

# **Playing the Good Behavior Game During a Peer-Tutoring Intervention: Effects on Behavior and Reading Fluency of Tutors and Tutees With Behavioral Problems**

**Markus Spilles**

**Tobias Hagen**

**Thomas Hennemann**

*University of Cologne, Germany*

---

*Children with externalized behavioral problems (BP) are at increased risk of being affected by a lack of social skills and academic engagement, and reading difficulties can frequently occur. Peer-tutoring (PT) interventions demonstrably promote both positive behavior and academic outcomes. To foster these effects, using interdependent group-reward contingencies appears to be a promising approach – especially for children with BP. In the current single-case study, a paired and repeated reading strategy was supplemented with the German version of the Good Behavior Game (GBG). A multiple-baseline design was used for evaluation. The study examined how the academic engagement, respectful behavior, disruptive behavior and reading fluency of four tutors and four tutees with BP (second and third grades) developed when the GBG was played in addition to PT. The results reveal that the tutors and tutees showed more academic engagement and less disruptive behavior. With regard to the tutees' reading fluency, significant small to medium effects can be attributed to the GBG.*

---

**Keywords:** Peer Tutoring, Good Behavior Game, Reading Fluency, Externalized Behavioral Problems, Primary School Teaching, Inclusive Education

## **INTRODUCTION**

### ***Problem Areas of Students with Externalized Behavioral Problems***

The school situation of students with externalized behavioral problems (BP) can be assessed as very problematic. Students with BP are characterized by oppositional, aggressive, destructive, impulsive and hyperkinetic behavior, which can lead to considerable suffering and significant impairments in different areas of life (Döpfner, 2013). For example, studies show that there are serious learning disabilities that affect up to 50 % of this group (Reid, Gonzalez, Nordness, Trout, & Epstein, 2004; Klauer & Lauth, 1997; Visser, Büttner, &

*Insights into Learning Disabilities* is published by Learning Disabilities Worldwide (LDW). For further information about learning disabilities, LDW's many other publications and membership, please visit our website: [www.ldworldwide.org](http://www.ldworldwide.org).

Hasselhorn, 2018). Reading difficulties, in particular, appear to frequently occur (Fischbach, Schuchardt, Mähler, & Hasselhorn, 2010; Klicpera, Schabmann, Gasteiger-Klicpera, & Schmidt, 2017; Visser et al., 2018; Morgan, Farkas, Tufis, & Sperling, 2008). Furthermore, academic failure, as well as classroom disruptions and conflicts with fellow students, can lead to reduced social integration in inclusive school settings (Newcomb, Bukowski, & Pattee, 1993; Krull, Wilbert, & Hennemann, 2014; Huber & Wilbert, 2012). The comorbidity of these problems indicates a poor prognosis for students with BP regarding their school career and future life. In addition, dealing with BP often results in high stress for (inclusive) teachers (Avramidis & Norwich, 2002). According to the results of the BELLA study (Klasen, Meyrose, Otto, Reiss, & Ravens-Sieberer, 2017) between 4.5 % and 15.6 % of students show clinically significant symptoms of externalizing disorders (attention-deficit/hyperactivity disorder: 4.5 - 11.5 %; conduct disorder: 12.1 - 15.6 %) depending on age and gender. Therefore, approximately one to two out of ten children in school classes can be designated as behaviorally problematic – regardless of a categorical diagnosis. To address these challenges and promote academic learning as well as positive behavior of students with BP, teachers must know about effective and practical interventions, especially for teaching in inclusive settings.

### ***Peer Tutoring***

A promising approach to fostering both academic skills and positive behavior is peer tutoring (PT). PT can be described as “a class of practices and strategies that employ peers as one-on-one teachers to provide individualized instruction, practice, repetition, and clarification of concepts” (Utley & Mortweet, 1997, p. 9). A substantial research base shows that PT has led to improvements for students with and without BP and learning disabilities in different academic areas (Spencer, Simpson, & Oatis, 2009; Okilwa & Shelby, 2010; Bowman-Perrott et al., 2013; Spilles, Hagen, & Hennemann, 2018). Bowman-Perrott et al. (2013) (26 single case studies with a total of 938 students in grades 1 through 12) summarize moderate to large academic benefits ( $TauU = .75$ ) that can be attributed to PT and state: “Among students with disabilities, those with emotional and behavioral disorders benefitted most” (p. 39).

Since the success of PT’s bolstering of academic outcomes can be explained by characteristic features such as regular and immediate peer feedback, frequent opportunities to respond for tutees, and increased time on task (Ginsburg-Block & Fantuzzo, 1997; Bowman-Perrott et al., 2013), academic achievement strongly depends on tutors’ and tutees’ academic engagement and social interaction. For this reason, social-emotional and behavioral learning are key elements of PT. The positive effects of peer-assisted learning interventions (PT and small-group cooperative learning) on behavioral outcomes have therefore been ratified as well (Ginsburg-Block, Rohrbeck, & Fantuzzo, 2006; Bowman-

Perrott, Burke, Zhang & Zaini, 2014; Spilles et al., 2018). For example, Ginsburg-Block et al. (2006) (36 studies with elementary school children) examined the small to moderate weighted mean effect sizes of peer-assisted learning on social skills (cooperative skills, conflict resolution, helping behavior, social integration;  $d = .28$ ) and behavior related to learning (on-task behavior, task orientation, rule compliance;  $d = .45$ ).

In light of these effects, PT appears to be an appropriate approach to fostering academic and nonacademic competencies. However, children with BP are affected by serious behavioral deficits, which leads to the hypothesis that additional support is necessary to improve their social behavior and academic engagement during PT interventions as well as their academic skills development.

### ***Interdependent Group Reward Contingencies***

A common strategy to support positive academic, social and behavioral outcomes in the context of peer-assisted learning is to use interdependent group reward contingencies (IGRC). An IGRC is “a contingency in which reinforcement for each member of a group is dependent on that person’s meeting a performance criterion that is in effect for all members of the group” (Cooper, Heron, & Heward, 2007, p. 9). In popular methods such as the Peer-Assisted Learning Strategies (PALS; Fuchs, Fuchs, & Karns, 2001), tutors and tutees are assigned to one of two teams in class during PT. Students receive points for cooperative behavior and correctly completing reading activities, and these points are counted for their team. At the end of the week, the class celebrates the winning group as a reward. The use of such methods during peer-assisted learning interventions can be considered a significant predictor for academic (Slavin, 1990; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003) and social (Ginsburg-Block et al., 2006) outcomes.

Thus, using IGRC during PT seems to be a promising approach, especially for the benefit of students with BP. However, the implementation of many of the systems described in the research literature seems to be quite expensive for teachers’ daily business. Time-consuming classroom-based interventions can negatively affect teachers’ implementation of them (Elliott, Witt, Galvin, & Peterson, 1984; Witt, Elliott, & Martens, 1984), which underlines the need for practical alternatives. The Good Behavior Game, which is a comparable and evidence-based behavior-management strategy using IGRC, might be such an alternative.

### ***Good Behavior Game***

The Good Behavior Game (GBG) was first introduced by Barrish, Saunders, and Wolf in 1969. Major features (with some variations) of the GBG are as follows: “(a) assigning students to teams, (b) giving points to teams that exhibit inappropriate behaviors, and (c) rewarding the team that accumulated the lowest number of points (i.e., the team that exhibits the least amount of

problem behavior)” (Bowman-Perrott, Burke, Zaini, Zhang, & Vannest, 2016, p. 180). Actual meta-analysis supports that the GBG leads to prosocial behavior and reduces problem behaviors for students in several grades, especially for those with emotional and behavioral disorders (Flower, McKenna, Bunuan, Muething, & Vega, 2014; Bowman-Perrott et al., 2016). Bowman-Perrott et al. (2016) evaluated 21 single-case studies with 1580 students in prekindergarten through Grade 12. The overall effect on behavioral variables was  $TauU = .82$  (strong effect). The GBG was especially effective in reducing disruptive and off-task behaviors. Flower et al. (2014) analyzed 22 articles about the effects of the GBG and found that moderate to large effects on challenging behaviors were also immediate. In addition to its effectiveness, the GBG – played in a simple version like in that in the present study – was also found to be a classroom intervention accepted by both teachers and students (Tingstrom, 1994).

### ***Research Gap***

Despite these significant effects on students’ behavior, interventions that explicitly combine PT and the GBG are actually rare. A systematic search in respective databases (Academic Search Complete, ERIC, PSYINDEX, MEDLINE, PsycARTICLES, PsycINFO) (combined keywords: peer tutoring, good behavior game) revealed only one study (Dion et al., 2011). In that experiment, 58 first-grade classrooms were randomly assigned to three conditions: PT, PT and GBG, or control group. In both experimental groups, a French adaptation of the first-grade reading PALS was implemented for six months at three times per week (30 min). In one of these groups, the GBG was additionally implemented during a daily reading lesson (15 min), which teachers considered to be the most important reading lesson of the day. The GBG was explicitly not played during the PT activities. As a result, both interventions were effective in helping students improving their reading skills; however, not all students benefited equally. The effect sizes were smaller for generally inattentive students than for their attentive classmates (teachers rated students’ classroom behavior before the intervention). Although during the GBG, both attentive and inattentive students were highly focused (direct observation), students in the GBG condition did not obtain higher reading results.

### ***Research Questions***

In summary, it can be assumed that studies evaluating (namely) the GBG in PT settings are a desideratum. The research questions of the present study are therefore as follows:

1) Does playing the GBG during a PT intervention (in contrast to PT only) increase academic engagement and social behavior and decrease the disruptive behavior of students with BP? The GBG leads to prosocial behavior and reduces problem behaviors for students with BP (Flower et al., 2014; Bowman-

Perrott et al., 2016). Since PT itself has an effect on behavioral variables (Ginsburg-Block et al., 2006), small to medium effect sizes are expected.

2) Does playing the GBG during a PT intervention (in contrast to PT only) enhance the academic skills development (in the present study: reading fluency) of students with BP? The use of IGRC during PT interventions can be considered a significant predictor for academic outcomes (Slavin, 1990; Rohrbach et al., 2003). If the GBG (an IGRC) leads, for example, to increased academic engagement, small to medium effects on reading fluency are also expected since PT has positive effects on academic skills.

## METHOD

### *Setting and Participants*

The study took place in two inclusive primary schools in the district of a large city in the state of North Rhine Westphalia, Germany. One school participated with two third-grade classes and one second-grade class. The other school was involved with one learning group that was specially composed for the study (second- to third-grade students; here, the intervention was performed exactly the same, but not within an already-existing class). The intervention was performed with the whole class/learning group, but only four tutors and four tutees participated in the evaluation procedures.

To identify tutors and tutees with BP, the German short version of the “Integrated Teacher Report Form” (ITRF; Volpe et al., 2018) was used. The ITRF is a universal problem-focused behavioral screener loading on the factors “academic productivity/disorganization” (APD; e.g., “does not complete class-work on time”, “does not correct own work”, “does not participate in class”) and “oppositional/disruptive behavior” (OD; e.g., “disrupts others”, “uses inappropriate language”, “has conflicts with peers”). The ITRF has a high internal consistency (APD:  $\alpha = .91$ ; OD:  $\alpha = .87$ ; total:  $\alpha = .91$ ) and provides cut-off scores for at-risk students through a comparison with the German-language version of the Teacher Report Form (TRF) of the Child Behavior Checklist (CBCL-TRF; Döpfner, Plück, & Kinnen, 2015) (attention problems: APD cut-off score  $\geq 10$ , oppositional defiant problems: OD cut-off score  $\geq 5$ , externalizing problems: total cut-off score  $\geq 13$ ). In the present study, the class teachers rated all their students on a 4-point Likert-type scale ranging from 0 (behavior is not of concern) to 3 (behavior is of strong concern) before the intervention (APD: 8 items; OD: 8 items). After that, in each class/learning group, the tutor and the tutee with the highest problem scores were chosen to be rated after each tutoring session (behavior) and to participate in the curriculum-based measurement (reading fluency).

To determine the students’ reading competency the reading fluency screening “Salzburger Lesescreening 1-4” (SLS 1-4; Mayringer & Wimmer,

2005) was used before the intervention. Based on the number of correctly evaluated statements (e.g., “A week has seven days.”) within three minutes, reading skills can be tested economically with the SLS 1-4 to identify children with low reading speed. Therefore, the screening enables a comparison with respective age norms. In the current study, the SLS 1-4 results were used in each group to assign students to fixed roles (tutors and tutees) that lasted throughout the whole intervention. Pairs were selected based on the SLS 1-4 scores (ranking students from highest to lowest, dividing the resulting list in half, pairing the highest-performing student in the higher half with the highest-performing student in the lower half, etc.) and on the teacher assessment.

For various reasons (illness, school absenteeism, extreme problematic behavior), the students could not participate in all intervention sessions. An overview of the sample is given in table 1.

**Table 1. Participants**

Name	Age	Role	Grade	APD	OD	Total	Reading Skills
Patrick	9;0	Tutor	3a	4	*15	*19	above average
Saida	7;7	Tutor	2	0	*5	5	almost average
Olaf	10;0	Tutor	3b	*13	1	*14	average
John	8;4	Tutor	LG	*16	*8	*24	average
Sarah	8;10	Tutee	3a	*12	*7	*19	weak
Gina	7;6	Tutee	2	*13	1	*14	very weak
Alina	9;9	Tutee	3b	*21	*14	*35	below average
Mehmet	9;0	Tutee	LG	5	4	9	below average

**Notes.** 3a/b: both third-grade classes, 2: second-grade class, LG: learning group. Integrated Teacher Report Form: APD: academic productivity/disorganization, OD: oppositional/disruptive behavior. \*risk-scores.

**Design**

A multiple-baseline design was used in the present single-case study to evaluate the GBG (Kazdin, 2011). The PT intervention was implemented for ten weeks (20 min per day, 3 times per week) between January and March 2018. Each group began playing the GBG in addition to PT at various points in time selected by the principle of chance. The implementation of a minimum of five

and a maximum of 25 sessions for each phase (PT versus PT + GBG) was determined. There was no data collection before PT started (baseline).

### ***Interventions***

**Lautlesetandems (reading tandems) (LLT).** For the PT intervention, the paired and repeated reading strategy LLT (Rosebrock, Nix, Rieckmann, C., & Gold, 2011) was put into practice. LLT is a successful method for enhancing the reading fluency of German-speaking children (Gold, Behrendt, Lauer-Schmalz, & Rosebrock, 2013; Lauer-Schmalz, Rosebrock, & Gold, 2014) and adolescents (Rosebrock, Rieckmann, Nix, & Gold, 2010; with BP: Paal, Hintz, & Marx, 2017). During the implementation, two students of different reading competencies read a text aloud and synchronously four times. The tutor listens carefully to the tutee's reading performance and provides correction if mistakes are made and no immediate self-correction is possible. If the tutee feels increasingly confident, he or she begins to read aloud on his or her own. One week before the intervention began, the LLT was introduced to the classes and the learning group, as proposed by Rosebrock et al. (2011). In addition, the GBG rules (see table 2) were discussed in detail during that week and placed within the range of vision in class during the LLT. Rewarding the students for appropriate behavior did not begin before the GBG was introduced. As reading material, students received 3 different reading books with short stories and factual texts. The different versions contained texts of different lengths and difficulties. Difficulties were analyzed by calculating the LIX (Björnsson, 1968) of each text. The LIX was then multiplied by the number of words. After that, all texts were ranked with regard to the resulting index (LIX\*length). The easiest version contained texts starting with LIX = 12 and 47 words. The hardest version contained texts up to LIX = 45 and 830 words. Teachers decided which version students received by themselves, oriented by the reading competency of each tutee.

**KlasseKinderSpiel (KKS).** After a few PT sessions, the KKS (Hillenbrand & Pütz, 2008) (a German adaption of the GBG) was also played during the LLT. For that task, pairs were assigned to one of two teams oriented by seating arrangement. The teachers counted fouls (inappropriate behaviors) that were matched to the KKS rules and discussed with the students beforehand in detail (see table 2). At the end of each LLT session, the teachers rewarded the teams that had accumulated five or fewer fouls. The teachers decided on their own what type of reward they delivered to the winning team (e.g., the students of the winning team could decide which games the whole class played after the intervention, or they could grab a secret reward out of a treasure chest). Pair constellations were changed every two to three weeks for pedagogical reasons. In this way, students came into contact with different class members, and KKS teams were not constant over time, which could have been problematic for the class climate. For new pairings, teachers used the described ranking list (see

below) and ensured that differences in reading skills between tutors and tutees were comparable to the prior constellations.

**Table 2. Behavior Definition**

Categories	Rules	Fouls/Single-Item Scale Definitions
Academic Engagement	We work in a concentrated and autonomous way.	stays on task looks at the reading material works autonomously
Respectful Behavior	We speak in a friendly manner, and we listen to each other.	speaks in a friendly manner to the partner or to the teacher listens to the partner or to the teacher asks and answers politely
Disruptive Behavior	We work calmly at our seats.	calls out is out of seat without permission acts aggressively

*Notes.* Fouls referring to “academic engagement” and “respectful behavior” were defined as the negative formulation of the single-item scale’s wording.

**Instruments**

**Direct Behavior Rating (DBR).** Immediately after each PT session, the teachers rated the academic engagement (AE), respectful behavior (RB) and disruptive behavior (DB) of the tutors and tutees with BP on a 6-point Likert-type scale ranging from 0 (did not occur) to 5 (always occurred). The operationalization of each single-item scale (SIS) was based on the papers written by Chafouleas et al. (2013) and Volpe and Briesch (2012) and matched to the GBG rules (see table 2). DBR appears to be an appropriate way of evaluating the behavioral development of students in classroom settings (Huber & Rietz, 2015). The categories AE and DB in particular show moderate to good validity (Huber & Rietz, 2015).

**Curriculum-Based Measurement (CBM).** In addition, directly after every PT session, undergraduates performed the Lernfortschrittsdiagnostik Lesen (LDL; Walter, 2010) with each tutor and tutee. The LDL is an instrument for testing reading fluency based on 28 parallel tests (parallel test reliability:  $r = .91$ ). The students were asked to read one text as accurately and fluently as possible in one minute. After that, the number of incorrectly read words was subtracted from the number of all read words. Comparing the LDL to the SLS 1-4 (Mayringer & Wimmer, 2005) leads to a high criterion validity ( $r = .94$ ) (Walter, 2010).



### ***Implementation Fidelity***

Correct implementation of the LLT and the KKS was ensured by conducting a teacher training before the study (two days for 6 hours each day). Also in each group, two undergraduates who were familiar with both interventions participated during the sessions. They assisted the teachers (i.e., in training tutors and tutees) and reflected with them upon difficulties. They also completed a self-created implementation-checklist based on the guidelines presented by Rosebrock et al. (2011) (LLT) and Hillenbrand & Pütz (2008) (KKS) every week. The implementation fidelity was as follows (LLT/KKS): 3a (74 % / 96 %), 3b (72 % / 100 %), 2 (52 % / 96 %), learning group (66 % / 97 %). The low fidelity values referring to the LLT are explainable by the fact that focusing on specific aspects such as reflecting upon appropriate feedback procedures every session (one question of the checklist) became less important over time. Correct paired and repeated reading procedures were implemented by nearly every group, with fidelity of 90-100 %. Only the second-grade class had a lower fidelity, of only 60 %.

### ***Data Analysis***

To evaluate the effects of the KKS regarding the student's behavior and academic skill development, a comparison of both phases (A: LLT versus B: LLT + KKS) was conducted by charging different nonoverlap effect size indices and significance tests depending on the data characteristics. All evaluation procedures were adapted to the fact that reading fluency, AE and RB were expected to increase and DB to decrease.

For all three DBR categories (AE, RB, DB) and the CBM, Pearson's Phi was translated from the percentage of all non-overlapping data (PAND; Parker, Hagan-Burke, & Vannest, 2007), and the nonoverlap of all pairs (NAP; Parker & Vannest, 2009) was calculated. Since reading fluency is likely to increase steadily over time, here, the percentage of data exceeding a median trend (PEM-T; Wolery, Busick, Reichow, & Barton, 2010) was additionally calculated by comparing the data in phase B with the regression-based (not median-based) trend of phase A.

Significance testing was performed by computing diverse randomization tests (Ferron & Ware, 1995). The DBR data (ordinal scale level) were analyzed by comparing the medians of phases A and B, and the CBM data (interval scale level) by comparing the means of both phases. Due to the small range of all three SIS and the few possible permutations of the median differences (small p-values are difficult to find), in addition to the nonoverlap effect sizes, the results of the related significance tests are also presented.

Because of the quantity of the DBR data, the results cannot be shown for all children individually. The effects are only presented for tutors and tutees overall. Further, no plots are mapped (if desired, all data can be requested from

the first author for own visualization). The CBM effect size indices are reported for each student, and individual plots are also presented. Individual randomization test results were not calculated because of the few permutations per student.

## RESULTS

### *Behavior (DBR)*

Overarching DBR outcomes are presented in table 3. The median development leads to the hypothesis that when the KKS was played in addition to PT, the tutors and tutees showed increased AE and RB as well as decreased DB. However, a significant difference is not supported by the results of the median-based randomization test. The smallest value ( $p = .119$ ) can be observed for the tutees' AE. Also notable is the fact that the tutors' and tutees' behavior during the LLT did not seem to be very inappropriate, even when the KKS was not played. Nevertheless, the large range in both phases complicates the interpretability of the findings. The calculated nonoverlap effect size indices reveal small to medium and significant effects on the tutors' and tutees' AE and DB that might be attributed to the KKS. No effects can be observed for the students' RB.

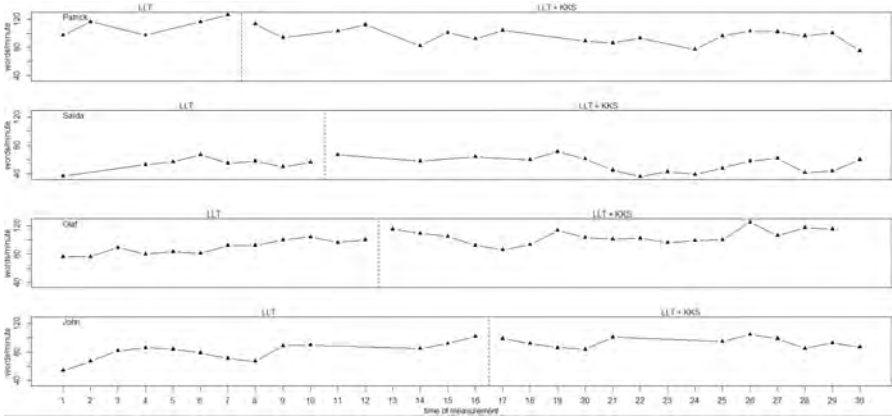
### *Reading Fluency (CBM)*

Reading fluency development is plotted in figure 1 (tutors) and figure 2 (tutees). Visual inspection indicates that notable trends occur for all students during phase A (PT only) as expected. A developmental change in phase B (PT + KKS) is not detectable. In addition, the trend test results (PEM-T) do not show any increased improvements in phase B. The Phi and NAP results give the impression that the tutees benefited moderately from playing the KKS in addition to PT. Overarching effects on the tutors' reading fluency are quite small. Overall, a clear statement cannot be obtained because the results are quite heterogeneous on an individual level. For this reason, the mean differences between both phases should be analyzed. On average, in phase B, the tutors and tutees were able to correctly read 5.13 and 6.92 more words per minute in contrast to phase A. The randomization test result indicates that the tutees' improvement in this context cannot be attributed to chance ( $p = .049$ ). Based on the collected data, it appears that playing the KKS in addition to PT led to a significant improvement in reading fluency for those students.

**Table 3. Behavior Outcomes**

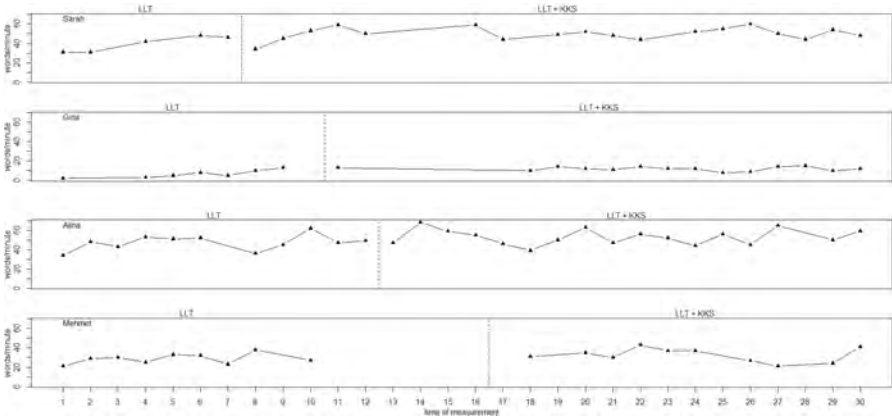
Group	Behavior	Md (A) (Min/Max)	Md (B) (Min/Max)	p (RT)	PAND			NAP		
					Phi	$\chi^2$	p	NAP	z	p
Tutors	AE	2.5 (1/5)	3.0 (1/5)	.746	*.24	5.31	.021	**70.05	5.26	.000
	RB	3.0 (2/5)	4.0 (1/5)	.734	.03	.09	.764	*62.05	4.69	.000
	DB	2.0 (0/4)	0.0 (0/4)	.693	**49	21.84	.000	**71.38	5.44	.000
Tutees	AE	3.0 (1/5)	4.0 (1/5)	.119	*.29	6.64	.010	*64.29	4.59	.000
	RB	3.5 (1/5)	4.0 (1/5)	.663	-.28	4.21	.040	48.06	3.43	.001
	DB	1.0 (0/4)	0.0 (0/4)	.408	**38	11.42	.000	*61.41	4.39	.000

**Notes.** AE: academic engagement, RB: respectful behavior, DB: disruptive behavior. Phase A: Lautlesetandems (paired and repeated reading strategy). Phase B: Lautlesetandems & KlasseKinderSpiel (German adaption of the Good Behavior Game). Mdn: Median. RT: median-based randomization test. \*small positive effect, \*\*moderate positive effect.



**Figure 1. LDL Plots (Tutors)**

**Notes.** LLT: Lautlesetandems (paired and repeated reading strategy). KKS: Klasse-KinderSpiel (German adaption of the Good Behavior Game).



**Figure 2. LDL Plots (Tutees)**

**Notes.** LLT: Lautlesetandems (paired and repeated reading strategy). KKS: Klasse-KinderSpiel (German adaption of the Good Behavior Game).

**Table 4. LDL Outcomes**

Name	n(A)	n(B)	M(A) (SD)	M(B) (SD)	p (RT)	Phi	NAP	PEM-T
Patrick	5	18	110.40 (12.90)	95.44 (10.72)	-	-.28	17.78	0.00
Saida	8	16	54.13 (8.49)	53.63 (10.97)	-	-.31	*54.30	18.75
Olaf	12	17	89.08 (9.78)	104.53 (10.20)	-	***.50	**86.52	11.77
John	13	11	80.62 (12.85)	93.27 (7.17)	-	** .33	**81.12	9.09
Tutors	38	62	81.63 (20.18)	86.76 (22.43)	.543	*.17	*65.31	9.90
Sarah	5	18	40.00 (8.57)	50.00 (6.50)	-	** .31	**86.67	0.00
Gina	7	14	6.57 (3.95)	11.86 (2.07)	-	***.57	**86.73	0.00
Alina	11	17	47.27 (7.85)	53.00 (8.08)	-	*.10	**67.91	23.53
Mehmet	9	10	28.67 (5.32)	32.60 (7.24)	-	*.26	**66.11	50.00
Tutees	32	59	31.94 (16.74)	38.86 (17.87)	.049	** .32	**75.54	18.38

**Notes.** n: Number of measurements. Phase A: Lautlesetandems (paired and repeated reading strategy). Phase B: Lautlesetandems & KlasseKinderSpiel (German adaption of the Good Behavior Game). M: Mean. RT: mean-based randomization test. \*small positive effect, \*\*moderate positive effect, \*\*\*strong positive effect.

## DISCUSSION

### *Main Findings*

The aim of the present study was to examine the effects of the GBG on the behavior of second- and third-grade students with BP during a PT intervention. Related to this goal, the impact of playing the GBG during PT on students' academic skills development was also investigated. Therefore, the reading fluency intervention LLT was implemented for ten weeks and supplemented at various points in time (class-specific) with a German adaption of the GBG (KKS). After every PT session, the teachers rated the AE, RB and DB of four tutors and tutees with the most problematic externalizing behavior. A CBM was also conducted by undergraduates immediately after PT.

The results of the DBR show that the tutors' and tutees' AE increased when the KKS was played in addition to the LLT. It was also found that the children were less disruptive. These effects can be assessed as small to medium and significant when taking the nonoverlap effect size indices into consideration. However, the results of the randomization test do not support a significant median difference between both phases. Regarding the students' RB, no mentionable change could be observed. A considerable increase in reading fluency that can be attributed to the additional implementation of the KKS may be presumed only for the tutees. With a probability of 95.1 %, the overarching mean difference between both phases cannot be explained only by chance, even if the effect sizes are quite heterogeneous on an individual level.

An explanation for the moderate impact of the KKS might be derived from the fact that PT itself has a positive influence on the development of students' academic skills (e.g., Bowman-Perrott et al., 2013), as well as on their behavior (Ginsburg-Block et al., 2006; Bowman-Perrott et al. 2014). Therefore, only small to medium effects were anticipated. Regarding social and behavioral outcomes, Bowman-Perrott et al. (2014) (examining the effects of 20 PT interventions) found that "the effect size for studies of disruptive/off-task behaviors using rewards [...] was higher than that in those not using rewards [...]. The opposite was true for studies addressing social skills/social interactions" (p. 275). However, in that meta-analysis, no explicit focus was set on IGRC. In contrast, Ginsburg-Block et al. (2006) reported that using IGRC (explicit focus) was found to moderate social outcomes but not behavior outcomes. The GBG is very effective in reducing disruptive and off-task behaviors (Bowman-Perrott et al., 2016). In summary, more research is necessary to examine the effects of playing the GBG during PT. The present study leads to the hypothesis that AE and DB are affected but not RB. However, the missing effect in the students' RB might also be proportionately traceable to the reduced validity of that DBR category (Huber & Rietz, 2015). In confirmation of this supposition, teachers in

the present study reported that RB was the most challenging to rate in contrast to AE and DB. Regarding academic skills development (reading fluency), the positive effect of the GBG that might be cautiously assumed for the tutees of the present study is consistent with the findings of Slavin (1990) and Rohrbeck et al. (2003) that report higher effect sizes for peer-assisted interventions using IGRC.

To explain the variance in the results, individual student characteristics might be informative. As the results indicate, the tutees profited more strongly from playing the KKS in addition to the LLT with regard to their reading fluency gains than did the tutors. A comprehensible explanation that associates role features with the impact of the KKS was not apparent. It is possible that the KKS supported the tutees in dealing with the tutors' feedback. Feedback is one of the most important key features for the effectiveness of PT (Bowman-Perrott et al., 2013). However, receiving immediate, corrective and high-frequency feedback may also cause defiant or aggressive reactions, especially in children with BP. Playing the KKS could have led to increased self-regulation so that the tutees with BP benefitted more from the reading intervention. The fixed role assignment and the tutees' lower reading competence could also have led to a greater increase in their reading fluency overall, which might have had an influence on the calculated nonoverlap effect size indices. Furthermore, various behavioral problems provide deeper insight. Within the group of tutors, mentionable changes in reading fluency that might be associated with the KKS only occurred for Olaf and John. These students also showed the highest risk scores in APD. Within the group of tutees, the APD risk scores are higher overall than for the tutors, supporting the hypothesis that the KKS is especially an essential supplementing key feature to PT for those children who have problems in their learning behavior.

### ***Limitations and Future Directions for Research***

There were several limitations to this study. The small sample, missing values, trends during the A phase (PT only) and the lack of any interrater reliability and social validity measures limit the explanatory power and the generalizability of the results. To further validate these effects, larger, more differentiated studies should be implemented. Future studies might be further concerned with the potential moderators of the effectiveness of playing the GBG/KKS during PT (e.g., student characteristics, different types of rewards), transfer effects (student behavior outside the intervention) and teacher experiences (e.g., self-efficacy, the perception of stress, acceptance).

### ***Conclusion***

In summary, playing the GBG/KKS during PT interventions appears to be a practical approach for teachers to promote both positive behavior and academic outcomes for students with BP, even if the results of the present study do not clearly attribute positive effects to the KKS. Teachers as well as under-

graduates reported that playing the KKS led to considerably reduced stress levels in class. They also mentioned that this effect was especially strong when using unknown rewards such as grabbing a secret reward out of a treasure chest (this procedure appeared to be more motivating for the students than determining rewards beforehand). It can be concluded that combining both is a practical way to simultaneously implement an evidence-based behavior management strategy and a reading intervention. Therefore, future studies should validate the effects suggested by the present findings.

## REFERENCES

- Avramidis, E., & Norwich, B. (2002). Teachers' attitudes towards integration/inclusion: A review of the literature. *European Journal of Special Needs Education, 17*, 129–147.
- Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good Behavior Game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis, 2*, 119–124.
- Bowman-Perrott, L., Burke, M. D., Zaini, S., Zhang, N., & Vannest, K. (2016). Promoting positive behavior using the good behavior game. A meta-analysis of single-case research. *Journal of Positive Behavior Interventions, 18*, 180–190.
- Bowman-Perrott, L., Burke, M. D., Zhang, N., & Zaini, S. (2014). Direct and collateral effects of peer tutoring on social and behavioral outcomes: A meta-analysis of single-case research. *School Psychology Review, 43*, 260–285.
- Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review, 42*, 39–55.
- Björnsson, C. H. (1968). *Läsbarhet*. Stockholm: Liber.
- Chafouleas, S. M., Jaffery, R., Riley-Tillman, T. C., Christ, T. J., & Sen, R. (2013). The impact of target, wording, and duration on rating accuracy for direct behavior rating. *Assessment for Effective Intervention, 39*, 39–53.
- Cooper, J. O., Heron, T. E., & Heward, W. L., (2007). *Applied Behavior Analysis* (2nd ed.). Upper Saddle River, NJ: Pearson.
- Dion, E., Roux, C., Landry, D., Fuchs, D., Wehby, J., & Dupéré, V. (2011). Improving attention and preventing reading difficulties among low-income first-Graders: A randomized study. *Prevention Science, 12*, 70–79.
- Döpfner, M. (2013). Klassifikation und Epidemiologie psychischer Störungen. In F. Petermann (Ed.), *Lehrbuch der Klinischen Kinderpsychologie* (7th ed., pp. 31-56). Göttingen: Hogrefe.
- Döpfner, M., Plück, J., & Kinnen, C. (2015). *CBCL/6-18R, TRF/6-18R, YSR/11-18R – Deutsche Schulalter-Formen der Child Behavior Checklist von Thomas M. Achenbach*. Göttingen, Germany: Hogrefe.
- Elliott, S. N., Witt, J. C., Galvin, G. A., & Peterson, R. (1984). Acceptability of positive and reductive behavioral interventions: Factors that influence teachers' decisions. *Journal of School Psychology, 22*, 353–360.
- Ferron, J. & Ware, W. (1995). Analyzing single-case data: The power of randomization tests. *The Journal of Experimental Education, 63*, 167–178.
- Fischbach, A., Schuchardt, K., Mähler, C., & Hasselhorn, M. (2010). Zeigen Kinder mit schulischen Minderleistungen sozio-emotionale Auffälligkeiten? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42*, 201–210.



- Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., & Vega, R. (2014). Effects of the good behavior game on challenging behaviors in school settings. *Review of Educational Research, 84*, 546–571.
- Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergartners' mathematical development: Effects of peer-assisted learning strategies. *Elementary School Journal, 101*, 495–510.
- Ginsburg-Block, M. D. & Fantuzzo, J. W. (1997). Reciprocal peer tutoring: An analysis of “teacher” and “student” interactions as a function of training and experience. *School Psychology Quarterly, 12*, 134–149.
- Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology, 98*, 732–749.
- Gold, A., Behrendt, S., Lauer-Schmaltz, M., & Rosebrock, C. (2013). Förderung der Leseflüssigkeit in dritten Grundschulklassen. In C. Rosebrock & A. Bertschi-Kaufmann (Eds.), *Literalität erfassen: bildungspolitisch, kulturell, individuell* (pp. 203-218). Weinheim: Beltz Juventa.
- Huber, C., & Rietz, C. (2015). Direct Behavior Rating (DBR) als Methode zur Verhaltensverlaufdiagnostik in der Schule: Ein systematisches Review von Methodenstudien. *Empirische Sonderpädagogik, 7*, 75–98.
- Huber, C., & Wilbert, J. (2012). Soziale Ausgrenzung von Schülern mit sonderpädagogischem Förderbedarf und niedrigen Schulleistungen im gemeinsamen Unterricht. *Empirische Sonderpädagogik, 4*, 147–165.
- Hillenbrand, C., & Pütz, K. (2008). *KlasseKinderSpiel. Spielerisch Verhaltensregeln lernen*. Hamburg: Edition Körber Stiftung.
- Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.
- Klauer, K. J., & Lauth, G. W. (1997). Lernbehinderungen und Leistungsschwierigkeiten bei Schülern. In F. E. Weinert (Ed.), *Psychologie des Unterrichts und der Schule* (pp. 701-738). Göttingen: Hogrefe.
- Klasen, F., Meyrose, A.-K., Otto, C., Reiss, F., & Ravens-Sieberer (2017). Psychische Auffälligkeiten von Kindern und Jugendlichen in Deutschland: Ergebnisse der BELLA-Studie. *Monatsschrift Kinderheilkunde, 165*, 402–407.
- Klicpera, C., Schabmann, A., Gasteiger-Klicpera, B., & Schmidt, B. (2017). *Legasthenie – LRS*. München: Reinhardt.
- Krull, J., Wilbert, J., & Hennemann, T. (2014). Soziale Ausgrenzung von Erstklässlerinnen und Erstklässlern mit sonderpädagogischem Förderbedarf im Gemeinsamen Unterricht. *Empirische Sonderpädagogik, 6*, 59–75.
- Lauer-Schmaltz, M., Rosebrock, C., & Gold, A. (2014). Lautlesetandems in der Grundschule. *Didaktik Deutsch, 37*, 44–61.
- Mayringer, H., & Wimmer, H. (2005). *SLS 1-4 – Salzburger Lesescreening für die Klassenstufen 1-4*. Bern: Hans Huber.
- Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities, 41*, 417–436.
- Newcomb, A. F., Bukowski, W. M., & Pattee, L. (1993). Children's peer relations: A meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin, 113*, 99–128.
- Okilwa, N. S. A., & Shelby, L. (2010). The effects of peer tutoring on academic performance of students with disabilities in Grades 6 through 12: A synthesis of the literature. *Remedial and Special Education, 31*, 450–463.

- Paal, M., Hintz, A.-M., & Marx, L. T. (2017). „Wir trainieren für ein Hörbuch!“. *Zeitschrift für Heilpädagogik*, 68, 193–207.
- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *Journal of Special Education*, 40, 194–204.
- Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, 40, 357–367.
- Reid, R., Gonzalez, J. E., Nordness, P. D., Trout, A., & Epstein, M. H. (2004). A meta-analysis of the academic status of students with emotional/behavioral disturbance. *The Journal of Special Education*, 38, 130–144.
- Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, 95, 240–257.
- Rosebrock, C., Nix, D., Rieckmann, C., & Gold, A. (2011). *Lese flüssigkeit fördern. Lautleseverfahren für die Primar- und Sekundarstufe*. Seelze: Kallmeyer/Klett.
- Rosebrock, C., Rieckmann, C., Nix, D., & Gold, A. (2010). Förderung der Leseflüssigkeit bei leseschwachen Zwölfjährigen. *Didaktik Deutsch*, 28, 33–58.
- Slavin, R. E. (1990). *Cooperative learning: Theory, research, and practice*. Englewood Cliffs, NJ: Prentice Hall.
- Spencer, V. G., Simpson, C. G., & Oatis, T. L. (2009). An update on the use of peer tutoring and students with emotional and behavioural disorders. *Exceptionality Education International*, 19, 2–13.
- Spilles, M., Hagen, T., & Hennemann, T. (2018): Tutorielle Leseverfahren mit Grundschulkindern mit externalisierenden Verhaltensproblemen. *Empirische Sonderpädagogik*, 10, 39–71.
- Tingstrom, D. H. (1994). The good behavior game: An investigation of teachers' acceptance. *Psychology in the Schools*, 31, 57–65.
- Utley, C. A., & Mortweet, S. L. (1997). Peer-mediated instruction and interventions. *Focus on Exceptional Children*, 29, 1–23.
- Visser, L., Büttner, G., & Hasselhorn, M. (2018, online first). Komorbidität spezifischer Lernstörungen und psychischer Auffälligkeiten: ein Literaturüberblick. *Lernen und Lernstörungen*. <https://doi.org/10.1024/2235-0977/a000246>
- Volpe, R. J., & Briesch, A. M. (2012). Generalizability and dependability of single-item and multiple-item direct behavior rating scales for engagement and disruptive behavior. *School Psychology Review*, 41, 246–261.
- Volpe, R. J., Casale, G., Mohiyeddini, C., Grosche, M., Hennemann, T., Briesch, A. M., & Daniels, B. (2018). A universal behavioral screener linked to personalized classroom interventions: Psychometric characteristics in a large sample of German schoolchildren. *Journal of School Psychology*, 66, 25–40.
- Walter, J. (2010). *Lernfortschrittsdiagnostik Lesen (LDL)*. Göttingen: Hogrefe.
- Witt, J. C., Elliott, S. N., & Martens, B. K. (1984). Acceptability of behavioral interventions used in classrooms: The influence of amount of teacher time, severity of behavior problem, and type of intervention. *Behavioral Disorders*, 9, 95–104.
- Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education*, 44, 18–28.

**AUTHORS' NOTE**

Correspondence concerning this article should be addressed to: Markus Spilles, Department of Special Education and Rehabilitation, University of Cologne, Klosterstr. 79b, 50931 Cologne, Northrhine-Westfalia, Germany, Phone: 0049-221 4701866, EMail: markus.spilles@uni-koeln.de.