

***Case Study 1: An Evidence-Based Practice Review Report***

***Theme: School Based Interventions for Learning***

***How effective is the Good Behavior Game at reducing disruptive behaviour in primary aged children?***

**Summary**

This literature review aims to evaluate the effectiveness of the Good Behavior Game intervention in reducing the disruptive behaviour of primary-aged children. The Good Behavior Game is an interdependent group contingency intervention, delivered class-wide. It can be used to increase desirable behaviour and reduce undesirable or inappropriate behaviour. Five studies met inclusion criteria and were reviewed according to Gough's (2007) Weight of Evidence Framework, Kratochwill's (2003) Coding Protocol and the 'Quality Indicators Within Single-Subject Research' by Horner et al. (2005). The studies indicated that the Good Behavior Game is an effective intervention for reducing disruptive behaviour of primary aged children.

## **Introduction**

### *What is the Good Behavior Game?*

The Good Behavior Game (GBG) is a classroom-based interdependent group contingency, which has been used to target a variety of behaviours (Lannie & McCurdy, 2007). The GBG consists of four key components: expectations of behaviour, being a team member, behaviour monitoring and reinforcement (Poduska & Kurki, 2014). The GBG involves teams in the same classroom competing against each other and works to promote appropriate behaviour. The class teacher splits the class into small teams and explains the rules of the game. Detailed criteria for target behaviour are shared with the class, for example 'no hitting' or 'no calling out'. A numerical criterion for the number of instances of inappropriate behaviour is kept hidden from the students until the end of the game (Lannie & McCurdy, 2007). The students work together to manage their own behaviour and that of their peers (McGoey, Schneider, Rezzetano, Prodan, & Tankersley, 2010). A recording sheet listing each team is used. Any instances of inappropriate behaviour observed by the teacher is marked against the team that the student who displayed the behaviour belongs too. At the end of the game, the winning team(s) are the ones whose marks fall at or below the criteria.

The team-based aspect of the GBG promotes positive peer relations and acceptance (Witvliet, van Lier, Cuijpers, & Koot, 2009). Teams are rewarded as a whole for the adherence to the rules by each student in the team. Over time, rewards change from tangible to intangible (Poduska & Kurki, 2014). The frequency and duration of each game increases steadily from 10 minutes, three times a week, over the course of the year. The game can also be applied to different lessons and different activities (Poduska et al., 2008).

The GBG has been found to be effective in reducing disruptive behaviour (McGoey et al., 2010; Ruiz-Olivares, Pino, & Herruzo, 2010) and increasing on-task behaviour (Lannie & McCurdy, 2007). Despite this, there is limited evidence of the effectiveness of this intervention in UK schools. Interventions that are recommended to schools must be evidence based. Therefore, this systematic review aims to identify whether the GBG is an effective intervention for reducing disruptive behaviour.

The use of the GBG as a behaviour management strategy has many advantages. Firstly, as it is a strategy rather than a curriculum, it does not require extra school learning time to implement (Poduska & Kurki, 2014). Secondly, reinforcements can be easily found by teachers in the school environment, for example stickers or extra choice-time (McGoey et al., 2010). Additionally, it equips the teacher with an easy to implement behaviour management tool that can be used in different lessons and activities (Lannie & McCurdy, 2007). Despite the positive aspects of the GBG, an issue raised by Wright and McCurdy (2011) regarding the intervention is with regards to the peer pressure and the need to conform that is placed upon the students taking part in the game.

### Psychological Basis

The GBG draws upon principles of Operant conditioning (Skinner, 1938) which states that behaviour modifications occur through reinforcements. Appropriate or desirable behaviour receives positive reinforcement, which leads to an increase in the desired behaviour. Inappropriate or undesirable behaviour, such as disruptive behaviour, receives a punishment or sanction. According to operant conditioning, this punishment should lead to a decrease in undesirable behaviour. The GBG also draws upon principles of Social Learning

Theory (Bandura, 1971). Children taking part in the GBG learn appropriate behaviours, through the modelling of said behaviours by their class peers.

### Relevance to EP Practice

According to a Department for Education report (2016), the number and rate of permanent and fixed-period exclusions has increased from 2013/14 to 2014/15. In total, 5,800 permanent exclusions and 302,980 fixed-period exclusions were issued across all stated-funded primary, secondary and special schools in England in the year 2014/15. The most common reason for both permanent and fixed-period exclusions was persistent disruptive behaviour (DfE, 2016). Persistent disruptive behaviour accounted for 27.3% of exclusions in primary schools, 26.4% in secondary schools and 20.7% in special schools.

Disruptive behaviour can often be a result of a child struggling with a task if it has not been delivered at the right level (Coombes, Chan, Allen, & Foxcroft, 2016). This can lead to the child being labelled as having a behavioural problem. A consequence of this is that the child may become socially excluded from peers through limited time in the classroom environment. This can lead to disaffection and decreased academic engagement, a decline that has been observed in vulnerable groups such as boys, children from ethnic minorities and low socioeconomic status (Skinner, Furrer, Marchand, & Kindermann, 2008). Behavioural problems in school can manifest in later life as mental health problems, alcohol and drugs misuse and unemployment, due to not completing a satisfactory level of schooling (Coombes et al., 2016).

Disruptive behaviour not only affects the individual pupil, but can also have an effect on the teacher and the class as a whole. Educational Psychologists (EPs) can play a role in supporting teachers to manage behaviour, by providing classroom behaviour management strategies. Hart (2010) identifies seven key elements that can contribute to effective classroom behaviour management strategies:

1. Rules that are simple to understand and are framed positively.
2. Appropriate behaviour is verbally reinforced.
3. Undesired behaviour is responded to subtly.
4. Positive relationships are formed between staff members and their students.
5. Teachers set high expectations for their students from the beginning.
6. Teachers have clear, consistent approaches to chronic misbehaviour.
7. A clean, organised and stimulating classroom environment is provided.

The GBG can be used by EPs as an intervention to help teachers who are having difficulties managing disruptive behaviour in their classrooms. The GBG involves some of the key elements highlighted above, such as clear rules, subtle responses to undesired behaviour, and high expectations of student behaviour outlined from the beginning.

To ensure that effective teaching and learning can occur, it is necessary to find appropriate and effective ways to reduce disruptive behaviours in classrooms. This review aims to establish whether an intervention based on behaviour modification and the social learning theory, can have a positive impact on disruptive behaviour.

Review Question

How effective is the Good Behavior Game at reducing disruptive behaviour in primary aged children?

**Critical review of the evidence base**

Literature search

A systematic literature search of the following electronic databases was conducted in January 2017 using the search terms listed in Table 1:

- ERIC (EBSCO)
- PsycArticles
- PsycInfo
- Web of Science

Table 1

*Search terms and results*

Search terms	ERIC Results	PsycArticles Results	PsycInfo Results	Web of Science Results
Good Behavior Game AND Disruptive Behavio*	20	23	43	63

This produced 20, 23, 43 and 63 articles in ERIC (EBSCO), PsycArticles, PsycInfo and Web of Science respectively. Inclusion and exclusion criteria were created to screen the articles for relevance to the review question. These criteria, with rationale, are presented in Table 2. Inclusion criteria 1 from Table 2 was applied to the initial search results, by selecting the filter ‘peer-reviewed journal’ where available. Figure 1 displays the selection process of the

studies to be included in the review. Titles and abstracts of 131 studies were screened using the inclusion criteria in Table 2. 38 duplicates and 79 studies that did not meet inclusion criteria 2, 4, 5, 6 and 7 were removed. Full text screening was conducted on 14 studies where a further 9 studies were removed. Studies that were removed following full text screening are presented in Appendix 1. The remaining 5 studies that were included in the systematic review are presented in Table 3.

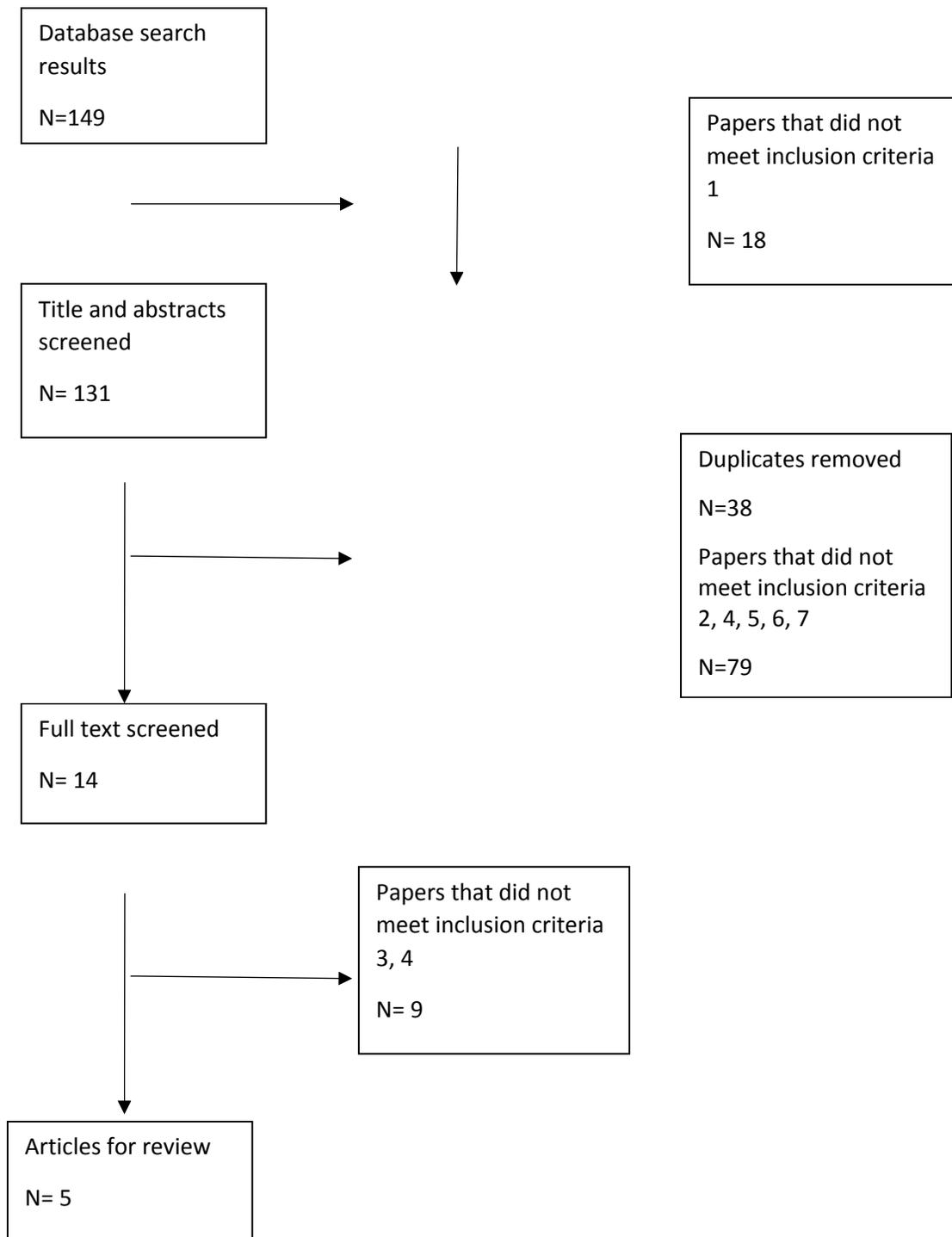


Figure 1. Database search and selection process.

Table 2

*Inclusion and exclusion criteria*

	<b>Inclusion</b>	<b>Exclusion</b>	<b>Rationale</b>
1) Type of publication	The study must appear in a peer-reviewed journal	The study has not been peer reviewed	Research that has been rigorously reviewed is, as a result, of higher quality
2) Language	The study must be published in English	The study has not been published in English	This ensures that the study can be understood by the reviewer
3) Research design	The study must include primary empirical quantitative data	The study does not include primary empirical data  The study data is qualitative	This review is looking at empirical data to allow the comparison of effectiveness of intervention
4) Intervention	The study uses the Good Behavior Game intervention  The study targets disruptive behaviour	The Good Behavior Game is not used  The study does not target disruptive behaviour	The review question is looking at the effectiveness of this particular intervention  The review question is looking specifically at the effects of the GBG on disruptive behaviour
5) Participants	Participants must be primary aged attending a mainstream school	Participants are not primary aged attending a mainstream school	This review is looking at participants in this particular age range attending a mainstream school
6) Setting	The intervention must be classroom-based	The intervention is not classroom-based	This review is looking at the effectiveness of GBG as a whole class intervention
7) Year	The study must be published after 1997	The study has been published before 1997	This ensures that the most current studies are being reviewed

Table 3

*Final list of included studies*

---

**Included studies**

---

Lannie, A.L., & McCurdy, B.L. (2007). Preventing Disruptive Behavior in the Urban Classroom: Effects of the Good Behavior Game on Student and Teacher Behavior. *Education and Treatment of Children, 30*(1), 85-98.

McGoey, K.E., Schneider, D.L., Rezzetano, K.M., Prodan, T., & Tankersley, M. (2010). Classwide Intervention to Manage Disruptive Behavior in the Kindergarten Classroom. *Journal of Applied School Psychology, 26*(3), 247-261.

Ruiz-Olivares, R., Pino, M. J., & Herruzo, J. (2010). Reduction of disruptive behaviors using an intervention based on the Good Behavior Game and the Say-Do-Report Correspondence. *Psychology in the Schools, 47*(10), 1046-1058.

Witvliet, M., van Lier, P.A.C., Cuijpers, P., & Koot, H.M. (2009). Testing Links Between Childhood Positive Peer Relations and Externalizing Outcomes Through A Randomized Controlled Intervention Study. *Journal of Consulting and Clinical Psychology, 77*(5), 905-915.

Wright, R.A., & McCurdy, B.L. (2011). Class-Wide Positive Behavior Support and Group Contingencies: Examining a Positive Variation of the Good Behavior Game. *Journal of Positive Behavior Interventions, 14*(3), 173-180.

---

**Weight of Evidence**

The Weight of Evidence (WoE) framework (Gough, 2007) was used to evaluate the five included studies. The studies were evaluated based on their methodological quality (WoE A), their methodological relevance to the research question (WoE B) and their topic relevance to the research question (WoE C). The overall WoE (WoE D) was calculated by averaging the scores obtained for WoE A, B and C. Table 4 summarises the WoE ratings for

each study. An adapted version of the Kratochwill (2003) protocol for coding group-based designs was used to evaluate the methodological quality of the study by Witvliet et al. (2009). The ‘Quality Indicators Within Single- Subject Research’ by Horner et al. (2005) was used to evaluate the methodological quality of the four other studies, which had a single-subject design. Completed protocols can be found in Appendix 2. An explanation of the WoE criteria and ratings are presented in Appendix 3.

Table 4

*Critical appraisal for quality and relevance (WoE A, B, C, D)*

<b>Studies</b>	<b>WoE A Methodological Quality</b>	<b>WoE B Methodological Relevance</b>	<b>WoE C Topic Relevance</b>	<b>WoE D Overall Weight of Evidence</b>
Lannie & McCurdy (2007)	Medium (1.71)	Low (1)	Low (1)	Low (1.24)
McGoey et al. (2010)	Medium (2.14)	Low (1)	Medium (2)	Medium (1.71)
Ruiz-Olivares et al. (2010)	Medium (1.86)	Low (1)	Low (1)	Low (1.29)
Witvliet et al. (2009)	Medium (1.67)	High (3)	High (3)	High (2.56)
Wright & McCurdy (2011)	Medium (1.86)	Low (1)	Medium (2)	Medium (1.62)

Low: Below 1.4; Medium: 1.5-2.4; High 2.5 or above

## Participants

Three of the studies were conducted in the United States (McGoey et al., 2010; Lannie & McCurdy, 2007; Wright & McCurdy, 2011), one in the Netherlands (Witvliet et al., 2009) and one in Spain (Ruiz-Olivares et al., 2010). The collated sample included 886 participants aged between 5 and 10 years old. Demographic information such as age, gender, ethnicity and socioeconomic status was reported by two studies (Witvliet et al., 2009; Ruiz-Olivares et al., 2010). Lannie and McCurdy (2007) and Wright and McCurdy (2011) reported only the gender of the participants. Information about special educational needs or difficulties such as Attention Deficit Hyperactivity Disorder (ADHD), which could affect the prevalence of disruptive behaviours in the classes, have not been reported by any of the five studies.

Classes were selected as a result of administration identifying that the class teacher was experiencing difficulties with classroom management (Lannie & McCurdy, 2007; Wright & McCurdy, 2011) or as a result of the class teacher raising concerns about the disruptive behaviour in their class (McGoey et al., 2010). Ruiz-Olivares et al. (2010) randomly selected a school, from which a class tutor requested the intervention, due to high incidences of disruptive behaviour in his class. There is no indication of how the 30 schools were chosen in the study by Witvliet et al. (2009). However, the classes within the schools were randomly assigned to either the intervention or the control group.

The description of participants and selection process for each study contributed to the WoE A rating found in Table 4.

## Design

Of the five studies included in this review, only one (Witvliet et al., 2009) used an experimental design. Witvliet et al. (2009) used random assignment to allocate participants to the intervention group or the control group, which did not receive any intervention. A single-subject design was used in the other four studies, where the single subject represented a whole class. An ABAB withdrawal design (Lannie & McCurdy, 2007; McGoey et al., 2010), ABAC withdrawal design (Wright & McCurdy, 2011) and a multiple baseline design (Ruiz-Olivares et al., 2010) was used by the other four studies. Witvliet et al. (2009) was the only study to include a control group. However, the baseline condition in the four single subject design studies acted as a within-participant control. This can be considered to be similar to a Treatment As Usual condition in a group design (Horner et al., 2005). The baseline condition should also typically include five or more data points, to demonstrate a predictable pattern (Horner et al., 2005). Lannie and McCurdy (2007) only included 4 data points in their baseline phase. The other three single-subject design studies included 5 or more data points.

Due to the random allocation in Witvliet et al.'s (2009) study, a high weighting for methodological relevance was assigned, according to the criteria for WoE B outlined in Appendix 3. The four other studies received a low weighting for methodological relevance, as there was no random assignment or control group included in the studies.

## Intervention

Intervention implementation varied between the five studies. Lannie and McCurdy (2007) and Witvliet et al. (2009) used the GBG intervention without any adaptations or comparison interventions. Wright and McCurdy (2011) compared the GBG and a positive variation called the Caught Being Good Game (CBGG). McGoey et al. (2010) used an adaptation of the GBG called the Good Behaviour Game Plus Merit (GBG-PM). Finally, Ruiz-Olivares et al. (2010) combined the GBG with the Say not-not Do-Request not (Sn-nD-Rn) correspondence training.

Witvliet et al. (2009) implemented the GBG intervention in three stages. In the first stage, the game was played three times a week to introduce it. In the next stage, the duration, applicable settings and targeted behaviours of the GBG were increased. In the final stage, generalisability to other settings was encouraged by telling the children that the rules of the GBG still applied even when the game was not being played. The length and frequency of the games has not been reported in the study. However, the intervention was implemented for 2 years. Witvliet et al. (2009) received a WoE C rating of 'high' as it met all of the criteria outlined in Appendix 3.

Lannie and McCurdy (2007) implemented the GBG for 30 minutes once a day during the maths lesson. The class teacher reviewed the game rules before the start of each GBG session. A recording sheet was displayed next to the blackboard and a numerical criterion for the session was kept hidden until the end of the game. A mark was made on the recording sheet after each instance of disruptive behaviour. If a team received less marks than the criterion, they received a reward. At various times during the maths lesson, a 10-

minute observation was conducted. Intervention integrity was improved through feedback provided to the teacher from the experimenter, which addressed the ten steps of the game integrity checklist.

Wright and McCurdy (2011) implemented the GBG or a positive variation, the Caught Being Good Game (CBGG) in two classes, once a day during a 40 minute language arts class. Each class was divided into four teams for both conditions. In the GBG, the teacher recorded instances of disruptive behaviour. Similar to the study by Lannie and McCurdy (2007), any team whose points were less than the set criterion earned a reward. In the CBGG version, the teacher recorded instances of on-task behaviour. A weekly criterion was set and teams received an award if their weekly points total exceeded this criterion. Both classes received both interventions in a different order. The kindergarten class (ages 5-6 years) received the GBG intervention first and then the CBGG intervention. The fourth grade class (ages 9-10 years) received the CBGG intervention first followed by the GBG intervention.

McGoey et al. (2010) implemented the Good Behaviour Game Plus Merit (GBG-PM) to three classes, each of which was divided into teams. Similar to the CBGG, the GBG-PM includes the opportunity for bonus points to be earned for positive behaviour through achieving class behaviour goals. In addition, rule violations resulted in the team losing a sticker. The team with the most stickers at the end of the day received a reward.

Ruiz-Olivares et al. (2010) implemented a combination of the GBG and the Say not-not Do-Request not (Sn-nD-Rn) correspondence training to a first grade class (ages 6-7 years).

Two teachers, who had been trained in the GBG and the Sn-nD-Rn correspondence, delivered the intervention. Four or five game sessions were implemented each day. Participants were told the rules of the game and then worked to achieve fewer than four marks for disruptive behaviour. Additionally, the participants received training in Sn-nD-Rn correspondence. Firstly, the participants were asked what they were going to do. For example, the teacher would ask “team 1, are you going to fight” and the team would reply, “No, we are not going to fight” (Ruiz-Olivares et al., 2010). Secondly, they were given the opportunity to do what they had said they would do, i.e. to not fight. Finally, as a team, the participants reported whether they had done what they said they would, i.e. they did not fight.

The intervention in all of the studies was delivered by either a trained psychologist or a trained teacher. Furthermore, in all studies, classes that were to receive the intervention were chosen after identification of noticeable levels of disruptive behaviour. As the GBG is targeting disruptive behaviour in these studies, it is important that this behaviour was already present in the classes.

All studies delivered the intervention to the whole class. This is vital as the GBG has been designed as a whole class intervention. Only, Witvliet et al. (2009) implemented the intervention in more than one school. Using a randomised control trial, they implemented the intervention to 47 classes in 30 elementary schools. This allows the comparison of intervention results across many different settings. As Witvliet et al. (2009) met all the criteria outlined in Appendix 3 for WoE C (topic relevance), it received a rating of ‘high’. In contrast, McGoey et al. (2010) implemented the intervention in three different classes in the

same school. Similarly, the intervention was implemented in two classes in the same school in Wright and McCurdy's (2011) study. The studies by Wright and McCurdy (2011) and McGoey et al. (2010) received a WoE C rating of 'medium', as they have met all of the criteria for this weighting as outlined in Appendix 3. Lannie and McCurdy (2007) implemented the intervention in only one class. As a result, despite meeting some of the criteria for a rating of 'high' or 'medium', this study can only receive a rating of 'low' for WoE C, as it does not meet all of the criteria for the other two weightings. Similarly, Ruiz-Olivares et al. (2010) only included one class, with two teachers and also received a rating of 'low' for WoE C.

### Measures

Four of the studies, (Lannie & McCurdy, 2007; McGoey et al., 2010; Ruiz-Olivares et al., 2010; Wright & McCurdy, 2011), used direct observations to assess the disruptive behaviour of the participants. Lannie and McCurdy (2007) used partial interval recording to record whether the participant was engaging in defined disruptive behaviour at any time during the interval. 45 student observations were conducted by the experimenter in a 10-minute observation period, with a 10 second interval for each observation.

Observations in the study by McGoey et al. (2010) were conducted by four trained graduates and undergraduates, using an adapted version of the Social Behavior Observation System of the Early Screening Procedure. A partial interval recording system was used to record the frequency of anti-social behaviour in 15 second intervals, for the duration of the 20-30 minute observation period.

Ruiz-Olivares et al.'s (2010) study involved two independent observers recording every instance of disruptive behaviour during 10-minute intervals. Observations in Wright and McCurdy's (2011) study were conducted during a 20-minute session. One student was observed for a 15 second interval, at the end of which the next student was observed. This continued for 80 observation intervals.

All four of the single subject design studies (Lannie & McCurdy, 2007; McGoey et al., 2010; Ruiz-Olivares et al., 2010; Wright & McCurdy, 2011) reported the inter-observer agreement level. Lannie and McCurdy (2007) reported a mean inter-observer agreement level of 81.4% for on-task behaviour and 82.8% for disruptive behaviour. McGoey et al. (2010) reported a mean agreement level of 96.1%. Ruiz-Olivares et al. (2010) reported a mean agreement level of 88.07%. Finally, Wright and McCurdy (2011) reported a mean agreement level of 98.09% for on-task behaviour and 98.73% for disruptive behaviour.

Witvliet et al. (2009) used five measures in their study. Cronbach's  $\alpha$  for The Problem Behavior at School Interview scale was 0.91 for the seven items assessing oppositional behaviour, 0.90 for the twelve items assessing conduct problems, and 0.80 for the four items assessing prosocial behaviour. The 11-item Social Problems scale of the Teacher's Report Form had a reported Cronbach's  $\alpha$  of 0.73.

### Findings

Percentage of all non-overlapping data (PAND) was used to calculate the effect sizes for four of the studies and the results are shown in Table 5, with the WoE D ratings.

The results provided from the included studies indicate that the GBG is an effective intervention for reducing disruptive behaviour. Effect sizes of 90% and over are considered highly effective, 70%-89% moderately effective, 50%-69% questionable effect and below 50% indicates an ineffective intervention. All studies had an effect size which indicated at least a moderate effect for baseline 1 vs intervention 1 and baseline 2 vs intervention 2. Ruiz-Olivares et al. (2010) had an effect size of 100% for both classes. Rates of disruptive behaviour decreased from baseline to intervention and remained below intervention levels during the follow up period.

This study implemented a variation of the GBG which involved the Say-Do-Report (S-D-R) Correspondence training. McGoey et al. (2010) found varying results for the three classrooms in their study. During baseline in classroom A, the levels of disruptive behaviour was displaying a downward trend. However, the intervention was still implemented. Baseline 2 levels of disruptive behaviour in classroom B remained at the same level as that of the intervention 1 phase immediately before it. The authors hypothesised that the teacher in classroom B had continued to use some of the strategies of the game in her daily teaching.

Table 5

Percentage of all non-overlapping Data (PAND) for Good Behavior Game interventions

Study	Baseline 1 vs Intervention 1	Baseline 2 vs Intervention 2	WoE D
Lannie & McCurdy (2007)	70%	85%	Low (1.24)
McGoey et al. (2010) Classroom A	94%	87%	Medium (1.71)
McGoey et al. (2010) Classroom B	80%	73%	Medium (1.71)
McGoey et al. (2010) Classroom C	74%	86%	Medium (1.71)
Ruiz-Olivares et al. (2010) Teacher 1	100%	N/A	Low (1.29)
Ruiz-Olivares et al. (2010) Teacher 2	100%	N/A	Low (1.29)
Wright & McCurdy (2011) Kindergarten class	93%	N/A	Medium (1.62)
Wright & McCurdy (2011) Fourth grade class	100%	N/A	Medium (1.62)

Lannie and McCurdy (2007) implemented the GBG without any adaptations and also demonstrated positive effects of the intervention in reducing disruptive behaviour. Wright and McCurdy (2011) found that both the GBG and the CBGG reduced disruptive behaviour. Witvliet et al. (2009) reported an effect size of the mean difference (Cohen’s *d*) of 0.45 between the control group and the GBG group after two years of intervention. This is a small effect according to Cohen’s (1988) descriptors, which can be found in Table 6.

Table 6

*Cohen’s descriptors for effect sizes*

Small	Medium	Large
0.2	0.5	0.8

**Conclusion and Limitations**

This literature review aimed to evaluate the effectiveness of the Good Behavior Game in reducing disruptive behaviour in primary aged children. Five studies were included in the review and have produced promising evidence for the effectiveness of the intervention. The study by Wietvliet et al. (2009) used a randomised control trial design and implemented the intervention in 47 classes across 30 schools. The use of a control group and implementation across many schools, as well as other predefined criteria resulted in an overall Weight of Evidence D rating of ‘high’ being allocated to this study. The other four studies, Lannie and McCurdy (2007), McGoey et al. (2010), Ruiz-Olivares et al. (2010) and Wright and McCurdy (2011), used a single-case design. These studies lacked a control group and as a result, this is a limitation of this review. Little information was given about the participants’ socioeconomic status, academic abilities and if they had special educational needs or disability. This poses issues with generalisability to the wider population. Similarly, the lack of implementation across more than one school also means that generalising the effectiveness of the intervention should be done with caution. Follow up data was only collected for one of the studies, raising concerns as to whether the effects of the intervention are long-term.

This review limited the age range to primary aged children. Additionally, none of the studies that were included were conducted in the United Kingdom. Future research is required to establish whether the GBG is effective with secondary aged children and children in the UK education system. Furthermore, studies involving larger sample sizes such as randomised controlled trials and the inclusion of follow up studies would extend the evidence base.

## **Recommendations**

The GBG is an intervention that Educational Psychologists can propose to schools who are having difficulties managing disruptive behaviour. Following initial training, the class teacher can implement the intervention daily and it can work alongside current classroom rules. Costings have not been reported in the included studies and this will be a consideration for the senior leadership team of the school to weigh up. Despite this, reduced disruptive behaviour gives rise to a more positive and effective learning environment which would be beneficial to both students and staff.

## **References**

Bandura, A. (1971). *Social Learning Theory*. New York, NY: General Learning Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Coombes, L., Chan, G., Allen, D., & Foxcroft, D.R. (2016). Mixed-methods Evaluation of the Good Behaviour Game in English Primary Schools. *Journal of Community and Applied Social Psychology*, 26, 369-387.

Department for Education. (2016). *Permanent and Fixed Period Exclusions in England: 2014-2015*. London: Department for Education.

Gough, D. (2007). Weight of Evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22(2), 213–228.

Hart, R. (2010). Classroom behaviour management: educational psychologist's views on effective practice. *Emotional and Behavioural Difficulties*, 15(4), 353-371.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education. *Exceptional Children*, 71(2), 165-179.

Kratochwill, T. R. (2003). Evidence-Based Practice: Promoting Evidence-Based Interventions in School Psychology. *School Psychology Quarterly*, 18(4), 389-408.

Lannie, A.L., & McCurdy, B.L. (2007). Preventing Disruptive Behavior in the Urban Classroom: Effects of the Good Behavior Game on Student and Teacher Behavior. *Education and Treatment of Children, 30*(1), 85-98.

McGoey, K.E., Schneider, D.L., Rezzetano, K.M., Prodan, T., & Tankersley, M. (2010). Classwide Intervention to Manage Disruptive Behavior in the Kindergarten Classroom. *Journal of Applied School Psychology, 26*(3), 247-261.

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health, 57*(7), 527–529.

Poduska, J.M., Kellam, S.G., Wang, W., Brown, C.H., Ialongo, N.S., & Toyinbo, P. (2008). Impact of the Good Behavior Game, a universal classroom-based behavior intervention, on young adult service use for problems with emotions, behavior, or drugs or alcohol. *Drug and Alcohol Dependence, 95*, 29-44.

Poduska, J.M., & Kurki, A. (2014). Guided by Theory, Informed by Practice: Training and Support for the Good Behavior Game, a Classroom-Based Behavior Management Strategy. *Journal of Emotional and Behavioral Disorders, 22*(2), 83-94.

Ruiz-Olivares, R., Pino, M. J., & Herruzo, J. (2010). Reduction of disruptive behaviors using an intervention based on the Good Behavior Game and the Say-Do-Report Correspondence. *Psychology in the Schools, 47*(10), 1046-1058.

Skinner, B.F. (1938). *The Behaviour of Organisms*. New York: Appleton-Century.

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and Disaffection in the Classroom: Part of a Larger Motivational Dynamic? *Journal of Educational Psychology, 100*(4), 765-781.

Witvliet, M., van Lier, P.A.C., Cuijpers, P., & Koot, H.M. (2009). Testing Links between Childhood Positive Peer Relations and Externalizing Outcomes Through A Randomized Controlled Intervention Study. *Journal of Consulting and Clinical Psychology, 77*(5), 905-915.

Wright, R.A., & McCurdy, B.L. (2011). Class-Wide Positive Behavior Support and Group Contingencies: Examining a Positive Variation of the Good Behavior Game. *Journal of Positive Behavior Interventions, 14*(3), 173-180.

**Appendix 1: Excluded studies at full text screening**

Excluded studies	Reason for exclusion
<p>Bradshaw, C.P., Zmuda, J.H., Kellam, S.G., &amp; Jalongo. (2009). Longitudinal Impact of Two Universal Preventive Interventions in First Grade on Educational Outcomes in High School. <i>Journal of Educational Psychology</i>, 101(4), 926-937.</p>	<p>Criteria 5. The participants in this study are not primary aged.</p>
<p>Embry, D.D. (2002). The Good Behavior Game: A Best Practice Candidate as a Universal Behavioral Vaccine. <i>Clinical Child and Family Psychology Review</i>, 5(4), 273-297.</p>	<p>Criteria 3. This study does not use primary empirical data. It is a description and review of the Good Behavior Game.</p>
<p>Flower, A., McKenna, J.W., Bunuan, R.L., Muething, C.S., &amp; Vega Jr, R. (2014). Effects of the Good Behavior Game on Challenging Behaviours in School Settings. <i>Review of Educational Research</i>, 84(4), 546-571.</p>	<p>Criteria 3. This study does not use primary empirical data.</p>
<p>Hemelt, S.W., Roth, K.B., &amp; Eaton, W.W. (2013). Elementary School Interventions: Experimental Evidence on Postsecondary Outcomes. <i>Educational Evaluation and Policy Analysis</i>, 35(4), 413-436.</p>	<p>Criteria 4. This study does not target disruptive behaviour.</p>
<p>Leflot, G., van Lier, P.A.C., Onghena, P., &amp; Colpin, H. (2010). The Role of Teacher Behavior Management in the Development of Disruptive Behaviors: An Intervention Study with the Good Behavior Game. <i>Journal of Abnormal Child Psychology</i>, 38, 869-882.</p>	<p>Criteria 4. This study is looking at the role teacher behaviour management plays in the development of disruptive behaviour.</p>
<p>Leflot, G., van Lier, P.A.C., Onghena, P., &amp; Colpin, H. (2013). The role of children's on-task behavior in the prevention of aggressive behavior development and peer rejection: A randomized controlled study of the Good Behavior Game in Belgian elementary classrooms. <i>Journal of School Psychology</i>, 51, 187-199.</p>	<p>Criteria 4. This study is looking at the effects of on-task behaviour in preventing aggressive behaviour.</p>

Newcomer, A.R., Roth, K.B., Kellam, S.G., Wang, W., Ialongo, N.S., Hart, S.R., Wagner, B.M., & Wilcox, H.C. (2016). Higher Childhood Peer Reports of Social Preference Mediates the Impact of the Good Behavior Game on Suicide Attempt. *Prevention Science, 17*, 145-156.

Criteria 4. This study does not target disruptive behaviour.

Poduska, J.M., Kellam, S.G., Wang, W., Brown, C.H., Ialongo, N.S., & Toyinbo, P. (2008). Impact of the Good Behavior Game, a universal classroom-based behavior intervention, on young adult service use for problems with emotions, behavior, or drugs or alcohol. *Drug and Alcohol Dependence, 95*, 29-44.

Criteria 4. This study does not target disruptive behaviour.

Poduska, J.M., & Kurki, A. (2014). Guided by Theory, Informed by Practice: Training and Support for the Good Behavior Game, a Classroom-Based Behavior Management Strategy. *Journal of Emotional and Behavioral Disorders, 22*(2), 83-94.

Criteria 4. This study does not target disruptive behaviour.

---

## Appendix 2: Coding Protocol

### Coding Protocol: Group-Based Design

**Domain:** School- and community-based intervention programs for social and behavioral problems  
Academic intervention programs  
Family and parent intervention programs  
**School-wide and classroom-based programs**  
Comprehensive and coordinated school health services

**Name of Coder(s):** \_\_\_\_\_

**Date:** February 12<sup>th</sup> 2017

**M / D / Y**

**Full Study Reference in APA format:** Witvliet, M., van Lier, P.A.C., Cuijpers, P., & Koot, H.M. (2009). Testing Links Between Childhood Positive Peer Relations and Externalizing Outcomes Through A Randomized Controlled Intervention Study. *Journal of Consulting and Clinical Psychology*, 77(5), 905-915.

**Intervention Name (description from study):** The Good Behaviour Game

**Study ID Number (Unique Identifier):** 4

**Type of Publication:** (Check one)

Book/Monograph

Journal article

Book chapter

Other (specify):

**I. General Characteristics**

**A. General Design Characteristics**

A1. Random assignment designs (if random assignment design, select one of the following)

- A1.1  Completely randomized design
- A1.2  Randomized block design (between-subjects variation)
- A1.3  Randomized block design (within-subjects variation)
- A1.4  Randomized hierarchical design

A2. ~~Nonrandomized designs (if nonrandom assignment design, select one of the following)~~

- A2.1  ~~Nonrandomized design~~
- A2.2  ~~Nonrandomized block design (between-participants variation)~~
- A2.3  ~~Nonrandomized block design (within-participants variation)~~
- A2.4  ~~Nonrandomized hierarchical design~~
- A2.5  ~~Optional coding of Quasi-experimental designs (see Appendix C)~~

A3. Overall confidence of judgment on how participants were assigned (select one of the following)

- A3.1  Very low (little basis)
- A3.2  Low (guess)
- A3.3  Moderate (weak inference)
- A3.4  High (strong inference)
- A3.5  Very high (explicitly stated)
- A3.6  N/A
- A3.7  Unknown/unable to code

**B. Statistical Treatment/Data Analysis (answer B1 through B6)**

- B1. Appropriate unit of analysis      **yes**     no
- B2. Familywise error rate controlled    **yes**     no     N/A
- B3. Sufficiently large N                    **yes**     no

Statistical Test: Mean difference

\_\_ level: 0.05

ES: Small

N required: 393

B4. Total size of sample (start of the study): 758

N

B5. Intervention group sample size: 501

N

B6. Control group sample size: 257

N

**For studies using qualitative research methods, code B7 and B8**

**B7. Coding**

~~B7.1 Coding scheme linked to study's theoretical-empirical basis (select one)~~     ~~yes~~     ~~no~~

~~B7.2 Procedures for ensuring consistency of coding are used (select one)~~     ~~yes~~     ~~no~~

Describe procedures: \_\_\_\_\_

B7.3 Progression from abstract concepts to empirical exemplars is clearly articulated (select one)  yes  no

B8. Interactive process followed (select one)  yes  no

Describe process: \_\_\_\_\_

**C. Type of Program** (select one)

- C1.  Universal prevention program
- C2.  Selective prevention program
- C3.  Targeted prevention program
- C4.  Intervention/Treatment
- C5.  Unknown

**D. Stage of the Program** (select one)

- D1.  Model/demonstration programs
- D2.  Early stage programs
- D3.  Established/institutionalized programs
- D4.  Unknown

**E. Concurrent or Historical Intervention Exposure** (select one)

- E1.  Current exposure
- E2.  Prior exposure
- E3.  Unknown

**II. Key Features for Coding Studies and Rating Level of Evidence/ Support**

(3=Strong Evidence 2=Promising Evidence 1=Weak Evidence 0=No Evidence)

**A. Measurement** (answer A1 through A4)

A1. Use of outcome measures that produce reliable scores for the majority of primary outcomes. The table for Primary/Secondary Outcomes Statistically Significant allows for listing separate outcomes and will facilitate decision making regarding measurement (select one of the following)

- A1.1  Yes \_\_\_
- A1.2  No \_\_\_
- A1.3  Unknown/unable to code

A2. Multi-method (select one of the following)

- A2.1  Yes \_\_\_
- A2.2  No \_\_\_
- A2.3  N/A \_\_\_
- A2.4  Unknown/unable to code

A3. Multi-source (select one of the following)

- A3.1  Yes \_\_\_
- A3.2  No \_\_\_
- A3.3  N/A \_\_\_
- A3.4  Unknown/unable to code

A4. Validity of measures reported (select one of the following)

- A5.1  Yes validated with specific target group
- A5.2  In part, validated for general population only
- A5.3  No \_\_\_
- A5.4  Unknown/unable to code

**Rating for Measurement** (select 0, 1, 2, or 3):  3  2  1  0

**B. Comparison Group**

B1. Type of Comparison Group (select one of the following)

- B1.1  Typical contact
- B1.2  Typical contact (other) specify:
- B1.3  Attention placebo
- B1.4  Intervention elements placebo
- B1.5  Alternative intervention
- B1.6  Pharmacotherapy B1.1
- B1.7  No intervention
- B1.8  Wait list/delayed intervention
- B1.9  Minimal contact
- B1.10  Unable to identify comparison group

**Rating for Comparison Group** (select 0, 1, 2, or 3):  3  2  1  0

B2. Overall confidence rating in judgment of type of comparison group (select one of the following)

- B2.1  Very low (little basis)
- B2.2  Low (guess)
- B2.3  Moderate (weak inference)
- B2.4  High (strong inference)
- B2.5  Very high (explicitly stated)
- B2.6  Unknown/Unable to code

B3. Counterbalancing of Change Agents (answer B3.1 to B3.3)

- B3.1  By change agent
- B3.2  Statistical
- B3.3  Other \_\_\_\_\_

B4. Group Equivalence Established (select one of the following)

- B4.1  Random assignment
- B4.2  Posthoc matched set
- B4.3  Statistical matching
- B4.4  Post hoc test for group equivalence

B5. Equivalent Mortality (answer B5.1 through B5.3)

- B5.1  Low Attrition (less than 20% for Post)
  - B5.2  Low Attrition (less than 30% for follow-up)
  - B5.3  Intent to intervene analysis carried out
- Findings \_\_\_\_\_

**C. Primary/Secondary Outcomes Are Statistically Significant**

C1. Evidence of appropriate statistical analysis for **primary outcomes** (answer C1.1 through C1.3)

- C1.1  Appropriate unit of analysis (rate from previous code)
- C1.2  Familywise/experimentwise error rate controlled when applicable (rate from previous code) C1.3
- Sufficiently large N (rate from previous code)

C2. Percentage of **primary outcomes** that are significant (select one of the following)

- C2.1  Significant primary outcomes for at least 75% of the total primary outcome measures for each key construct
- C2.2  Significant primary outcomes for between 50% and 74% of the total primary outcome measures for each key construct
- C2.3  Significant primary outcomes for between 25% and 49% of the total primary outcome measures for any key construct

**Rating for Primary Outcomes Statistically Significant** (select 0, 1, 2, or 3):  3  2  1  0

C3. Evidence of appropriate statistical analysis for **secondary outcomes** (answer C3.1 through C3.3)

- C3.1  Appropriate unit of analysis
- C3.2  Familywise/experimentwise error rate controlled when applicable (rate from previous code)

C3.3  Sufficiently large N (rate from previous code)

C4. Percentage of **secondary outcomes** that are significant (select one of the following)

C4.1

C4.2  Significant secondary outcomes for at least 75% of the total secondary outcome measures for each key construct

Significant secondary outcomes for between 50% and 74% of the total secondary outcome measures for each key construct

C4.3  Significant secondary outcomes for between 25% and 49% of the total secondary outcome measures for any key construct

**Rating for Secondary Outcomes Statistically Significant** (select 0, 1, 2, or 3):  3  2  1  0

C5. Overall Summary of Questions Investigated

C5.1 Main effect analyses conducted (select one)  yes  no

C5.2 Moderator effect analyses conducted (select one)  yes  no

Specify results: \_\_\_\_\_

C5.3 Mediator analyses conducted (select one)  yes  no

Specify results: \_\_\_\_\_

C. Primary/Secondary Outcomes Statistically Significant (only list  $p \leq .05$ )

(list primary outcomes first in alphabetical order, followed by secondary outcomes in alphabetical order)

Outcomes	Primary vs. Secondary	Who Changed	What Changed	Source	Treatment Information	Outcome Measure Used	Reliability	ES	(1-)
Outcome #1:	Primary Secondary Unknown	Child Teacher Parent/sign. adult Ecology Other Unknown	Behavior Attitude Knowledge Other Unknown	Self Report Parent Report Teacher Report Observation Test Other Unknown					
Outcome #2	Primary Secondary Unknown	Child Teacher Parent/sign. Adult Ecology Other Unknown	Behavior Attitude Knowledge Other Unknown	Self Report Parent Report Teacher Report Observation Test Other Unknown					
Outcome #3:	Primary Secondary Unknown	Child Teacher Parent/sign. Adult Ecology Other Unknown	Behavior Attitude Knowledge Other Unknown	Self Report Parent Report Teacher Report Observation Test Other Unknown					
Outcome #4:	Primary Secondary Unknown	Child Teacher Parent/sign. Adult Ecology Other Unknown	Behavior Attitude Knowledge Other Unknown	Self Report Parent Report Teacher Report Observation Test Other Unknown					
Outcome #5:	Primary Secondary Unknown	Child Teacher Parent/sign. Adult Ecology Other Unknown	Behavior Attitude Knowledge Other Unknown	Self Report Parent Report Teacher Report Observation Test Other Unknown					

Null Findings/Negative Outcomes Associated with the Intervention (listed alphabetically by outcome)

Outcomes	Primary vs. Secondary	Who Was Targeted for Change	What Was Targeted For Change	Source	Note null/negative outcomes	Outcome Measure Used	Reliability	ES
Outcome #1:	Primary Secondary Unknown	Child Teacher Parent/sign. Adult Ecology Other Unknown	Behavior Attitude Knowledge Other Unknown	Self Report Parent Report Teacher Report Observation Test Other Unknown				
Outcome #2	Primary Secondary Unknown	Child Teacher Parent/sign. Adult Ecology Other Unknown	Behavior Attitude Knowledge Other Unknown	Self Report Parent Report Teacher Report Observation Test Other Unknown				
Outcome #3:	Primary Secondary Unknown	Child Teacher Parent/sign. Adult Ecology Other Unknown	Behavior Attitude Knowledge Other Unknown	Self Report Parent Report Teacher Report Observation Test Other Unknown				
Outcome #4:	Primary Secondary Unknown	Child Teacher Parent/sign. Adult Ecology Other Unknown	Behavior Attitude Knowledge Other Unknown	Self Report Parent Report Teacher Report Observation Test Other Unknown				
Outcome #5:	Primary Secondary Unknown	Child Teacher Parent/sign. Adult Ecology Other Unknown	Behavior Attitude Knowledge Other Unknown	Self Report Parent Report Teacher Report Observation Test Other Unknown				

Group Design

Type of Data Effect Size is Based On	Confidence Rating in ES Computation
(check all that apply) <input type="checkbox"/> Means and SDs <input type="checkbox"/> $t$ value or $F$ value <input type="checkbox"/> Chi-square ( $df=1$ ) <input type="checkbox"/> Frequencies or proportions (dichotomous) <input type="checkbox"/> Frequencies or proportions (polytomous) <input type="checkbox"/> Other (specify): <input type="checkbox"/> Unknown	(select one of the following) <input type="checkbox"/> Highly estimated (e.g., only have $N$ p-value) <input type="checkbox"/> Moderate estimation (e.g., have complex but complete statistics) <input type="checkbox"/> Some estimation (e.g., unconventional statistics that require conversion) <input type="checkbox"/> Slight estimation (e.g., use significance testing statistics rather than descriptives) <input type="checkbox"/> No estimation (e.g., all descriptive data is present)

D. Educational/Clinical Significance

Outcome Variables:	Pretest	Posttest	Follow Up
<b>D1. Categorical Diagnosis Data</b>	Diagnostic information regarding inclusion into the study presented: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Positive change in diagnostic criteria from pre to posttest: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Positive change in diagnostic criteria from posttest to follow up: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
<b>D2. Outcome Assessed via continuous Variables</b>		Positive change in percentage of participants showing clinical improvement from pre to posttest: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Positive change in percentage of participants showing clinical improvement from posttest to follow up: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
<b>D3. Subjective Evaluation:</b> The importance of behavior change is evaluated by individuals in direct contact with the participant.	Importance of behavior change is evaluated: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Importance of behavior change from pre to posttest is evaluated positively by individuals in direct contact with the participant: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Importance of behavior change from posttest to follow up is evaluated positively by individuals in direct contact with the participant: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
<b>D4. Social Comparison:</b> Behavior of participant at pre, post, and follow up is compared to normative data (e.g., a typical peer).	Participant's behavior is compared to normative data <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Participant's behavior has improved from pre to posttest when compared to normative data: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown	Participant's behavior has improved from posttest to follow up when compared to normative data: <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown

Rating for Educational/Clinical Significance (select 0, 1, 2, or 3):  3  2  1  0

E. Identifiable Components (answer E1 through E7)

E1. Evidence for primary outcomes (rate from previous code):  3  2  1  0

E2. Design allows for analysis of identifiable components (select one)  yes  no

E3. Total number of components:

N

E4. Number of components linked to primary outcomes:

N

Additional criteria to code descriptively:

E5. Clear documentation of essential components (select one)  yes  no

E6. Procedures for adapting the intervention are described in detail (select one)  yes  no

E7. Contextual features of the intervention are documented (select one)  yes  no

**Rating for Identifiable Components** (select 0, 1, 2, or 3):  3  2  1  0

### F. Implementation Fidelity

F1. Evidence of Acceptable Adherence (answer F1.1 through F1.3)

F1.1  Ongoing supervision/consultation

F1.2  Coding intervention sessions/lessons or procedures

F1.3  Audio/video tape implementation (select F1.3.1 or F1.3.2):

F1.3.1  Entire intervention

F1.3.2  Part of intervention

F2. Manualization (select all that apply)

F2.1  Written material involving a detailed account of the exact procedures and the sequence in which they are to be used

F2.2  Formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used

F2.3  Written material involving an overview of broad principles and a description of the intervention phases

F2.4  Formal or informal training session involving an overview of broad principles and a description of the intervention phases

F3. Adaptation procedures are specified (select one)  yes  no  unknown

**Rating for Implementation Fidelity** (select 0, 1, 2, or 3):  3  2  1  0

### G. Replication (answer G1, G2, G3, and G4)

G1.  Same Intervention

G2.  Same Target Problem

G3.  Independent evaluation

**Rating for Replication** (select 0, 1, 2, or 3):  3  2  1  0

### H. Site of Implementation

H1. School (if school is the site, select one of the following options)

H1.1  Public

- H1.2  Private
- H1.3  Charter
- H1.4  University Affiliated
- H1.5  Alternative
- H1.6  Not specified/unknown

H2. Non School Site (if it is a non school site, select one of the following options)

- H2.1  Home
- H2.2  University Clinic
- H2.3  Summer Program
- H2.4  Outpatient Hospital
- H2.5  Partial inpatient/day Intervention Program
- H2.6  Inpatient Hospital
- H2.7  Private Practice
- H2.8  Mental Health Center
- H2.9  Residential Treatment Facility
- H2.10  Other (specify): \_\_\_\_\_
- H2.11  Unknown/insufficient information provided

**Rating for Site of Implementation** (select 0, 1, 2, or 3):  3  2  1  0

**I. Follow Up Assessment**

- Timing of follow up assessment: specify N/A
- Number of participants included in the follow up assessment: specify N/A
- Consistency of assessment method used: specify N/A

**Rating for Follow Up Assessment** (select 0, 1, 2, or 3):  3  2  1  0

**III. Other Descriptive or Supplemental Criteria to Consider**

**A. External Validity Indicators**

A1. Sampling procedures described in detail  yes  no

Specify rationale for selection: \_\_\_\_\_

Specify rationale for sample size: \_\_\_\_\_

A1.1 Inclusion/exclusion criteria specified  yes  no

A1.2 Inclusion/exclusion criteria similar to school practice  yes  no

A1.3 Specified criteria related to concern  yes  no

A2. Participant Characteristics Specified for Treatment and Control Group

Participants from Treatment Group	Grade/age	Gender	Ethnicity or Multi-ethnic	Ethnic Identity	Race(s)	Acculturation	Primary Language	SES	Family Structure	Locale	Disability	Functional Descriptors
<input type="checkbox"/> Child/Student <input type="checkbox"/> Parent/caregiver <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other												
<input type="checkbox"/> Child/Student <input type="checkbox"/> Parent/caregiver <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other												
<input type="checkbox"/> Child/Student <input type="checkbox"/> Parent/caregiver <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other												
<input type="checkbox"/> Child/Student <input type="checkbox"/> Parent/caregiver <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other												

Participants from Control Group	Grade/age	Gender	Ethnicity or Multi-ethnic	Ethnic Identity	Race(s)	Acculturation	Primary Language	SES	Family Structure	Locale	Disability	Functional Descriptors
<input type="checkbox"/> Child/Student <input type="checkbox"/> Parent/caregiver <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other												
<input type="checkbox"/> Child/Student <input type="checkbox"/> Parent/caregiver <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other												
<input type="checkbox"/> Child/Student <input type="checkbox"/> Parent/caregiver <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other												
<input type="checkbox"/> Child/Student <input type="checkbox"/> Parent/caregiver <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other												

A3. Details are provided regarding variables that:

A3.1 Have differential relevance for intended outcomes  yes  no

Specify: gender

A3.2 Have relevance to inclusion criteria  yes  no \_\_\_

Specify: \_\_\_\_\_

A4. Receptivity/acceptance by target participant population (treatment group)

Participants from Treatment Group	Results (What person reported to have gained from participation in program)	General Rating
<input type="checkbox"/> Child/Student <input type="checkbox"/> Parent/caregiver <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other		<input type="checkbox"/> Participants reported benefiting overall from the intervention  <input type="checkbox"/> Participants reported not benefiting overall from the intervention
<input type="checkbox"/> Child/Student <input type="checkbox"/> Parent/caregiver <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other		<input type="checkbox"/> Participants reported benefiting overall from the intervention  <input type="checkbox"/> Participants reported not benefiting overall from the intervention
<input type="checkbox"/> Child/Student <input type="checkbox"/> Parent/caregiver <input type="checkbox"/> Teacher <input type="checkbox"/> School <input type="checkbox"/> Other		<input type="checkbox"/> Participants reported benefiting overall from the intervention  <input type="checkbox"/> Participants reported not benefiting overall from the intervention

A5. Generalization of Effects:

A5.1 Generalization over time

A5.1.1 Evidence is provided regarding the sustainability of outcomes after intervention is terminated   
 yes  no \_\_\_

Specify: \_\_\_\_\_

A5.1.2 Procedures for maintaining outcomes are specified  yes  no

Specify: \_\_\_\_\_

A5.2 Generalization across settings

A5.2.1 Evidence is provided regarding the extent to which outcomes are manifested in contexts that are different from the intervention context  yes  no

Specify: \_\_\_\_\_

A5.2.2 Documentation of efforts to ensure application of intervention to other settings  yes  no

Specify: \_\_\_\_\_

A5.2.3 Impact on implementers or context is sustained  yes  no

Specify: \_\_\_\_\_

**A5.3 Generalization across persons**

Evidence is provided regarding the degree to which outcomes are manifested with participants who are different than the original group of participants for with the intervention was evaluated

yes  no \_\_\_\_\_

Specify: \_\_\_\_\_

**B. Length of Intervention** (select B1 or B2)

B1.  Unknown/insufficient information provided

B2.  Information provided (if information is provided, specify one of the following:)

B2.1 weeks \_\_\_\_\_  
N

B2.2 months \_\_\_\_\_  
N

B2.3 years 2 years

B2.4 other \_\_\_\_\_

**C. Intensity/dosage of Intervention** (select C1 or C2)

C1.  Unknown/insufficient information provided

C2.  Information provided (if information is provided, specify both of the following:)

C2.1 length of intervention session \_\_\_\_\_  
N

C2.2 frequency of intervention session \_\_\_\_\_  
N

**D. Dosage Response** (select D1 or D2)

D1.  Unknown/insufficient information provided

D2.  Information provided (if information is provided, answer D2.1)

D2.1 Describe positive outcomes associated with higher dosage: \_\_\_\_\_

**E. Program Implementer** (select all that apply)

- E1.  Research Staff
- E2.  School Specialty Staff
- E3.  Teachers
- E4.  Educational Assistants
- E5.  Parents
- E6.  College Students
- E7.  Peers
- E8.  Other
- E9.  Unknown/insufficient information provided

**F. Characteristics of the Intervener**

- F1.  Highly similar to target participants on key variables (e.g., race, gender, SES)
- F2.  Somewhat similar to target participants on key variables
- F3.  Different from target participants on key variables

**G. Intervention Style or Orientation** (select all that apply)

- G1.  Behavioral
- G2.  Cognitive-behavioral
- G3.  Experiential
- G4.  Humanistic/interpersonal
- G5.  Psychodynamic/insight oriented
- G6.  other (specify): \_\_\_\_\_
- G7.  Unknown/insufficient information provided

**H. Cost Analysis Data** (select G1 or G2)

- H1.  Unknown/insufficient information provided
- H2.  Information provided (if information is provided, answer H2.1)

H2.1 Estimated Cost of Implementation: \_\_\_\_\_

**I. Training and Support Resources** (select all that apply)

- I1.  Simple orientation given to change agents
- I2.  Training workshops conducted

# of Workshops provided 3

Average length of training an afternoon

Who conducted training (select all that apply)

- I2.1  Project Director
- I2.2  Graduate/project assistants

I2.3  Other (please specify): Licensed GBG supervisor

I2.3  Unknown \_\_\_\_\_

I3.  Ongoing technical support

I4.  Program materials obtained

I5.  Special Facilities

I6.  Other (specify): \_\_\_\_\_

### J. Feasibility

J1. Level of difficulty in training intervention agents (select one of the following)

J1.1  High \_\_\_\_\_

J1.2  Moderate \_\_\_\_\_

J1.3  Low \_\_\_\_\_

J1.4  Unknown \_\_\_\_\_

J2. Cost to train intervention agents (specify if known): \_\_\_\_\_

J3. Rating of cost to train intervention agents (select one of the following)

J3.1  High \_\_\_\_\_

J3.2  Moderate \_\_\_\_\_

J3.3  Low \_\_\_\_\_

J3.4  Unknown \_\_\_\_\_

### Summary of Evidence for Group-Based Design Studies

Indicator	Overall Evidence Rating NNR= No Numerical Rating or 0-3	Description of Evidence (Strong, Promising, Weak, No/limited evidence, or descriptive ratings)
<b>General Characteristics</b>		
General Design Characteristics	NNR	
Statistical Treatment	NNR	
Type of Program	NNR	
Stage of Program	NNR	
Concurrent/Historical Intervention Exposure	NNR	
<b>Key Features</b>		
Measurement	1	
Comparison Group	2	
Primary/Secondary Outcomes are Statistically Significant	-	
Educational/clinical significance	-	
Identifiable Components	0	
Implementation Fidelity	2	
Replication	-	
Site of Implementation	-	
Follow Up Assessment Conducted	0	

<b>Descriptive or Supplemental Criteria</b>		
External validity indicators	NNR	
Length of Intervention	NNR	
Intensity/dosage	NNR	
Dosage Response	NNR	
Program Implementer	NNR	
Characteristics of the Intervener	NNR	
Intervention Style/Orientation	NNR	
Cost Analysis Data Provided	NNR	
Training and Support Resources	NNR	
Feasibility	NNR	

**Horner et al., (2005). The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education**

**Quality Indicators Within Single-Subject Research**

**Article Reference:** Lannie, A.L., & McCurdy, B.L. (2007). Preventing Disruptive Behavior in the Urban Classroom: Effects of the Good Behavior Game on Student and Teacher Behavior. *Education and Treatment of Children*, 30(1), 85-98.

**Study ID:** 1

**Description of Participants and Setting**

Participants are described with sufficient detail to allow others to select individuals with similar characteristics; (e.g., age, gender, disability, diagnosis).

- Yes
- No
- N/A
- Unknown/Unable to Code

The process for selecting participants is described with operational precision.

- Yes
- No
- N/A
- Unknown/Unable to Code

Critical features of the physical setting are described with sufficient precision to allow replication.

- Yes
- No
- N/A
- Unknown/Unable to Code

**Overall Rating of Evidence:**  3  2  1  0

**Dependent Variable**

Dependent variables are described with operational precision.

- Yes
- No
- N/A
- Unknown/Unable to Code

Each dependent variable is measured with a procedure that generates a quantifiable index.

- Yes

- No
- N/A
- Unknown/Unable to Code

Measurement of the dependent variable is valid and described with replicable precision.

- Yes
- No
- N/A
- Unknown/Unable to Code

Dependent variables are measured repeatedly over time.

- Yes
- No
- N/A
- Unknown/Unable to Code

Data are collected on the reliability or inter-observer agreement associated with each dependent variable, and IOA levels meet minimal standards

- Yes
- No
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence:  3  2  1  0

### ***Independent Variable***

Independent variable is described with replicable precision.

- Yes
- No
- N/A
- Unknown/Unable to Code

Independent variable is systematically manipulated and under the control of the experimenter.

- Yes
- No
- N/A
- Unknown/Unable to Code

Overt measurement of the fidelity of implementation for the independent variable is highly desirable.

- Yes
- No
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence:  3  2  1  0

**Baseline**

The majority of single-subject research studies will include a baseline phase that provides repeated measurement of a dependent variable and establishes a pattern of responding that can be used to predict the pattern of future performance, if introduction or manipulation of the independent variable did not occur.

- Yes**
- No**
- N/A**
- Unknown/Unable to Code**

Baseline conditions are described with replicable precision.

- Yes**
- No**
- N/A**
- Unknown/Unable to Code**

Overall Rating of Evidence:  3  2  1  0

**Experimental Control/internal Validity**

The design provides at least three demonstrations of experimental effect at three different points in time.

- Yes**
- No**
- N/A**
- Unknown/Unable to Code**

The design controls for common threats to internal validity (e.g., permits elimination of rival hypotheses).

- Yes**
- No**
- N/A**
- Unknown/Unable to Code**

The results document a pattern that demonstrates experimental control.

- Yes**
- No**
- N/A**
- Unknown/Unable to Code**

Overall Rating of Evidence:  3  2  1  0

**External Validity**

Experimental effects are replicated across participants, settings, or materials to establish external validity.

- Yes
- No
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence:  3  2  1  0

**Social Validity**

The dependent variable is socially important.

- Yes
- No
- N/A
- Unknown/Unable to Code

The magnitude of change in the dependent variable resulting from the intervention is socially important.

- Yes
- No
- N/A
- Unknown/Unable to Code

Implementation of the independent variable is practical and cost effective

- Yes
- No
- N/A
- Unknown/Unable to Code

Social validity is enhanced by implementation of the independent variable over extended time periods, by typical intervention agents, in typical physical and social contexts.

- Yes
- No
- N/A
- Unknown/Unable to Code

Overall Rating of Evidence:  3  2  1  0

**Average WoE A across the 7 judgement areas:**

Sum of X / N = 12/7 = 1.71

X = individual quality rating for each judgement area

N = number of judgement areas

Overall Rating of Evidence:  3  2  1  0

**Appendix 3: Weight of Evidence**

Weight of Evidence A: Methodological Quality

The Horner et al. (2005) Coding Protocol was used to evaluate the methodological quality of the four single case design studies. The remaining group-based design study was evaluated using Kratochwill’s (2003) Coding Protocol. Upon completion of the Horner et al. (2005) Coding Protocol, an average score across the 7 sections was taken, to give a numerical rating for WoE A. The criteria for a rating of ‘high’, ‘medium’, or ‘low’ is presented in the table below.

	<b>WoE A Score</b>	<b>Criteria</b>
Horner et al. (2005) Coding Protocol	High	Average score of 2.5 or above across the 7 sections
	Medium	Average score of 1.5- 2.4 across the 7 sections
	<b>Low</b>	<b>Average score of 1.4 or less across the 7 sections</b>

Upon completion of the Kratochwill (2003) Coding Protocol, the study was given a score from 0 to 3 for each section where applicable. These scores were then averaged out to give an overall score for the study (between 0 and 3). The scores for measurement, comparison group and implementation fidelity were included in the overall score. This overall score was used to determine the study’s methodological weighting. The criteria for a rating of ‘high’, ‘medium’, or ‘low’ is presented in the table below.

	WoE A Score	Criteria
Kratochwill (2003) Coding Protocol	High	Overall average score of 2.5 or above
	Medium	Overall average score of 1.5- 2.4
	<b>Low</b>	<b>Overall average score of 1.4 or less</b>

Weight of Evidence B: Methodological Relevance

WoE B considers how suitable the research design is in addressing the review question. To evaluate this, a typology of evidence criteria was applied to the studies (Petticrew & Roberts, 2003).

High (3)	Medium (2)	Low (1)	Rationale
Randomised Control Trial studies	Cohort Studies Quasi-experimental studies	Case control studies Non-experimental evaluations Survey Qualitative research	Research questions addressing effectiveness are most suited to these types of evidence. (Petticrew & Roberts, 2003).

Weight of Evidence C: Relevance to review question

Weight of Evidence C considers the relevance of the focus of the evidence to the review question. As the review question is looking at the effectiveness of the Good Behaviour Game at reducing disruptive behaviour, there are two important criteria. Firstly, the intervention should be implemented in a class with high levels of

disruptive behaviour. Secondly, the intervention should be delivered by a trained psychologist or a trained class teacher. Thirdly, findings can be more widely generalised if implemented in multiple schools. Finally, the Good Behaviour Game was created as a whole class behaviour management approach. As a result, a higher rating will be applied to studies that have used the intervention in this way. The table below presents the criteria that is required for a ‘high’, ‘medium’ or ‘low’ rating. For a study to receive a rating of ‘high’, ‘medium’ or ‘low’, all criteria for that weighting must be met.

<b>High (3)</b>	<b>Medium (2)</b>	<b>Low (1)</b>
<b>Intervention is delivered by a class teacher or psychologist trained in the intervention</b>	<b>Intervention is delivered by a teacher with minimal training on the intervention</b>	<b>Intervention is delivered by a Teaching Assistant or a Learning Support Assistant</b>
<b>Intervention is implemented in a class with children that have demonstrated significant, repeated disruptive behaviour</b>	<b>Intervention is implemented in a school with children that have demonstrated low levels of disruptive behaviour</b>	<b>Intervention is implemented in a class where no disruptive behaviour has been demonstrated</b>
<b>Intervention is delivered to the whole class</b>	<b>Intervention is delivered to a small group</b>	
<b>Intervention is implemented in more than one school</b>	<b>Intervention is implemented in more than one class in the same school</b>	<b>Intervention is implemented in one class</b>

**Appendix 4: Mapping the field**

Study	Location	Participants	Intervention	Design	Measures	Primary Outcomes
Witvliet et al. (2009)	Netherlands	758 children (mean age 6.0 years SD 0.46)	1) Good behavior game (GBG)  2) Good behavior game and parent intervention (GBG+PI)	Random assignment to one of the three conditions : GBG, GBG+PI or the control group	Problem Behaviour at School Interview  UCINET network analysis software program  Social Problems scale of the Teacher's Report Form	Participants in the intervention group had significantly lower levels of externalizing behavior after 2 years ( $d=0.45$ )
McGoey et al. (2010)	United States	54 children aged 5-6 years	Good behavior game plus merit	Single-subject ABAB design	Adapted version of the Social Behavior Observation System of the Early Screening Procedure  Behavior Intervention Rating Scale-Teacher Version	Reduction in disruptive behavior in two of the classes
Lannie & McCurdy (2007)	United States	22 children aged 6-7 years	Good behavior game	ABAB withdrawal design	Observation	Reduction in disruptive behavior

---

Wright & McCurdy (2011)	United States	17 children aged 5-6 years  20 children aged 8-9 years	Good behavior game and Caught being good game	ABAC withdrawal design	Observation	Reduction in disruptive behavior
Ruiz-Olivares, Pino & Herruzo (2010)	Spain	15 children aged 6-7 years	Good behavior game Say not-not do-Request no	Multiple baseline design across situations	Observation	Reduction in disruptive behaviour

---