

Case study 1: An Evidence-based practice review report.

Theme: School/Setting Based Interventions for Social, Emotional and Mental Health.

How effective is the Good Behaviour Game at improving social and behavioural outcomes for primary aged pupils?

Summary

The Good Behaviour Game (GBG) is an interdependent group contingency intervention that is delivered at the whole-class level. The intervention has been originally used to reduce disruptive behaviour in schools but has recently been adapted to target a variety of other behaviours such as encouraging positive social behaviour (Bowman-Perrot et al., 2015). A systematic literature review was conducted to determine the effectiveness of the GBG on improving social and behavioural outcomes in primary aged pupils. Eight studies were selected based on the inclusion criteria and were evaluated in line with Gough's (2007) Weight of Evidence Framework and the coding protocol devised by Gersten et al. (2005) to enable a critical review of the evidence. The review found insufficient evidence to conclude that the GBG is an effective intervention for improving the social and behavioural outcomes for primary aged pupils. This is demonstrated by inconsistent significant results with negligible effect sizes. Limitations of the review are highlighted, along with recommendations for future research.

Introduction

The Good Behaviour Game

compromising valuable teaching time (Chan et al., 2012). Additionally, it offers
teach

reported by the Office for Standards in Education (2014) that approximately one hour of learning is lost daily due to disruptive behaviour in the classroom. In the United Kingdom (UK), the rate of exclusions has increased from 2018 to 2019, with persistent disruptive behaviour accounting for 31% of fixed-term exclusions and 35% of permanent exclusions across state-funded schools (DfE, 2021). Not only can disruptive behaviour impact upon pupil outcomes, it can also interfere with the learning of others and has led to the reduction of teachers in the profession (Ofsted, 2014).

Additionally, pupils displaying disruptive behaviour also tend to demonstrate poor relationships with others and low social competence (Hukkelberg et al., 2019). Low social competence has been associated with poorer emotional wellbeing in childhood and can lead to disruptive behaviour being displayed in school. Furthermore, children who receive social skills support are less likely to develop psychiatric disorders in adulthood (Sewell, 2019; Coombes et al., 2016).

As well as a rise in exclusion rates, there has been increasing concerns in regards to children's mental health, with more children experiencing social, emotional and behavioural difficulties in schools (DHSC & DfE, 2017). This emphasises the need for schools to deliver evidence-based interventions to

individual, group and universal level (DHSC & DfE, 2017; DfE & DoH, 2015).

The GBG provides a universal intervention that can be used as a behavioural management system to decrease disruptive behaviour (Barrish et al., 1969) and increase positive social behaviour (Sewell, 2020). Therefore, the purpose of this will review will be of relevance to EP practice as the findings will provide an overview of the evidence-

Critical Review of the Evidence Base

Literature Search

A systematic literature search was carried out on 5th January 2021 using the following three databases: PsycINFO, Web of Science and Educational Resources Information Center (ERIC). Table 1 outlines the search terms used in the literature search.

Table 1.

Search Terms for Literature Search

Intervention		Context		Participants
“Good Behaviour Game”	AND	School	AND	Pupil
OR				OR
GBG				Student
				OR
				Children

Note:

terms of the same concept (e.g. pupil OR student). Quotation marks yield results for the exact phrase of conc

Article Screening

Figure 1 represents a flow diagram of the literature search and article screening process. Where possible, search results were filtered to studies that had been peer reviewed (PsycINFO and ERIC) and published between 2015 to 2021 (PsycINFO and Web of Science). The rationale for this is outlined in the inclusion and exclusion criteria in Table 2. This yielded a total of 276 studies. After the removal of duplicates, the titles and abstracts of 262 studies were screened against the inclusion and exclusion criteria to determine eligibility for review. This led to the exclusion of 243 studies, with full-text screening carried out on the remaining 19 studies. Studies excluded at the full-text screening

stage are presented in Appendix A with the reasons for exclusion. A total of 8 studies met the eligibility criteria and were selected for in depth review (see Table 3 for included studies)

Mapping the Field

A description of the eight studies included in this systematic literature review is presented in Table 4.

Table 4.

Mapping the Field

Author	Location	Sample & Participant Characteristics	Study Type & Control Group	Measures	Outcomes
Ashworth et al. (2020)	United Kingdom	<p><i>N</i> = 3084 pupils (77 schools)</p> <p>Primary pupils aged between 6-7 years (male: 52.6%, female: 47.4%)</p> <p>Male: 52.6%, Female: 47.4</p>	<p>Cluster Randomised Control Trial</p> <p>Intervention: The GBG Control</p>		

Pupils aged 5 - 13 years
(M = 10.08 years)
attending a special
primary education.

All pupils had been
referred to the setting for
significant social,

Ialongo et al. (2019)	United States	<p><i>N</i> = 5611 (9 schools)</p> <p>Elementary pupil across Kindergarten to 5th grade</p> <p>Male: 50.81%, Female: 49.19%</p>	Clus 5 reC q70.92 302.95 (<p>The Social Health Profile Social Competence Scale</p>
-----------------------	---------------	--	----------------------------	--

Table 5.

Weight of Evidence Ratings

WoE C rating as the intervention was conducted across afterschool programs in the US, thus lackii0.000,(

et al. (2016) also used the Social Preference Procedure, whilst Split et al. (2016) used the Asocial Behaviour subscale of the Child Behaviour Scale and Peer Nomination Assessments as additional measures for social and behavioural outcomes. However, only the internal consistency was reported for the PBSI ($\alpha > .87$) and the Child Behaviour Scale ($\alpha = .90$). Finally, all eight studies did not report data on the validity of measures used, which is reflected in WoE A ratings (see Table 5).

Outcomes

Each study provided a variety of outcomes, with five studies reporting an effect size using Cohen's d (Ashworth et al., 2020; Bradshaw et al., 2020, Hart et al., 2021; Smith et al., 2018; Split et al., 2016) and three studies providing only inferential statistics, which was accounted for in WoE A ratings (see Table 5). In order to ensure consistency amongst studies, this review calculated the

Study	Sample size	Relevant Measure(s)	Behaviour Outcome	Cohen's d^1	Effect Size Descriptor	Significance value	95% CI	Overall Judgement Rating
-------	-------------	---------------------	-------------------	---------------	------------------------	--------------------	--------	--------------------------

Study	Sample size	Relevant Measure(s)	Behaviour Outcome	Cohen's d^1	Effect Size Descriptor	Significance value	95% CI	Overall Judgement Rating (WoE D)
Streimann et al. (2020)	N=708	<u>Teacher reported:</u> Strengths and Difficulties Questionnaire (SDQ)	Conduct problems	$d = -0.07$	Negligible	$p < .05^*$	-0.22 - 0.07	Low
			Peer problems	$d = 0.13$	Negligible	$p > .05$	-0.01 - 0.28	
			Hyperactivity	$d = 0.0969$				

¹According to Cohen (1988) an effect size considered negligible = < 0.2 , small = 0.2, medium = 0.5 and large = 0.8

*indicates a significant effect less than or equal to 0.05.

Four of the eight studies found that pupils in the intervention group demonstrated a significant reduction in problem behaviour at post-test, compared to the control group. However, the effect sizes were negligible ($d < 0.2$). Additionally, the type of problem behaviour varied considerably across studies. Split et al. (2016) found a reduction in pupil-reported oppositional behaviour but not in teacher-reported ratings. Additionally, non-significant effects were found for both teacher and pupil-reported measures for hyperactivity. This suggests that oppositional behaviours may be perceived differently by teachers and pupils. Streimann et al. (2020) found a significant reduction in teacher-reported conduct problems but not for hyperactivity. Two studies that used a general measure of behavioural problems (Breeman et al., 2016; Ialongo et al., 2019) found a significant reduction in behaviour at post-test compared to the control group. However, this was not the case for Hart et al. (2021) and Ashworth et al. (2020) who used similar measures.

In regards to social behaviours, one study reported a significant improvement in prosocial behaviours (Smith et al., 2018), demonstrating a negligible effect size ($d = 0.08$). Non-significant effects were found for social competence, peer relations, withdrawn from peers and prosocial skill across the remaining studies.

Three studies did not report any significant effects for both behavioural and social outcomes (Ashworth, et al., 2020; Bradshaw et al., 2020; Hart et al., 2021). Bradshaw et al. (2020) and Hart et al. (2021) received lower WoE C ratings for fidelity (Hart et al., 2021) and origins of the intervention (Bradshaw et al., 2020), thus findings may account for these limitations (see Appendix D, Table 2 for WoE C ratings).

Despite some of the studies demonstrating significant differences between the intervention and control group, a negligible effect size indicates that the differences are too small to infer that improvements in social and behavioural outcomes are a result of the GBG (Thompson, 2007).

Conclusions and

measures typically employed in such designs (Dunlap et al., 1996). Additionally, SCDs are not suitable for answering how effective an intervention is (Petticrew & Roberts, 2003). Therefore, this review provides more repre

References:

Ashworth, E., Humphrey, N., & Hennessey, A. (2020). Game Over? No Main or Subgroup Effects of the Good Behavior Game in a Randomized Trial in English Primary Schools. *Journal of Research on Educational*

Social Problems of Children with Psychiatric Disorders in Special Education Settings. *Journal of Positive Behavior Interventions*, 18(3), 156–167.

<https://doi.org/http://dx.doi.org/10.1177/1098300715593466>

Chan, G., Foxcroft, D., Smurthwaite, B., Coomes, L., & Allen, D. (2012). Improving child behaviour management: An evaluation of the Good Behaviour Game in UK primary schools. *Oxford: Oxford Brookes University*.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum

Coomes, L., Chan, G., Allen, D., & Foxcroft, D. R. (2016). Mixed-methods Evaluation of the Good Behaviour Game in English Primary Schools. *Journal of Community & Applied Social Psychology*, 26(5), 369–387. <https://doi.org/http://dx.doi.org/10.1002/casp.2268>

Department for Education, & Department of Health (DfE & DoH). (2015). *Special educational needs and disability code of practice: 0 to 25 years*. GOV.UK. <https://www.gov.uk/government/publications/send-code-of-practice-0-to-25>.

Department of Education (DfE). (2021). *Permanent and fixed-period exclusions in England*. GOV.UK. <https://explore-education-statistics.service.gov.uk/find-statistics/permanent-and-fixed-period-exclusions-in-england>.

Department of Health and Social Care, & Department of Education (DHSC & DfE). (2017). *Transforming children and young people's mental health*

provision: a green paper. GOV.UK.

<https://www.gov.uk/government/consultations/transforming-children-and-young-peoples-mental-health-provision-a-green-paper>.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1(2), 170.

Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., & Vega Jr, R. (2014). Effects of the Good Behavior Game on challenging behaviors in school settings. *Review of Educational Research*, 84(4), 546-571.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71(2), 149-164.

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.

- Leflot, G., van Lier, P. A., Onghena, P., & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention study with the good behavior game. *Journal of Abnormal Child Psychology*, 38(6), 869-882.
- Office for Standards in Education (Ofsted). (2014). *Below the radar: low-level*. London: Ofsted.
- Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology & Community Health*, 57(7), 527-529.
- Sewell, A. (2020). An Adaption of the Good Behaviour Game to Promote Social Skill Development at the Whole-Class Level. *Educational Psychology in Practice*, 36(1), 93–109.

Streimann, K., Selart, A., Trummal, A., Karin, S., Anne, S., & Aire, T. (2020).

Effectiveness of a Universal, Classroom-Based Preventive Intervention (PAX GBG) in Estonia: a Cluster-Randomized Controlled Trial.

Prevention Science, 21(2), 234–244.

<https://doi.org/http://dx.doi.org/10.1007/s11121-019-01050-0>

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence

intervals for effect sizes. *Psychology in the Schools*, 44(5), 423-432.

Tingstrom, D. H., Sterling-Turner, H. E., & Wilczynski, S. M. (2006). The good

behavior game: 1969-2002. *Behavior Modification*, 30(2), 225-253.

Wilson, D. B. (n.d). *Practical Meta-Analysis Effect Size Calculator*. Campbell

Collaboration. [https://www.campbellcollaboration.org/research-](https://www.campbellcollaboration.org/research-resources/effect-size-calculator.html)

[resources/effect-size-calculator.html](https://www.campbellcollaboration.org/research-resources/effect-size-calculator.html).

Appendices

Appendix A.

Table 1.

Studies excluded at full-text screening based on exclusion criteria

Study reference	Exclusion Criteria number(s)
1. Coombes, L., Chan, G., Allen, D., & Foxcroft, D. R. (2016). Mixed methods evaluation of the good behaviour game in English primary schools. <i>Journal of community & applied social psychology, 26</i> (5), 369-387.	4
2. Donaldson, J. M., Wiskow, K. M., & Soto, P. L. (2015). Immediate and distal effects of the good behavior game. <i>Journal of Applied Behavior Analysis, 48</i> (3), 685-689.	4
3. Groves, E. A., & Austin, J. L. (2019). Does the Good Behavior Game Evoke Negative Peer Pressure? Analyses in Primary and Secondary Classrooms. <i>Journal of Applied Behavior Analysis, 52</i> (1), 3–16.	4
4. Lynne, S., Radley, K. C., Dart, E. H., Tingstrom, D. H., Barry, C. T., & Lum, J. D. K. (2017). Use of a technology-enhanced version of the good behavior game in an elementary school setting. <i>Psychology in the Schools, 54</i> (9), 1049–1063.	4 & 5
5. McHugh, D. M. B., Radley, K. C., Tingstrom, D. H., Dart, E. H., & Barry, C. T. (2019). The Effects of Tootling via ClassDojo on Pupil Behavior in Elementary Classrooms. <i>School Psychology Review, 48</i> (1), 18–30.	4 & 5
6. Ortiz, J., Bray, M. A., Biliias-Lolis, E., & Kehle, T. J. (2017). The Good Behavior Game for Latino English Language Learners in a Small-Group Setting. <i>International Journal of School & Educational Psychology, 5</i> (1), 26–38.	4 & 5
7. Sewell, A. (2020). An adaption of the Good Behaviour Game to promote social skill development at the whole-class level. <i>Educational Psychology in Practice, 36</i> (1), 93-109.	4
8. Sondey, J., Taurel, N., Khem, C., Negre, L., Birocchi, S., & Reynaud-Maurupt, C. (2019). The Good Behavior Game: when the classroom becomes the playground for life skills (Toulon area). <i>European Journal of Public Health, 29</i> .	4
9. Stratton, K. K., Gadke, D. L., & Morton, R. C. (2019). Using the Good Behavior Game with High School Special Education Pupils: Comparing Pupil- and Teacher-Selected Reinforcers. <i>Journal of Applied School Psychology, 35</i> (2), 105–121.	4 & 5
10. Torok, M., Rasmussen, V., Wong, Q., Werner-Seidler, A., Bridianne, O., Toumbourou, J., & Alison, C. (2019). Examining the impact of the Good Behaviour Game on emotional and behavioural problems in primary school children: A case for	4

integrating well-being strategies into education. *Australian Journal of Education*, 63(3), 292–306.

11. Wu, Y. Q., Chartier, M., Ly, G., Phanlouong, A., Shelby, T., Weenusk, J., Murdock, N., Munro, G., & Sareen, J. (2019). Qualitative case study investigating PAX-good behaviour game in first nations communities: insight into school personnel's perspectives in implementing a whole school approach to promote youth mental health. *BMJ Open*, 9(9).
-

Appendix B.

Example of WoE A Coding Protocol

Coding protocol: Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71(2), 149-164.

Note. This protocol was adapted so that the questions are relevant to the research question. Question wording will be strikethrough (e.g. ~~example~~) and appropriately re-worded if it is not relevant to the review. Explanations of adaptation will be identified through italics.

Reference of the study: Ashworth, E., Humphrey, N., & Hennessey, A. (2020). Game Over? No Main or Subgroup Effects of the Good Behavior Game in a Randomized Trial in English Primary Schools. *Journal of Research on Educational Effectiveness*, 13(2), 298-321.

Essential Quality Indicators

A. Quality indicators for describing participants

Was sufficient information provided about the participants involved in the study? ~~to determine whether the participants demonstrated the difficulties presented?~~ *(the intervention is a universal programme so for the purpose of the review, this question will focus on whether the study has described the participant characteristics involved in the study)*

Yes

No

Unknown/Unable to Code

Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?

Yes

No

Unknown/Unable to Code

Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?

Yes; partially, Teachers allocated to intervention condition provided training and support to implement GBG.

No

Unknown/Unable to Code

B. Quality indicators for implementation of the intervention and description of comparison conditions

Was the intervention clearly described?

Yes

No

Unknown/Unable to Code

Was the fidelity of implementation described and assessed?

Yes

No

Unknown/Unable to Code

Was the nature of services provided in comparison conditions described?

Yes

No

Unknown/Unable to Code

C. Quality indicators for outcome measures

Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance?

Yes

No; one measure assessed social and behavioural outcomes

Unknown/Unable to Code

Were outcomes for capturing the intervention's effect measured at the appropriate times?

Yes

No

Unknown/Unable to Code

D. Quality indicators for data analysis

Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study?

Yes

No

Unknown/Unable to Code

Did the research report include not only inferential statistics but also effect size calculations?

Yes

No

Unknown/Unable to Code

Desirable Quality Indicators

Was data available on attrition rates among intervention samples?

Yes

No

Unknown/Unable to Code

Was severe overall attrition (30% or more) avoided? Is attrition comparable across samples?

Yes

No

Unknown/Unable to Code

Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures?

Yes

No; only internal consistency reliability reported (coefficient .87)

Appendix C.

Weight of Evidence B: Methodological Relevance

The WoE B determines how relevant the methodology used within each study is in answering the review question on the effectiveness of the Good Behaviour Game on social and behavioural outcomes for primary aged pupils. This was evaluated using Petticrew and Roberts (2003) typology of evidence criteria (see Table 1 and 2 for criteria and rationale). All criteria must be met to fulfil rating (see Table 3 for WoE B ratings).

Table 1.

WoE B Rating Criteria

Table 2.

Rationale for WoE B Criteria

Criteria	Rationale
Design	Randomised control trials are identified as a high-quality research design to examine an interventions effectiveness (Petticrew & Roberts, 2003) A comparison group allows for intervention effects to be compared against another intervention or no intervention

Table 3.

WoE B Rating for Reviewed Studies

Study	WoE B Rating
Ashworth et al. (2020)	3
Bradshaw et al. (2020)	3
Breeman et al. (2016)	3
Hart et al. (2020)	3
Ialongo et al. (2019)	3
Smith et al. (2018)	3
Split et al. (2016)	3
Streimann et al. (2020)	3

Appendix D.

Weight of Evidence C: Relevance to review question

- 2 Teachers receive training but no continued support from researchers

Streimann et al. (2020)	2	2	3	2	2.25 (High)
-------------------------------	---	---	---	---	----------------
