore, it can be assumed that not all of the nominated students were at risk for EBD. Although research has indicated that teachers are competent in identifying students in their classrooms with problem behaviors (Lane and Menzies, 2005), we believe that in addition to teacher nomination, future research should use other methods to identify students with challenging behavior (i.e., systematic behavioral assessment in the baseline).

Second, teachers both delivered the intervention and rated students' performance. The 'double burden' of teaching and rating as well as the teachers' acceptance of the intervention could affect their ratings. However, despite limited associations between behavioral change and acceptability, research has demonstrated the sensitivity of the DBR-SIS completed by implementing teachers (Chafouleas et al., 2012; Smith et al., 2018). Furthermore, we were unable to conduct systematic direct observations by trained observers or video recordings. As such, we decided to use the DBR-SIS as an efficient tool with acceptable reliability, validity and sensitivity within our aims.

Third, our dependent variables were broad categories combining different behaviors. These target behaviors enabled us to compare our results, particularly with the findings of the existing meta-analysis. On the other hand, students' individual changes in specific behaviors could not be tested. Furthermore, we only investigated one potential moderator for the impact of the intervention. However, the differentiated analysis of these risks for two key aspects of classroom behaviors extends previous research independently of the need for further investigation.

## Implications for research and practice

Analyzing the responses to the GBG using behavioral progress monitoring helps to identify students who need additional support on tier 2 in a MTSS and to recognize at an early stage if positive effects are decreasing over time for individual students (Donaldson et al., 2017; Moore et al., 2022). The selected combination of data analysis methods enables precise alignment with our aims and considers the characteristics of the data (Manolov and Moeyaert, 2017): Due to our large sample and the analytical method chosen, we were able to investigate the impact of the GBG on individual at-risk student classroom behaviors as well as its effectiveness across cases. We therefore believe that our study is a methodologically sound study about the GBG that could be used to substantiate the evidence classification of the GBG. Considering the fact that the single-case studies of the GBG work with small samples and use methods for data analysis that do not address the problem of autocorrelation, misestimates of effectiveness due to the methods chosen are plausible (Shadish et al., 2014), and the number of non-responders tends to be underestimated. The number of non-responders indicates

that we should be very careful with the transfer of results of studies investigating a class or a group as a single-case to the behavioral development of individual students with challenging behavior. Therefore, further single-case research with larger-than-usual samples and meta-analytical approaches are necessary to extend the findings regarding the impact of the GBG on at-risk students' classroom behaviors as well as further potential moderators. As shown in our study, externalizing risks seem to moderate the impact of the GBG on reducing disruptive behavior. Thus, our findings imply that in future intervention studies, the effects should be controlled for possible influences of externalizing risks. Furthermore, in addition to potential individual factors, functional characteristics (Maggini et al., 2017), environmental moderators, such as the classroom level of aggression (Waschbusch et al., 2019), peer factors, and school climate (Farrell et al., 2013), should be examined.

Overall, the results of our study suggest that the GBG facilitates at-risk students' behavioral development in inclusive settings in Germany. In particular, in our sample, students who were assessed by teachers as exhibiting high externalizing behavior problems can benefit. These results should encourage teachers to implement this classroom-based intervention and to monitor its effect using behavioral progress monitoring. Our findings also suggest that a combination of the behavioristic method of the GBG with cognitive methods such as self-monitoring (Bruhn et al., 2015) could be necessary to enable the effects regarding AE. In our study, treatment integrity was high, and teachers assessed the intervention as suitable for their daily work. However, whether the intervention can be sustainably implemented depends on several factors. Despite the simple rules of the game, coaching throughout the implementation and positive impacts, teachers do not see sufficient possibilities to integrate the GBG naturally in their daily work (Coombes et al., 2016). Furthermore, the teachers in our study likewise reported that they did not find the time to play the GBG due to school events or learning projects. In addition, some found it difficult to consistently implement behavioral progress monitoring. To increase sustainability, the maintenance of both the GBG and progress monitoring should be carefully planned and monitored. If this is successful, the GBG, in conjunction with behavioral progress monitoring, is an appropriate classroom-based intervention to improve at-risk students' classroom behaviors and to adjust students' supports data-based at an early stage.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

Ethical review and approval were not required in accordance with the local legislation and institutional requirements. Following the school law and the requirements of the ministry of education of the federal state North Rhine-Westphalia (Schulgesetz für das Land Nordrhein-Westfalen), school administrators decided in co-ordination with their teachers about participation in this scientific study. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements. Verbal informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

MG, GC, RV, JW, and TL: methodology. JW: software. JW and TL: formal analysis. GC and TL: investigation. TH, GC, and TL: resources. TL, GC, and JW: data curation. TL: writing – original draft and visualization. TH and RV: supervision. GC and TL: project administration. All authors: conceptualization and writing – review and editing.

## Funding

This study was supported by the Marbach Residency Program (Jacobs Foundation). We acknowledge support for the

## References

- Alresheed, F., Hott, B. L., and Bano, C. (2013). Single Subject Research: a Synthesis of Analytic Methods. *JOSEA* 2, 1–18.
- Barkmann, C., and Schulte-Markwort, M. (2012). Prevalence of Emotional and Behavioural Disorders in German Children and Adolescents: a Meta-Analysis. *J. Epidemiol. Community Health* 66, 194–203. doi: 10.1136/jech.2009.102467
- Barrish, H. H., Saunders, M., and Wolf, M. M. (1969). Good Behavior Game: Effects of Individual Contingencies for Group Consequences on Disruptive Behavior in a Classroom. *J. Appl. Behav. Anal.* 2, 119–124. doi: 10.1901/jaba.1969.2-119
- Barth, J. M., Dunlap, S. T., Dane, H., Lochman, J. E., and Wells, K. C. (2004). Classroom Environment Influences on Aggression, Peer Relations, and Academic Focus. *J. Sch. Psychol.* 42, 115–133. doi: 10.1016/j.jsp.2003.11.004
- Batsche, G. (2014). "Multi-Tiered Systems of Supports for Inclusive Schools," in *Handbook of Effective Inclusive Schools*, eds J. McLeskey, N. L. Waldron, F. Spooner, and B. Algozzine (New York, NY: Routledge), 183–196.
- Bettge, S., Ravens-Sieberer, U., Wietzker, A., and Hölling, H. (2002). Ein Methodenvergleich der Child Behavior Checklist und des Strengths and Difficulties Questionnaire. [Comparison of Methods Between the Child Behavior Checklist and the Strengths and Difficulties Questionnaire]. *Gesundheitswesen* 64, 119–124. doi: 10.1055/s-2002-39264
- Bowman-Perrott, L., Burke, M. D., Zaini, S., Zhang, N., and Vannest, K. (2016). Promoting Positive Behavior Using the Good Behavior Game: a Meta-Analysis of Single-Case Research. *J. Posit. Behav. Interv.* 18, 180–190. doi: 10.1177/1098300715592355
- Briesch, A. M., Casale, G., Grosche, M., Volpe, R. J., and Hennemann, T. (2017). Initial Validation of the Usage Rating Profile-Assessment for Use Within German Language Schools. *Learn. Disab.* 15, 193–207.
- Briesch, A. M., Chafouleas, S. M., and Riley-Tillman, T. C. (2010). Generalizability and Dependability of Behavior Assessment Methods to Estimate Academic Engagement: a Comparison of Systematic Direct Observation and Direct Behavior Rating. *School Psych. Rev.* 39, 408–421.
- Bruhn, A., McDaniel, S., and Kreigh, C. (2015). Self-Monitoring Interventions for Students With Behavior Problems: a Systematic Review of Current Research. *Behav. Disord.* 40, 102–121. doi: 10.17988/BD-13-45.1
- Casale, G., Grosche, M., Volpe, R. J., and Hennemann, T. (2017). Zuverlässigkeit von Verhaltensverlaufsdiagnostik über Rater und Messzeitpunkte bei Schülern mit externalisierenden Verhaltensproblemen. [Dependability of Direct Behavior Ratings Across Rater and Occasion in Students with Externalizing Behavior Problems]. *Empirische Sonderpädagogik* 9, 143–164.
- Casale, G., Hennemann, T., Volpe, R. J., Briesch, A. M., and Grosche, M. (2015). Generalisierbarkeit und Zuverlässigkeit von Direkten Verhaltensbeurteilungen des Lern- und Arbeitsverhaltens in einer inklusiven Grundschulklasse. [Generalizability and Dependability of Direct Behavior Ratings of Academically Engaged Behavior in an Inclusive Classroom Setting]. *Empirische Sonderpädagogik* 7, 258–268.
- Chaffee, R., Briesch, A. M., Johnson, A., and Volpe, R. J. (2017). A Meta-Analysis of Class-Wide Interventions for Supporting Student Behavior. *School Psych. Rev.* 46, 149–164. doi: 10.17105/SPR-2017-0015.V46-2

Article Processing Charge from the DFG (German Research Foundation, 491454339).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2022.917138/full#supplementary-material>

- Chafouleas, S. M. (2011). Direct Behavior Rating: a Review of the Issues and Research in Its Development. *Educ. Treat. Child.* 34, 575–591. doi: 10.1353/etc.2011.0034
- Chafouleas, S. M., Hagermoser Sanetti, L. M., Kilgus, S. P., and Maggin, D. M. (2012). Evaluating Sensitivity to Behavioral Change Using Direct Behavior Rating Single Item Scales. *Except. Child* 78, 491–505. doi: 10.1177/00144029120780406
- Christ, T. J., Riley-Tillman, T. C., and Chafouleas, S. M. (2009). Foundation for the Development and Use of Direct Behavior Rating (DBR) to Assess and Evaluate Student Behavior. *Assess. Eff. Interv.* 34, 201–213. doi: 10.1177/1534508409340390
- Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., et al. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remed. Spec. Educ.* 36, 220–234. doi: 10.1177/0741932514557271
- Coombes, L., Chan, G., Allen, D., and Foxcroft, D. R. (2016). Mixed-Methods Evaluation of the Good Behavior Game in English Primary Schools. *J. Com. Appl. Social Psych.* 26, 369–387. doi: 10.1002/casp.2268
- DeVries, J. M., Gebhardt, M., and Voß, S. (2017). An Assessment of Measurement Invariance in the 3- and 5-Factor Models of the Strengths and Difficulties Questionnaire: new Insights from a Longitudinal Study. *Pers. Individ. Diff.* 119, 1–6. doi: 10.1016/j.paid.2017.06.026
- Donaldson, J. M., Fisher, A. B., and Kahng, S. W. (2017). Effects of the Good Behavior Game on Individual Student Behavior. *Behav. Anal.* 17, 207–216. doi: 10.1037/bar0000016
- Fabiano, G. A., and Pyle, K. (2019). Best Practices in School Mental Health for Attention-Deficit/Hyperactivity Disorder: a Framework for Intervention. *School Ment. Health* 11, 72–91. doi: 10.1007/s12310-018-9267-2
- Farrell, A. D., Henry, D. B., and Bettencourt, A. (2013). Methodological Challenges Examining Subgroup Differences: Examples from Universal School-Based Youth Violence Prevention Trials. *Prev. Sci.* 14, 121–133. doi: 10.1007/s11212-011-0200-2
- Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., and Vega, R. (2014). Effects of the Good Behavior Game on Challenging Behaviors in School Settings. *Rev. Educ. Res.* 84, 546–571. doi: 10.3102/0034654314536781
- Foley, E. A., Dozier, C. L., and Lessor, A. L. (2019). Comparison of Components of the Good Behavior Game in a Preschool Classroom. *J. Appl. Behav. Anal.* 52, 84–104. doi: 10.1002/jaba.506
- Gebhardt, M., DeVries, J., Jungjohann, J., Casale, G., Gegenfurtner, A., and Kuhn, T. (2019). Measurement Invariance of a Direct Behavior Rating Multi Item Scale across Occasions. *Soc. Sci.* 8:46. doi: 10.3390/socsci8020046
- Goodman, A., Lamping, D. L., and Ploubidis, G. B. (2010). When to Use Broader Internalising and Externalising Subscales Instead of the Hypothesised Five Subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British Parents. *Teach. Child. J. Abnorm. Child Psychol.* 38, 1179–1191. doi: 10.1007/s10802-010-9434-x
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a Research Note. *J. Child. Psychol. Psychiatry* 38, 581–586. doi: 10.1111/j.1469-7610.1997.tb01545
- Groves, E. A., and Austin, J. L. (2017). An Evaluation of Interdependent and Independent Group Contingencies During the Good Behavior Game. *J. Appl. Behav. Anal.* 50, 552–566. doi: 10.1002/jaba.393
- Hagermoser Sanetti, L. M., and Kratochwill, T. R. (2009). Treatment Integrity Assessment in the Schools: an Evaluation of the Treatment Integrity Planning Protocol. *Sch. Psychol. Q.* 24, 24–35. doi: 10.1037/a0015431
- Hagermoser Sanetti, L. M., Collier-Meek, M. A., Long, A. C. J., Kim, J., and Kratochwill, T. R. (2014). Using Implementation Planning to Increase Teachers' Adherence and Quality to Behavior Support Plans. *Psychol. Sch.* 51, 879–895. doi: 10.1002/pits.21787
- Hanisch, C., Casale, G., Volpe, R. J., Briesch, A. M., Richard, S., Meyer, H., et al. (2019). Gestufte Förderung in der Grundschule: Konzeption eines mehrstufigen, multimodalen Förderkonzeptes bei expansivem Problemverhalten. [Multitiered system of support in primary schools: Introducing a multistage, multimodal concept for the prevention of externalizing behavior problems]. *Präv. Gesundheitsf.* 14, 237–241. doi: 10.1007/s11553-018-0700-z
- Ho, A. D., and Yu, C. C. (2015). Descriptive Statistics for Modern Test Score Distributions: skewness, Kurtosis, Discreteness, and Ceiling Effects. *Educ. Psychol. Meas.* 75, 365–388. doi: 10.1177/0013164414548576
- Huitema, B. E., and McKean, J. W. (2000). Design Specification Issues in Time-Series Intervention Models. *Educ. Psychol. Meas.* 60, 38–58. doi: 10.1177/00131640021970358
- Hunt, B. M. (2012). *Using the Good Behavior Game to Decrease Disruptive Behavior While Increasing Academic Engagement with a Head Start Population*. Unpublished doctoral dissertation. Hattiesburg: University of Southern Mississippi.
- Joslyn, P. R., Donaldson, J. M., Austin, J. L., and Vollmer, T. R. (2019). The Good Behavior Game: a Brief Review. *J. Appl. Behav. Anal.* 52, 811–815. doi: 10.1002/jaba.572
- Kauffman, J. M., and Landrum, T. J. (2012). *Characteristics of Emotional and Behavioral Disorders of Children and Youth* (10th ed.). Upper Saddle River, NJ: Pearson.
- Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., et al. (2008). Effects of a Universal Classroom Behavior Management Program in First and Second Grades on Young Adult Behavioral, Psychiatric, and Social Outcomes. *Drug Alcohol Depend.* 95, 5–28. doi: 10.1016/j.drugalcdep.2008.01.004
- Klipker, K., Baumgarten, F., Göbel, K., Lampert, T., and Hölling, H. (2018). Psychische Auffälligkeiten bei Kindern und Jugendlichen in Deutschland – Querschnittergebnisse aus KiGGS Welle 2 und Trends. [Mental Health Problems in Children and Adolescents in Germany – Results of the Cross-Sectional KiGGS Wave 2 Study and Trends]. *J. Health Monit.* 3, 37–45. doi: 10.17886/RKI-GBE-2018-077
- KMK (2011). *Inklusive Bildung von Kindern und Jugendlichen mit Behinderungen in Schulen (Beschluss vom 20.10.2011)*. [Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Federal Republic of Germany – Inclusive Education of Children and Adolescents with Disabilities in Schools of October 20, 2011]. Available online at: <https://www.kmk.org/themen/allgemeinbildende-schulen/inklusion.html> (accessed date October 20, 2011)
- Kratochwill, T. R., and Levin, J. R. (2010). Enhancing the Scientific Credibility of Single-Case Intervention Research: Randomization to the Rescue. *Psychol. Methods* 15, 124–144. doi: 10.1037/a0017736
- Lane, K. L., and Menzies, H. M. (2005). Teacher-Identified Students with and without Academic and Behavioral Concerns: Characteristics and Responsiveness. *Behav. Disord.* 31, 65–83. doi: 10.1177/019874290503100103
- Lobo, M. A., Moeyaert, M., Baraldi Cunha, A., and Babik, I. (2017). Single-Case Design, Analysis, and Quality Assessment for Intervention Research. *J. Neurol. Phys. Ther.* 41, 187–197. doi: 10.1097/NPT.0000000000000187
- Maggin, D. M., Pustejovsky, J. E., and Johnson, A. H. (2017). A Meta-Analysis of School-Based Group Contingency Interventions for Students with Challenging Behavior: an Update. *Remedial Spec. Educ.* 38, 353–370. doi: 10.1177/0741932517116900
- Manolov, R., and Moeyaert, M. (2017). Recommendations for Choosing Single-Case Data Analytical Techniques. *Behav. Ther.* 48, 97–114. doi: 10.1016/j.beth.2016.04.008
- Medland, M. B., and Stachnik, T. J. (1972). Good-Behavior Game: a Replication and Systematic Analysis. *J. Appl. Behav. Anal.* 5, 45–51. doi: 10.1901/jaba.1972.5-45
- Moeyart, M., Ferron, J., Beretvas, S., and Van den Noortgate, W. (2014). From a Single-Level Analysis to a Multilevel Analysis of Single-Case Experimental Designs. *J. Sch. Psychol.* 52, 191–211. doi: 10.1016/j.jsp.2013.11.003
- Moore, T. C., Gordon, J. R., Williams, A., and Eshbaugh, J. F. (2022). A Positive Version of the Good Behavior Game in a Self-Contained Classroom for EBD: effects on Individual Student Behavior. *Behav. Disord.* 47, 67–83.
- Mundy, L. K., Canterford, L., Tucker, D., Bayer, J., Romaniuk, H., Sawyer, S., et al. (2017). Academic Performance in Primary School Children with Common Emotional and Behavioral Problems. *J. Sch. Health* 87, 593–601. doi: 10.1111/josh.12531
- Pennington, B., and McComas, J. J. (2017). Effects of the Good Behavior Game Across Classroom Contexts. *J. Appl. Behav. Anal.* 50, 176–180. doi: 10.1002/jaba.357
- Poduska, J. M., and Kurki, A. (2014). Guided by Theory, Informed by Practice: Training and Support for the Good Behavior Game, a Classroom-Based Behavior Management Strategy. *J. Emot. Behav. Disord.* 22, 83–94. doi: 10.1177/1063426614522692
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.
- Saile, H. (2007). Psychometrische Befunde zur Lehrerversion des „Strengths and Difficulties Questionnaire“ (SDQ-L). Eine Validierung anhand soziometrischer Indizes. [Psychometric findings of the teacher version of “Strengths and Difficulties Questionnaire” (SDQ-L). Validation by means of socio-metric indices]. *Z. Entwicklungspsychol. Pädagog. Psychol.* 39, 25–31. doi: 10.1055/s-0033-1343321
- Schulgesetz für das Land Nordrhein-Westfalen (Schulgesetz NRW – SchulG) vom 15. Februar 2005 (GV. NRW. S. 102) zuletzt geändert durch Gesetz vom 23. Februar 2022 (GV. NRW. 2022 S. 250). [School Law for the State of North Rhine-Westphalia

(Schulgesetz NRW - SchulG) of February 15, 2005 (GV. NRW. p. 102) last amended by the Act of February 23, 2022 (GV. NRW. 2022 p. 250)].

Shadish, W. R., and Sullivan, K. J. (2011). Characteristics of Single-Case Designs Used to Assess Intervention Effects in 2008. *Behav. Res.* 43, 971–980. doi: 10.3758/s13428-011-0111-y

Shadish, W. R., Hedges, L. V., Pustejovsky, J. E., Rindskopf, D. M., Boyajian, J. G., and Sullivan, K. J. (2014). “Analyzing Single-Case Designs: d, G, Hierarchical Models, Bayesian Estimators, Generalized Additive Models, and the Hopes and Fears of Researchers about Analyses,” in *Single-Case Intervention Research. Methodological and Statistical Advances*, eds T. R. Kratochwill and R. L. Levin (Washington DC: American Psychological Association), 247–281.

Shadish, W. R., Rindskopf, D. M., Hedges, L. V., and Sullivan, K. J. (2013). Bayesian Estimates of Autocorrelations in Single-Case Designs. *Behav. Res. Methods* 45, 813–821. doi: 10.3758/s13428-012-0282-1

Simonsen, B., and Myers, D. (2015). *Classwide Positive Behavior Interventions and Supports. A Guide to Proactive Classroom Management*. New York, NY: Guilford Press.

Smith, R. L., Eklund, K., and Kilgus, S. P. (2018). Concurrent Validity and Sensitivity to Change of Direct Behavior Rating Single-Item Scales (DBR-SIS) within an Elementary Sample. *Sch. Psychol. Q.* 33, 83–93. doi: 10.1037/spq0000209

Smith, S., Barajas, K., Ellis, B., Moore, C., McCauley, S., and Reichow, B. (2021). A Meta-Analytic Review of Randomized Controlled Trials of the Good Behavior Game. *Behav. Modif.* 45, 641–666. doi: 10.1177/0145445519878670

Spilt, J. L., Koot, J. M., and van Lier, P. A. C. (2013). For Whom Does it Work? Subgroup Differences in the Effects of a School-Based Universal Prevention Program. *Prev. Sci.* 14, 479–488. doi: 10.1007/s11121-012-0329-7

Tanol, G., Johnson, L., McComas, J., and Cote, E. (2010). Responding to Rule Violations or Rule Following: a Comparison of Two Versions of the Good

Behavior Game with Kindergarten Students. *J. Sch. Psychol.* 48, 337–355. doi: 10.1016/j.jsp.2010.06.001

Van den Noortgate, W., and Onghena, P. (2003). Combining single case experimental data using hierarchical linear modeling. *Sch. Psychol. Q.* 18, 325–346. doi: 10.1521/scpq.18.3.325.22577

van Lier, P. A. C., Vuijk, P., and Crijnen, A. A. M. (2005). Understanding Mechanisms of Change in the Development of Antisocial Behavior: The Impact of a Universal Intervention. *J. Abnorm. Child. Psychol.* 33, 521–535. doi: 10.1007/s10802-005-6735-7

Volpe, R. J., Casale, G., Mohiyeddini, C., Grosche, M., Hennemann, T., Briesch, A. M., et al. (2018). A Universal Screener Linked to Personalized Classroom Interventions: Psychometric Characteristics in a Large Sample of German Schoolchildren. *J. Sch. Psychol.* 66, 25–40. doi: 10.1016/j.jsp.2017.11.003

Voß, S., Blumenthal, Y., Mahlau, K., Marten, K., Diehl, K., Sikora, S., et al. (2016). *Der Response-to-Intervention-Ansatz in der Praxis. Evaluationsergebnisse zum Rügener Inklusionsmodell*. [The Response-to-Intervention Approach in Practice. Evaluation Results on the Rügen Inclusion Model]. Münster: Waxmann.

Wang, M.-T., and Eccles, J. S. (2013). School Context, Achievement Motivation, and Academic Engagement: a Longitudinal Study of School Engagement Using a Multidimensional Perspective. *Learn. Instr.* 28, 12–23. doi: 10.1016/j.learninstruc.2013.04.002

Waschbusch, D. A., Breaux, R. P., and Babinski, D. E. (2019). School-Based Interventions for Aggression and Defiance in Youth: a Framework for Evidence-Based Practice. *School Ment. Health* 11, 92–105. doi: 10.1007/s12310-018-9269-0

Wilbert, J., and Lüke, T. (2018). *Scan: Single-Case Data Analyses for Single and Multiple Baseline Designs*. Available at: <https://r-forge.r-project.org/projects/scan> (accessed date 2016-06-23).

Wiskow, K. M., Ruiz-Olivares, R., Matter, A. L., and Donaldson, J. M. (2018). Evaluation of the Good Behavior Game with a Child with Fetal Alcohol Syndrome in a Small-Group Context. *Behav. Interv.* 33, 150–159. doi: 10.1002/bin.1515