Examining the Moderating Effects of Poverty on the Implementation and Outcomes of the Good Behaviour Game

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in Education in the Faculty of Humanities

2019

Kirsty Frearson

School of Environment, Education and Development

Contents

Word count: 69,639

List of Figures

List of Tables

Abstract

Examining the Moderating Effects of Poverty on the Implementation and Outcomes of the Good Behaviour Game

The Good Behaviour Game (GBG) is a US universal preventive intervention that has promising social and emotional outcomes for children, particularly males with high baseline levels of disruptive behaviour (e.g. Kellam et al., 2008; Petras et al., 2008; Poduska et al., 2008). The GBG has a large literature base with much of the previous research conducted in high poverty locations. However, there has been little investigation into how such a contextual factor could affect both the implementation of the intervention as well as its outcomes, particularly the differential outcomes for children experiencing poverty.

The present study utilised a cluster-randomised control trial design, with pupils in Year 3 (ages 7-8) as the target cohort at the start of the project (N= 3084). Seventy seven schools were randomly allocated to either the intervention condition (N=38) or the usual practice condition (N=39). Delivery of the GBG started in September 2015 and ended in July 2019. Teachers in both conditions rated pupils' disruptive behaviour using the *Teacher Observation of Children's Adaptation Checklist* (TOCA-C: Koth, Bradshaw, & Leaf, 2009) at yearly intervals throughout the study. Implementation (in this case fidelity and dosage) was assessed through a yearly structured observation and teachers recording intervention sessions through an electronic scoreboard. Poverty at the individual level was determined through a pupil's free school meal (FSM) status. School-level poverty was determined through grouping schools into moderate or high poverty categories based on the percentage of the population receiving FSM.

Using multilevel modelling, the study found an interaction between individual-level poverty, school-poverty and the GBG, suggesting that children who were experiencing poverty and attending a high poverty school would have a small increase in disruptive behaviour. Meanwhile, three implementation profiles were identified through cluster analysis: (low dosage, medium fidelity; medium dosage, medium fidelity, high dosage, high fidelity). However, it was determined that poverty did not moderate the implementation of the GBG. Further research should consider investigating the role of organisational capacity as a factor that influences the implementation of interventions such as the GBG when testing their effectiveness in school settings.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgements

First, I would like to thank all the participants who were involved in the research, particularly the teachers, whose time and dedication in the data collection process made this study possible.

I would also like to thank my first and second supervisors, Prof. Neil Humphrey and Dr. Michael Wigelsworth for their support and guidance throughout the project. I must also thank Dr. Alex Hennessey being the MLwin guru and Dr. Ann Lendrum for her wisdom and kindness.

Thanks to "the GeeBeeGeebies" Dr. Emma Ashworth, Kim Petersen and Louise Upton for providing many hours of entertainment and support in "the wellbeing laboratory" as well as during our data collection expeditions. I would also like to thank Tamsin McCaldin and Laura Goodfellow for their friendship and encouragement over the last few years.

Finally, I would like to dedicate this work to Uncle Gavin and Grandad Peter: I miss you both.

Key to Abbreviations

General terms
DB= Disruptive behaviour
SEND= Special Educational Needs & Disabilities
FSM= Free School Meals
PP= Pupil Premium
GBG= Good Behaviour Game
NQT= Newly Qualified Teacher
RQ= Research Question
NPD= National Pupil Database

Acronyms for organisations
DWP= Department for Work & Pensions
EEF= Education Endowment Foundation
JFR= Joseph Rowntree Foundation
WB= The World Bank
UN= United Nations

Acronyms for theories
CET= Classical Economic Theory
SCT= Social Capital Theory
SFT= Social Field Theory
SLT= Social Learning Theory
N-CET= Neo-Classical Economic Theory

Methodological terms
CII= Cumulative Intervention Intensity
C-RCT= Cluster Randomised Control Trial
HCA= Hierarchical Cluster Analysis
HLM= Hierarchical Linear Modelling
ICC= Intra-Class Coefficient
ITT= Intention To Treat
MDES= Minimal Detectable Effect Size
MI= Multiple Imputation
MLM= Multi-Level Modelling
RCT= Randomised Control Trial

Chapter 1
Introduction

1.1 Introduction to Chapter

The current chapter outlines the context of this thesis. First, the author examines the rising interest in the development and use of preventive interventions, before providing a brief history of the Good Behaviour Game (GBG). The present study utilises data from a wider evaluatory trial of the intervention, and so the context and critical differences between the two pieces of research are outlined. A brief overview of the researcher's involvement in the wider evaluation is also provided in order to provide further context to the reader. Then the author presents three key constructs that form part of the current research. These are poverty, disruptive behaviour (DB), and implementation. The aim of this section is to provide the reader with some background on these constructs before reading the literature review. Next, the rationale and research questions for the thesis are presented; the intended contribution to knowledge follows this. The final section of the current chapter lays out the presentation of the thesis, which consists of three main sections: the literature review, the study itself, and the discussion.

1.2 Growing interest in prevention interventions

The proliferation of the GBG literature base is not only down to the success of the intervention; it also reflects a growing interest in prevention research generally. Historically, policymakers and practitioners approached many social, behavioural, and health issues using the treatment model (Beresford, 2002). The treatment model, although necessary, is a reactive model where intervention occurs once the symptoms of maladaptive behaviour or outcome have manifested (Eisenberg & Neighbours, 2009). Attempting to reverse symptoms costs more money, time, resources compared to preventing them from occurring in the first place (Knapp, McDaid & Parsonage, 2011). Moreover, individuals with maladaptive symptoms will often experience a period of distress or ill health before accessing the services that intervene. The time at which services intervene poses an ethical dimension to the treatment model: is it right to let individuals at risk of adverse outcomes develop maladaptive behaviours or experience distress before receiving help? While the preventive model is unlikely to eradicate cases of behavioural, social, and health disorders, the number of individuals requiring treatment later on in development will reduce. Therefore, the burden on services that prescribe treatment will be smaller, and so access to these services would be improved (Eisenberg & Neighbours, 2009).

In the UK, the current landscape of prevention research is behind its western counterparts (Humphrey et al., 2016). The lack of intervention development within the UK contradicts the founding principles Beverage set out for the National Health Service (NHS), which was meant to be heavily invested in prevention practices (Gough, 2013). However, various political and financial decisions caused the service to become focused on treatment (Gough, 2013). After neo-liberalism introduced market principles to the NHS (and much of the state

services more generally) during the 1980s, New Labour attempted to realign services with a preventive focus. The late coalition government continued this focus and expanded upon to follow a similar policy model to the US, whereby government departments set up clearinghouses across key areas of policy to promote evidence-based practice (Cabinet Office, 2019). For the education sector, the Education Endowment Foundation (EEF) was set up to coordinate research and disseminate the findings of initiatives and best practice (EEF, 2018). Such efforts included preventive interventions. The rationale for the EEF was to inform schools of good practices and effective interventions as schools had greater autonomy over their budgets since the proceeding governments had made radical changes to the educational system. Whether these changes are good or bad is an entirely separate debate, but the fact remains that the government feel they have some responsibility to guide policy in this way, and it has led to increased interest in prevention research.

1.3 A brief history of the Good Behaviour Game

Barrish, Saunders and Wolf (1969) observed a teacher in the US playing a behaviour management game that later became the GBG. The teacher awarded infractions to groups of pupils that broke any of the outlined rules of the game during a class activity. In this instance, there were many rules the children had to follow in order to "win" the game (Embry, 2013). However, the researchers saw the potential in the overall concept demonstrated by the teacher, and this led to them producing a more formalised version of the game. Barrish and colleagues (1969) reduced the number of rules the children had to follow down to four. The developers based the rules on necessary behaviours that are universally expected by teachers within their classroom and phrasing them using positive terminology (Ford et al., 2014). From this, many single-case studies reported reductions in DB as a result of implementing the GBG (see Tingstrom, Sterling-Turner & Wilczynski, 2006).

From this, the American Institute of Research (AIR) formally developed the GBG into a bespoke preventive intervention. The developers provided technical assistance to the implementers in the form of GBG coaches (Ford et al., 2014). The role of the coaches was to identify areas where the implementation of the programme could be improved and work with the implementer to overcome classroom or personal barriers to delivering the intervention. Another addition to the formalised preventive programme was the use of "voice levels". The voice levels compliment the rule "we will work quietly" and allow the teacher to set an appropriate noise level according to the type of class activity taking place when the children play the game. AIR tested the effectiveness of the GBG through two large longitudinal RCTs in Baltimore, the US where a range of social and emotional outcomes was measured (Kellam et al., 2011; Poduska et al., 2008; Kellam et al., 2014). The researchers tracked the participants from grade 3 until young adulthood, with the children at the most risk of maladaptive issues including conduct disorder and mental health problems benefiting from the universal intervention. The impact of these findings not only

led AIR to export the programme internationally, but also adaptations of the GBG were developed by other researchers, all of which have also reported promising results (e.g. van Lier, 2002). Chapter 3 discusses the components of the intervention and its evidence base.

As the market for preventive interventions is more established outside of the UK, it is easier to import programmes compared to focusing on producing interventions. However, an essential factor that shapes outcomes is how suited an intervention is to the cultural setting it will operate within (Castro et al., 2004). Developers produce preventive interventions with a particular context in mind, which influences the development process (Castro et al., 2004). By exporting an intervention to a different context, stakeholders cannot assume it will be as effective as it was in its original context (Lendrum & Humphrey, 2012). Therefore, stakeholders should take into account the cultural fit of a programme when importing interventions from other countries. In the present study, the stakeholders' trialled the GBG, a US programme, in UK primary school classrooms. Although both countries share a common language, there are many cultural differences between the UK and the US, particularly with regards to their education systems. For instance, a key difference is that teachers within the US rarely have a teaching assistant (TA) in the classroom, whereas the majority of teachers within the UK will have a TA in the class to aid children who have special educational needs or disabilities (SEND). Such nuances can affect the extent to which programmes will be accepted and effective within new cultural contexts.

Developers can support schools that have different contextual or cultural needs by adapting the intervention to meet the demands of the schools or classrooms implementing the programme (Lendrum & Humphrey, 2012; Castro et al., 2004). As the GBG has been utilised and implemented for such a long time, it is no surprise that several versions of the intervention now exist. For instance, several single-case studies exist whereby the GBG has been modified to suit specific contexts within daily school routines (e.g. Sweizy, Matson & Box, 1993; Hynes, Smith & Perkins, 2009). A prominent example is the study conducted by McCurdy, Lannie and Barnabas (2009) where a school implemented the GBG within a cafeteria. Although the majority of these modified versions of the GBG provided promising results, the lack of publications that follow up on the research suggests that there is not a high demand for such modifications. This lack of further study may be due to the variety of licensed versions of the GBG on the intervention market.

AIR and the Paxis Institute (PAX) have produced the principal variants of the GBG. While it is difficult to deduce which form of the intervention was released first, the PAX GBG cites many of the studies produced by AIR to evidence the effectiveness of their programme (E.g. Embrey, 2009). While the underlying structure of the PAX GBG is the same, the core distinction between the two interventions is that PAX GBG exclusively uses intangible rewards whereas the AIR GBG gradually transitions from tangible to intangible rewards (see

section 4.5.1). Meanwhile, van Lier (2002) developed another variant of the programme, called *Taakspel,* for schools in The Netherlands and Belgium in response to the culturally specific needs of those schools. Although the rules of *Taakspel* are the same as the AIR version of the intervention, the way teachers handle infractions are very different. For further insight into the key differences between the three versions of the GBG, see section 4.5.1.

Before stakeholders invest heavily into an imported intervention, a pilot study should be conducted (Humphrey et al., 2016). This allows the developers to see how well a programme translates into a new context, it also allows for the developers to trial any adaptations made to the intervention specifically for the new context (Humphrey et al., 2016). AIR performed a pilot study in conjunction with Oxford County Council in order to assess the fit of the GBG within UK primary school classrooms (Coombes, Chan, Allen & Foxcroft, 2016). Overall teachers reported that they would continue to implement the GBG as part of their behaviour management practice in order to manage their classrooms. Chapter 3 further details the importance of cultural context and its influence on implementation.

1.4 Context of the wider trial and the present study

The reader should be informed of the thesis' context before the present study is explored in depth. The data used in this piece of research is from a wider RCT trial that the author was an acting-research assistant for. This unorthodox situation contributes to some of the strengths and weaknesses of the study. For instance, the author would not have feasibly been able to collect data from such a large sample if conducting the research alone. However, they would have more freedom in certain methodological decisions if the author had complete control of the study design as traditionally expected. The primary purpose of this section is to allow the author to share their role on the wider evaluatory trial and how this may have influenced the present study. Secondly, the author wishes to reassure the reader that the present study is a distinct and unique piece of research.

1.4.1 Separating the present study from the wider trial

The EEF commissioned an evaluation study conducted by a research team at the University of Manchester in collaboration with Mentor UK, a charity that aims to reduce substance abuse use amongst children and young people. The project evaluated the GBG using a cluster randomised control trial (C-RCT). Randomisation occurred at the school level. The evaluation study followed participants (*N*= 3093) for two academic years in order to determine with GBG would significantly reduce academic attainment, although DB outcomes were a secondary outcome of the study. The role of MentorUK was to oversee

the implementation of the GBG by providing the teachers that delivered the intervention training and technical support throughout the trial.

The present study uses data collected from the evaluation study conducted by The University of Manchester on behalf of the EEF. Despite utilising the same core data set as the evaluation study, the present thesis is a distinct and original piece of research for three main reasons:

1. The present study used DB as a primary outcome, while academic attainment was the primary outcome of the evaluation study.
2. The present study uses DB outcome data at two time-points whereas the evaluation only uses the final time point to draw conclusions.
3. The researcher investigates both the role of individual poverty and school-wide poverty as moderators for DB. The evaluation study only accounts for individual poverty.
4. The present study attempts to determine the extent school-wide poverty moderates the implementation of the GBG. Although the evaluation study also investigates the implementation practices of teachers delivering the intervention, the research team did not account for factors that could moderate implementation levels.

Chapter 5 presents a more in-depth discussion about the distinctiveness of the current study from the evaluation project by detailing the conductance of the evaluation study and contrasting the two pieces of research (see section 5.2).

1.4.2 The role of the researcher in the wider GBG evaluation

This section provides an overview of the unique position of the researcher concerning the wider trial and present study. By describing the researcher's contribution to the wider trial, the reader would gain a better understanding of the present study's context. As already explained, the EEF commissioned an evaluation of the GBG which assessed the intervention outcome as well as implementation outcomes. As part of the funding received by the evaluation team at The University of Manchester, a PhD scholarship opportunity was provided for two individuals. They would act as research assistants for the wider trial alongside their research which would include the data collected from the wider trial. While the opportunity to use data from the wider trial ensured the present study was robust, it also meant that there were limitations on the data that was available to the researcher. This meant that many methodological decisions were restricted by the needs of the wider trial. Such limitations are further explored in section 5.6.

Below is a description of the roles the researcher undertook to support the wider trial.

Recruitment

The researcher was heavily involved in the recruitment of participants in the wider trial. Alongside others, the researcher contacted schools within targeted geographic locations to promote the GBG evaluation to schools. Information was provided to interested parties to aid decision-making about participating in the study. More details on the recruitment process are given in section 5.5.1.

Instrument development

As the delivery of one session of the GBG had several stages, a structured observation schedule was designed to collect implementation data from teachers that implemented the GBG (see appendix A). The researcher was heavily involved in the designing and piloting of the observation schedule to ensure that the measurement of the sub-categories of implementation was robust and that the schedule was easy to use in the field. As data from the observation schedule was used within the present study, the researcher's involvement in the observation design ensured the schedule would be appropriate for their research. More information about the development of the structured observation schedule can be found in section 5.6.3.

Interview schedule development

The wider evaluation trial included a qualitative component to further investigate the implementation of the GBG in a UK classroom context. This was to ascertain whether the US intervention would be appropriate for the UK intervention market. The qualitative strand involved six case study schools that volunteered for further collaboration with the research team. A significant component of the case study was the use of semi-structured interviews with key stakeholders of the GBG within each of the case study schools. This meant that interview schedules had to be designed so they were appropriate for the range of participants involved in the case study. Although the researcher was not as heavily involved in the designing of the interview questions as other areas of the wider trial, they did contribute to editing the questions to ensure comprehension and usability within the field.

As the present study is quantitative in nature, there is no description of the wider trial's qualitative element because it does not directly relate to the researcher's objectives. For those interested in the qualitative strand of the wider trial, it is advised for the reader to see Humphrey and colleagues (2018) for further details.

Data collection

The researcher's primary contribution to the wider evaluation was collecting the data from the participating schools in both conditions of the RCT. As there were many milestones for the different outcomes being measured by the wider trial, the researcher was in constant

correspondence with the participating schools. Below is a brief description of how the researcher aided the data collection process by method.

Online questionnaires were sent to teachers in both conditions at baseline and the end of each academic year. It was the researcher's responsibility to provide technical assistance to ensure teachers were able to complete the questionnaires. Teachers provided behavioural data for each of the children in their class, data from one subscale of the *Teacher Observation of Classroom Adaptation- Checklist* (Koth et al., 2009) was used in the present study. Meanwhile, teachers in the GBG condition were also required to rate their implementation of the intervention using an online form.

The researcher visited GBG schools once a year to observe teachers implementing the intervention using the rubric developed especially for the trial. If a school had multiple forms within a cohort, the researcher would observe each teacher individually. As the teacher ran the intervention session, the researcher would follow the observation sheet marking whether the teacher was following the developer's intended delivery and making a note of any adaptations or reactions from the pupils.

Towards the end of the wider trial, all participating schools were visited in order to complete the Hodder group reading test and a battery of social and emotional learning inventories (see Humphrey, et al., 2018 for more details). This was because the wider trial was interested in academic outcomes alongside behavioural outcomes. The researcher administered the reading test by reading the instructions aloud to the class.

Finally, as part of the qualitative strand of the wider trial, the researcher conducted semi-structured interviews with key stakeholders of the GBG within 6 case study schools. The teachers implementing the intervention, the head-teacher or deputy head teacher, a focus group of children and some parents were interviewed twice a year about their opinions and experiences of the GBG. These interviews were often conducted on the school premises as part of a regular visit to the schools where the researcher would informally observe the teacher and class both during inside and outside the GBG intervention time.

Data analysis

The researcher was trained to analyse the interviews using thematic analysis as part of the implementation strand of the wider trial. Both a priori and posteriori themes were utilised to analyse the stakeholders' experiences of the GBG. As one of the lead qualitative analysts, the profiles and themes that were written up featured heavily in the qualitative section of the wider trial's write up (see Humphrey, et al., 2018 for the qualitative findings).

Summary of the researcher's experiences on the wider trial

As the above sections suggest, the researcher spent a considerable amount of time in the field, particularly within schools implementing GBG. Although the present study is quantitative in nature, the regular interactions with teachers implementing the GBG allowed the researcher to gain a rich understanding of the implementers' perceptions of the intervention. Although quantitative studies tend not to detail the wider responsibilities and experiences of the researcher, it is worth mentioning their unique position of working on the wider trial alongside their own research project before delving into the present study. There are two reasons for this, firstly, the researcher goes to length about justifying how their work is distinct and separate from the wider trial and providing the surrounding context may further support this argument. Secondly, despite being as objective as possible, the researcher is able to admit that their understanding of the findings of the study may be influenced by their involvement in the wider trial.

1.5 Core concepts in the present study

The thesis covers four broad areas in great detail. The first area explored is poverty, followed by DB, the production and testing of preventive interventions, and the implementation of said interventions. These areas are outlined below to aid the reader in future chapters.

1.5.1 Poverty

Poverty has always been a research interest for many stakeholders. There is a moral imperative to understand its effects to improve better outcomes for those who experience it and break the cycle of generational poverty. For others, the interest may be for financial reasons. For instance, higher levels of employment across a society make for a stronger economic market and reduce costs incurred by the state to provide financial aid to those who are not in work (Chang, 2014). The fact remains, poverty causes adverse outcomes at both an individual level and societal level (McLoyd, 1998; Picket & Wilkinson, 2009). For instance, individuals who experience poverty are at risk of having: a shorter life expectancy (Wilkinson, 1992), developing chronic physical illnesses (Anderson & Horvath, 2004), experiencing mental health problems (Murali & Oyebode, 2004), or engaging in criminal activity (Patterson, 1991). At a societal level, high rates of poverty can lead to an overstretching of resources, with many individuals requiring financial aid, housing, or access to GPs and other health services (Duncan & Brooks-Gunn, 1997). With this inability to meet public needs, impoverished areas are more likely to have poor housing conditions, low-quality local services, and a lack of opportunity (Putnam, 2000).

Both individual and societal levels of poverty put neighbourhoods and families at risk of creating a generational cycle whereby children raised in impoverished conditions are likely to remain in poverty at adulthood (Boyden & Cooper, 2007). However, what is poverty? To the layperson, one of two images usually come to mind: starving individuals in developing

countries, or the homeless person begging for change on the local high street. These are clear, disturbing and very moving examples of poverty. However, poverty is not limited to those lacking fundamental rights such as food, water and shelter (World Bank, 2017). Many individuals who are employed and access to hot running water and electricity are considered to fall below the poverty line (World Bank, 2017). Poverty lies on a continuum, and the moving examples mentioned above rest on the extreme end. However, the line between "poor" and "not-poor" is hard to place; where the powers that be draw this distinction affects which families or individuals receive help and which ones are left struggling despite not being recognised as experiencing poverty (Davis & Sanchez-Martinez, 2014).

With academics across different disciplines defending their definitions of poverty, it is making the distinction between those who require help and those who do not is increasingly difficult (Davis & Martinez, 2014). However, although there is unlikely to be a universally accepted definition for the term, all that research or work alongside individuals experiencing poverty agree that more resources and attention is required to better understand the concept at both the individual level and societal level. Chapter 2 explores the different perspectives that have developed to define and measure poverty (section 2.2). Ultimately, the purpose of investigating poverty within the literature review is twofold. First, the researcher wishes to establish a definition and theory that drives the present study. Secondly, the researcher wants to demonstrate how detrimental the effects of poverty can be for individuals, particularly those who experience it during childhood.

1.5.2 Disruptive behaviour

Extremely harmful behaviours that put the self or others at risk has always attracted attention from the academic sphere. The psychoanalytical concept of Thanatos was the first popularised explanation as to why individuals engage in destructive acts that the media clamour for and the state declare war against (Freud, 1920). Psychodynamic advocates refer to Thanatos as the "death principle", in which individuals engage in destructive acts to fulfil one's destiny to cease existing (Freud, 1920). While Freud's concepts now remain in the confines of specific therapeutic interventions, psychologists still seek explanations and ways of reversing or preventing individuals from committing unintelligible acts. Developmental psychologists have been seeking precursor signs in children to predict the likelihood of them developing antisocial behaviours that put themselves and others at risk of harm later in life (Tremblay, 2010).

It is accepted that most individuals do not suddenly decide to vandalise property or physically assault others. Longitudinal studies that have followed cohorts of children over time have measured aggressive and impulsive behaviours throughout their development (e.g. Loeber & Stoutham-Loeber, 1998). Such studies report the same pattern, repeatedly, with most infants displaying low intensity but high frequencies of disruptive behaviour. As they age, the rate of these behaviours reduces, until adolescence, where a considerable

proportion of individuals increase both the frequency and intensity of their disruptive behaviours (Tremblay, 2010). The same reduction is detected when adolescents transition into young adults. However, a small proportion of children increase the intensity and frequency of their DB throughout childhood. It is these individuals who are most likely to be diagnosed with conduct disorder or antisocial personality disorder and commit crimes later in life (Vaillancourt et al., 2007). Chapter 3 further explores the normative and non-normative developmental trends associated with DB (see section 3.2).

1.5.3 The implementation of preventive interventions

Although the RCT can be considered the "gold standard" of testing preventive interventions, a major threat to the method's integrity is the lack of insight into the processes it provides regarding "how" any changes happen (Humphrey et al., 2016). Not being able to see into the "black box" can lead to incorrect conclusions about null findings (Lendrum & Hymphrey, 2012). Researchers within prevention science refer to this inaccurate attribution of cause as a 'type three' error, as many studies concluded that interventions "failed" when in fact they were not being implemented correctly (Lendrum & Humphrey, 2012). Therefore, it is imperative to collect data on implementation as part of an RCT to assess whether intervention failure or implementation failure was the cause of any null findings.

The term implementation itself has been subject to controversy, with two perspectives influencing its conceptualisation and measurement. These are the macro- and micro-perspectives. The macro-perspective uses the developers' manual as a benchmark for successful implementation (Humphrey et al., 2016). Meanwhile, advocates for the micro-perspective argue implementation is comprised of subordinate constructs, and that these should be the focus of study. The subordinate constructs that makeup implementation are fidelity of delivery, quality of delivery, dosage, participant responsiveness, adaptations, programme reach, and programme (Durlak & DuPre, 2008). Within the micro-perspective of implementation, the developer's vision for intervention delivery is still an important aspect to consider when assessing successful implementation. However, the micro-perspective acknowledges that the implementer can enhance intervention effectiveness by deviating from the developer's recommendations (Humphrey et al., 2016).

Chapter 4 looks in further detail into the contributions of implementation science has made to researchers understanding of preventive intervention outcomes at conceptual and methodological levels. The chapter also highlights a bias towards certain aspects of implementation and concludes that researchers should consider investigating a wider variety of subordinate constructs when assessing the implementation of a preventive intervention.

1.6 The rationale for the present study

While empirical evidence suggests that the GBG is particularly effective for highly disruptive males (see Flower et al., 2014 for a review of the literature), there is little evidence for the impact of the programme on other risk groups. One such group is children experiencing poverty. The little evidence that is available has been mostly descriptive and not considered factors that may affect implementation. With primary schools serving their local neighbourhoods within the UK, the intake of these schools is likely to reflect the demography of the wider community. A relevant example is that the level of poverty within a neighbourhood is reflected in the percentage of children in receipt of FSM within a given school population. Besides a substantial evidence base highlighting the adverse outcomes associated with childhood poverty (see McLoyd, 1998, for a review), there is evidence that indicates that the level of poverty at broader levels, such as at the school or neighbourhood, add further risk to children's outcomes (see: Duncan & Brooks-Gunn, 1997).

Moreover, the infrastructure and services that serve populations in deprived areas are under pressure to meet the higher levels of demand from a subsection of the population requiring more access to services and resources compared to the average (Botvin, 2004). Therefore, schools serving deprived catchment areas are likely to struggle to meet the demands of their intake because of the higher levels of need, in the form of looked after children (LAC), SEND, mental health and behavioural issues (Botvin, 2004). Consequently, the intention to implement a new strategy, such as a preventive intervention, to meet pupils' needs may not become an established practice because of the time and resources required to implement it are already invested elsewhere in the school. Schools that have strong organisational structures are better able to manage their resources, as according to social disorganisation theory (SDT: Shaw, Mackay & Hayner, 1942; Bradshaw, Slawyer & O)'Brennan, 2009). Based on the above rationale, the current study investigates whether school level and individual level poverty interact to moderate the disruptive behaviour of children.

While this thesis has already proposed that schools within high-poverty catchments are likely to struggle with the delivery of a new intervention, a further reason to investigate implementation within the present study is due to a lack of knowledge regarding the delivery practices of the GBG. Despite a substantial evidence base, only a handful of studies report the extent to which teachers' implemented the programme (e.g. Dion et al., 2011). Of these studies, the majority used descriptive statistics to present the data, and none took into account of broader factors that could moderate implementation levels, such as school-level poverty. Therefore, the present study is an ideal opportunity to expand the GBG literature base by gaining insight into teachers' implementation levels of the programme and how school-level poverty may contribute to any variation in delivery.

1.6.1(RQ1) Does poverty moderate the effects of the Good Behaviour Game on children's disruptive behaviour?

a) Does the GBG reduce children's DB?
b) Are there differential effects of the GBG among children who are eligible for FSM?
c) Does school-level poverty moderate the effects of the GBG on DB?
d) Do school-level poverty and child-level FSM eligibility interact to moderate the effects of the GBG on DB?

The above questions are presented sequentially as it is important to establish the basic facts such as whether the intervention works for everyone before looking for subgroup effects. This allows for the identification of an appropriate explanation for the findings. For instance, if the preventive intervention worked for all students but that children experiencing poverty have an even more pronounced effect the explanation would be very different to finding intervention effects at the subgroup level rather than overall. Without these basic answers, we are at risk of presenting a less accurate explanation.

As the majority of RCT evidence for the GBG has not reported evidence for a main intervention effect, it is unlikely that for the present study will report such an effect. However, the researcher predicts that children that experience poverty that attend high poverty schools that implement the GBG.

1.6.2 (RQ2) Does poverty moderate the implementation of the Good Behaviour Game?

a) Are there distinct implementation profiles amongst teachers implementing GBG?
b) Does implementation vary as a function of school-level poverty?
c) Does implementation variability moderate the effects of the GBG on DB?
d) Do school-level poverty and implementation variability interact to moderate the effects of the GBG on disruptive behaviour?
e) Do school-level poverty, implementation variability and individual-level poverty interact to moderate the effects of the GBG on DB?

Much like the first RQ, there is a need to check the basic facts before going onto the more complicated analysis. RQ2 follows RQ1 because it is important to know in what context it will be asked. If the findings relating to RQ1 produce positive and significant results, then investigating implementation will be to see what factors are most beneficial. However, if RQ1 reveals null effects for the intervention, this part of the analysis would be seeking to determine if this was due to implementation failure, or intervention failure. It is important to note that for sub-questions c to d, any findings are exploratory as the sample was underpowered for a three level MLM that focuses on half the original sample (because the usual practice sample cannot be utilised in RQ2 analyses).

1.6.3 Contribution to knowledge

The first contribution to knowledge this thesis will provide is by developing an understanding as to whether poverty is a moderator of the GBG on DB outcomes. The justification for this piece of research is because there has been little research into the role poverty may have on outcomes, despite the fact that the majority of schools interested in implementing the intervention serve high poverty areas. While one study has attempted to achieve this aim, the researchers were unable to account for the clustered nature of the data set (Kellam et al., 1998). The present study addresses this by utilising a more appropriate analysis, in this case, multilevel modelling (MLM).

Another contribution to knowledge the thesis provides is utilising cluster analysis to test whether implementation profiles for preventive programmes can be generated. Using a cluster analysis requires the researcher to acknowledge that implementation as a construct is multi-faceted. Not only does that mean the micro-perspective of implementation has to be adopted, but the present study also recognises that individuals vary the extent to which they deliver the different subordinate constructs of implementation. Therefore, the present study also aims to persuade other researchers in the field of implementation to consider a more holistic approach when assessing intervention delivery.

Further conditional contributions to knowledge may be made if the researcher can report distinct and meaningful implementation profiles. One such contribution would be determining whether implementation profiles vary as a function of school-level poverty. The investigation of such an interaction could potentially highlight the vital role broader contexts play in influencing preventive intervention delivery, an area that is currently lacking a strong evidence base. Another contribution to knowledge the present thesis provides is an attempt to establish the extent to which school-level poverty, teacher implementation profiles, and individual-level poverty interact to influence the GBG's effect on DB. Finally, the present study seeks to move beyond the current "black box" paradigm within prevention science by focusing on the contextual factors that may influence intervention outcomes.

1.7 Presentation of the thesis

The present study consists of three main sections. These are: the literature review, the present study, and the discussion of the results. Each section is outlined below.

1.7.1 The literature review

The literature review consists of three chapters. The first chapter further explores the notion of poverty and seeks to clarify an appropriate definition of the term for the present study. The section also seeks to provide a theory that explains how poverty occurs. The choice of definition is vital to the present study as it should complement the proxy variable used to reflect poverty within the sample. Then, the poverty section outlines the many negative

effects it has on children's development in order to emphasise the need to investigate solutions such as preventive interventions and their viability for this vulnerable group.

The second chapter produces a definition for DB in order to clear up a long history of confusion surrounding the term. Following the same format as the previous chapter, the literature review then explores both the negative outcomes associated with DB as well as the associated risk factors. A key risk factor that is discussed is poverty, with the focus of this particular section highlighting the need for research similar to the present study. Then, the chapter introduces preventive interventions. This starts more generally, with an outline of the different types of preventive interventions, followed by the research processes required to produce an intervention that is ready to disseminate to the wider population. The section provides an important foundation for the reader to better understand the intervention tested in the present study, as well as the subsequent chapter. Finally, the chapter presents the GBG intervention and evaluates the empirical evidence in line with the objectives of the present study.

The final chapter of the literature review introduces the concepts of implementation science, particularly the dimensions that make up the term (e.g. Fidelity, quality, dosage, participant responsiveness, etc.). The associated literature for implementation variation is then scrutinised, alongside the factors that may be responsible for such variations. The literature review within this chapter takes a particular focus on school-level poverty as a possible explanation for some of the variability consistently described in the literature (e.g. SDT: Shaw, Mackay & Hayner, 1942). The proposal of school-level poverty as a potential moderator for implementation variability is then used to justify the exploratory nature of the second aim of the present study.

1.7.2 The present study

This part of the thesis consists of two chapters. The first is the methodology chapter, which presents the rationale and justification for the methods and analytical strategy used to answer the proposed research questions (outlined in section 1.5.1 and 1.5.2). The sample, procedure, and ethical considerations are also presented within this chapter. The second chapter is the results chapter. The data screening is presented before the inferential statistics that answer the two research questions.

1.7.3 The discussion

The discussion is a stand-alone chapter that explores the results reported in the previous chapter. The thesis presents the results for both RQs in line with the current understanding of the literature bases for the GBG intervention, preventive interventions generally, poverty, and implementation science. The researcher acknowledges the limitations of the present study at both conceptual and methodological levels, and provide suggestions for

improvements. In the penultimate section, the researcher proposes how the knowledge gained from the study can be utilised ad the research level, policy level and school level. Finally, the thesis is summarised, and an overall conclusion is drawn.

1.8 Chapter summary

At the start of the chapter, the reader was given an account of increased interest in preventive interventions and a brief history of the GBG. Then the researcher outlined the context the present study was situated in, as well as their role in the wider trial and clarified how their experiences as a research assistant may have influenced the direction and interpretations of the present study. Next, the reader was introduced to further contextual information regarding three concepts that would be investigated in the present study. The concepts were poverty, DB and implementation. In the penultimate section, the researcher presented the rationale, research questions and contribution to knowledge. The final section of the chapter laid out the presentation of the thesis.

Chapter 2
Poverty and its impact on children

2.1 Introduction to the chapter

The chapter opens with an exploration of two critical perspectives on poverty: the needs perspective and the income perspective. The suitability of each definition that sits within two perspectives of poverty is assessed. Next, the researcher undertakes an exploration of a variety of theories that attempt to explain the transmission of poverty. Again, each theory is scrutinised in order to determine which is most appropriate for the present study. As it is already well established that poverty moderates a wide variety of child outcomes, the extent to which the length of time spent in poverty, the age at which poverty is experienced, and the methodological decisions used in poverty-based research moderate the outcomes reported within the literature, are discussed. Finally, the literature review examines the school-level factors that can moderate childhood outcomes.

2.2 Defining Poverty

The present study must establish a definition of poverty before exploring the negative outcomes with which childhood poverty is associated. An established definition of poverty will allow the reader to have a better understanding of the context in which children are experiencing poverty develop when reviewing the literature in later sections of the chapter. However, multiple definitions exist for the term "poverty". Therefore it is necessary to critically examine different meanings in order to select the most appropriate one for the present study. The establishment of a definition is pertinent for later in the study when considering a suitable proxy to measure poverty. This section of the literature review focuses on two perspectives: the needs perspective, and the income perspective.

2.2.1 The needs perspective on poverty

The needs perspective operationalises the available definitions of poverty by providing visible indicators such as the low consumption of resources (Ringen, 1988). By proposing that poverty is a lack of access to resources, the needs perspective can describe the personal experiences of disenfranchised members of society and the living conditions they are likely to experience (Davis & Sanchez-Martinez, 2014). This section of the literature review explores three definitions of poverty within the within the needs perspective literature. These are social exclusion, deprivation, and consensual poverty.

Poverty as social exclusion

Social exclusion refers to members of society being unable to access the same opportunities or resources as their more privileged peers (Burchardt, Le Grand, & Piachaud, 1999). This definition includes access to income, work opportunities, services, political

involvement, and social relations (Khan, Combaz & McAslan Fraser, 2015). It is thought that these processes are naturally occurring within society cause alienation amongst certain groups of people from the benefits of participating in society (Levitas et al., 2007). While some researchers generally agree that social exclusion is a multidimensional construct, they often dispute which dimensions make up the construct (Peace, 2001; Silver, 1994; Sen, 2000). This lack of agreement suggests that social exclusion may be better described as a discourse, rather than a definition of poverty, particularly as different paradigms and political ideologies use the term for their explanations behind the power dynamics between the included and excluded (Silver, 1994). With a lack of consensus on dimensions that form the term, it would be difficult to operationalise within a classroom context. Therefore, the present study requires a more suitable definition.

Poverty as deprivation

Deprivation relates to the lack of resources, activities and facilities that are thought to be commonplace with a given society (Townsend, 1979). Examples of these include clothing; housing and household facilities; as well as adequate educational, work, and social conditions. Although deprivation is similar to social exclusion regarding multi-dimensionality, a considerable amount of research has operationalised the definition into a measurable and generalisable construct. Townsend (1979) developed 60 indicators that measured the different facets of deprivation in the 1960s in order to establish the nature and scope of poverty within the UK. Townsend's work has led to the UK government developing similar measures to produce yearly reports on the deprivation levels within the UK (Department for Communities and Local Government: DCLG, 2015). However, the lack of access to commonly available goods and resources is indicative of long-term poverty, which has led to some researchers suggesting that deprivation is a product of poverty, rather than a definition of poverty (Gordon, 2000; Whelan, Layte & Maître, 2004). While individuals experiencing poverty over a short period are likely to reduce their consumption of goods and resources until their economic condition improves, others who have been subject to poverty for a more extended period would continue to decrease their consumption in order to preserve resources and goods that are at their disposal (Callan, Nolan & Whelan, 1993). As deprivation has been conceptualised as a separate construct from poverty, utilising the term deprivation could create an unnecessary conflation of terms.

Perceived poverty

Most definitions of poverty produced in the literature do not take into account the public's perception of poverty. Perceived poverty, also known as consensual poverty, was a response to the rift between the public's understanding of poverty and the academics in privileged positions that typically generate definitions (Veit-Wilson, 1987). The term "perceived poverty" is operationalisable and is established through utilising a representative sample that rank access to services and resources as either "necessities" or "luxuries" (Mack & Lansley, 1985). Researchers then administer these responses in the form of a

checklist to a larger sample, whereby participants indicate whether they can afford/access these items. This process establishes a specific number of "necessary" items to form a poverty line. A major assumption of perceived poverty is that different demographics of society generally are in agreement as to whether an item should be considered a necessity (Gordon et al., 2000). However, research by Mackay (2005) disputes this notion. Evidence from the re-analysis of two large UK datasets created from the above method suggested there was very little agreement on the perceived necessity of items. Mackay also reported that many individuals selected more "luxury" indicators than "necessity" items. The selection of "luxury" indicators suggests that perceived poverty is actually about maintaining an average lifestyle, rather than meeting an individual's needs (Berthoud, Bryan, & Bardasi, 2004). The application of perceived poverty is therefore not appropriate as consumption habits are not an area of focus for the present study.

Criticisms of the needs perspective

The purpose of defining poverty is to be able to easily distinguish children experiencing poverty from the rest of the population, as well as determine an appropriate proxy to achieve this distinction within the sample of the present study. The definitions provided by the needs perspective are not conducive for the creation of arbitrary cut off points as both the research community and the public disagree on the nature and number of indicators that would comprise of poverty (Halleröd, Bradshaw, and Holmes, 1997; Mackay, 2004). Halleröd and colleagues (1997) use the following example to demonstrate how the strict application of cut off points based on the number of necessities lacking within a household could be inappropriate:

> *"…a family lacking three necessities, for example, a garden, a roast meat joint or its equivalent once a week and a washing machine, should have a right to be provided with these things. A family lacking just two necessities, for example heating to warm living areas of the home if it's cold and an indoor toilet, should not have the same right because they are not below the poverty line" (p.218).*

All three definitions described in this section, if applied as a measure of poverty, are in danger of overlooking households who do not access the most fundamental resources despite not falling under the poverty line. The needs perspective also assumes that the consumption of resources follows a logical process, in that essential items or services are purchased or accessed first, and that individuals buy luxury items if they can be afforded, but this is not the case. In the re-analysis of perceived poverty data sets, Mackay (2004) reported that a large proportion of participants who were considered in poverty were consuming "luxury" resources. The assumption that "necessary" items should be purchased first highlights a further question: is it up to policymakers or the academic community to direct the public on their consumption? Unfortunately, such issues are

beyond the scope of the present study. Instead, the concern is that households unable to access the resources they wish to consume will impact on the wellbeing of the children living within these environments. Therefore, the needs perspective of poverty is not appropriate for the current study.

2.2.2 The income perspective of poverty

Definitions of poverty generated from this perspective are concerned with generating discrete categories based on low income. While income varies annually and differs across nations, definitions from this perspective are widely adopted across the world: from individual nations such as the UK to international organisations such as The World Bank (WB). Contrary to popular belief, definitions of poverty based on cut-off points are not static. With variables such as housing prices, inflation, and the price of goods, continually fluctuating, definitions need to be flexible in order to account for changes over time. This section of the literature review will explore three definitions of poverty within the income perspective. These are absolute poverty, minimum budget, and relative poverty.

Absolute poverty

Absolute poverty is when an individual or household does not have the minimum income required to meet the most basic needs, which are universally agreed to be: food, clothing, sanitation, and shelter (United Nations: UN, 1995; WB, 2017). Although international organisations have cut-off points, for instance, WB (2017) uses one $1.90 per day; such thresholds are inappropriate for developed countries as the cost of consuming resources that meet basic needs for a single day cost more than one US dollar. The cut-off points set by international organisations are more appropriate for low-income countries where a good basic standard for living is not accessible to a large proportion of their respective populations compared to high-income countries (Casazza, 2015). However, developed countries such as the US and the UK generate their margins for absolute poverty based on social relevancy: in that a household's disposable income would be able to afford the very basic amenities required to meet the standards set by the UN and WB (Casazza, 2015). Developed states set their poverty margins as members of society tend to compare their income and resource consumption with one another (Ravallion & Chen, 2011). If prosperous nations set their cut-off points based on social relevance, this suggests the margins are relative rather than absolute, then it would be more appropriate to adopt a relative cut-off point for the present study explicitly.

Minimum budget standards

A minimum budget standard establishes the lowest income required for households to maintain a basic standard of living based on the actual market value of the services and goods Padley & Hirsch, 2017). This definition has played a vital role in the set-up of the welfare state, in that households falling below this minimum accepted standard are given a weekly income to meet the minimum standard (Beveridge, 1942). Meanwhile, the Joseph

Rowntree Foundation (JRF) calculate a minimum income standard annually by taking into account both the public's opinion and expert's opinion on the level of income is required to sustain a healthy and socially acceptable minimum standard of living (Padley & Hirsh, 2017). Although JRF uses a process called equivalisation to account for the differing costs required for different forms of household[1], more recent terms of UK government have used a similar, but different definition (see below) to establish cut off points that determine whether households can access the avenues government provide in terms of social security and benefits (McGuinness, 2018). Despite the ongoing debate regarding government's provision of adequate intervention for members of the public experiencing poverty, the proxy adopted by the present study (Free School Meal eligibility: FSM) is based on the definition adopted by the state. It would, therefore, be more representative of the current understanding of the nature of poverty in the UK by adopting the accepted definition.

Relative poverty

As already established in the absolute poverty section, wealthy nations have higher minimum standards that all members of the public should be able to achieve compared to developing countries. As affluent states' economic growth increase and decrease, based on factors such as trade and consumer spending, the minimum standard is subject to change too (Mack, 2016). Therefore, the rationale behind the definition is that a threshold set at a certain point below the national average income indicates that a household is unable to access the resources required to participate fully in society (Nolan, 2007). The majority of states (including the UK) that use relative poverty as their official definition when generating policy and government documents use 60% of the median income of the nation as the poverty-line (Lelkes & Gasior, 2011).

One major criticism of using a poverty-line based on the median is that the value of the threshold changes is based incomes of households in the middle range. Mack (2016) argues that during times of economic shrinkage, households at the lower end of the distribution scale appear to be "lifted" out of poverty, but it is just the reduction in income for wealthier households lowering the median value. However, the cut-off point is easy to understand, calculate, and requires minimal effort to update annually compared to other measures (Zheng, 2001). The 60% median threshold also allows for the comparison of poverty trends over time as well as the comparisons between other countries (Nolan, 2007). With this in mind, relative poverty is the best available definition of poverty for the present study.

Relative poverty and the present study

The principal reason the researcher chose relative poverty over other definitions of poverty was that the UK government uses the 60% median threshold to determine policy. The use

---

[1] See 'relative poverty in the present study' for a further explanation of the term 'equivalisation'.

of the definition within policy is particularly relevant in education research, as it is often beyond the scope of the researchers to collect income data on the participants. Therefore, proxies such as pupil premium (PP) or FSM eligibility are used to capture child poverty data. The combined income of the child's household determines their eligibility for the benefit. As households vary in size and composition, a process called equivalisation is used to adjust incomes for the different households in order to draw comparisons. Equivalisation is because a single person requires less money to enjoy a good standard of living compared to a lone parent with children or a couple with children. Table 2.1 demonstrates the raising of the poverty line when the household size increases, particularly in the number of children less than 14 years of age or number of adults living in the household.

Table 2.1
The poverty threshold for different household compositions adapted from Department for Work and Pensions (DWP: 2017).

N of children		Lone parent		Couples	
>14 years	≤14 years	Monthly Income[1]	Annual Income[1]	Monthly Income[1]	Annual Income[1]
1	0	£957	£11,484	£1326	£15,912
1	1	£1178	£14,136	£1547	£18,564
1	2	£1339	£16,068	£1768	£21,216
2	2	£1620	£19,440	£1989	£23,868

[1]Before housing costs.

As previously mentioned, it would be unfeasible to collect the necessary data to determine the poverty status of the participants in the current study, so the researcher decided that FSM would be the proxy of choice instead. For more information regarding whether FSM was a suitable proxy for the chosen definition of the study, see section 5.6.1.

2.3. Theories of poverty

Although a definition of poverty is useful for the present study, it does not explain as to why certain groups are at risk of economic hardship. Therefore, this section will establish an appropriate theory of poverty for the present study. The provision of a framework that explains the drivers of poverty can lend insight into children's proximal and distal outcomes, as well as a foundation for investigating the effects of poverty at the school level. Two branches of economic theory provide explanations for impoverishment: orthodox theories and heterodox theories.

2.3.1 Orthodox theories of poverty

Orthodox theories of poverty have been formulated using the critical tenets of mainstream economic principles (Davis & Sanchez-Martinez, 2014). The first orthodox theories of poverty were developed in the 19th century, forming the basis of formal investigations within the UK (David & Sanchez-Martinez, 2014). Orthodox economic models assume that wages

are reflective of an individual's level of productivity, which is defined as the rate at which goods or services can be produced (Chew, 1988)₂. Advocates of the orthodox theory argue that poverty is a consequence of either: low productivity or poor individual choices (Davis & Sanchez-Martinez, 2014). It is important to stress that low productivity is not a synonym for "laziness", as factors such as education, skills, and previous experience determine an individual's productivity level (Thurow, 1972).. Therefore, poverty as a consequence of low productivity may be out of the individual's control. Classical economic theory (CET) and neo-classical economic theory (N-CET) are the two models that fall under the orthodox school of thought.

Classical economic theory

CET argues that economic markets function optimally with minimal government interference (Chang, 2014), when economic growth curbs, advocates of CET proposed that naturally occurring mechanisms within the market will, in the long term, allow for economic recovery (Smith, 1976). For instance, if there is a reduction in growth, companies and institutions respond by reducing their outputs, resulting in their employees being less productive (Smith, 1976). CET also argues that an individual's wage is an accurate reflection of their productivity, so during economic hardship, wages must be reduced to reflect the decrease in output (Case, Fair & Oster, 2017). During this period, employees can either accept or reject these revisions; but by rejecting a pay reduction, individual's "self-select" into poverty (Kasper, 1967). However, many individuals find themselves in economic hardship despite being employed; this is particularly true within affluent states (Pradella, 2015).

Nevertheless, CET argues that rather than redistributing wealth, it is in the economic market's long-term interest to allow the capital class (which consists of individuals who own the means of production) to receive the majority of profits because of their ability to invest large amounts of money and stimulate further economic growth (Chang, 2014). Whereas the working class (the individuals who use the means of production to create the products which are sold for profit) do not have much influence within the economic market which restricts them to saving or investing (Ricardo, 1821; Say, Prinsep & Biddle 1832). On the surface, this argument sounds like it is advocating poverty amongst the labouring class, but this would only be valid if employers were purposely paying a wage that was below the poverty line.

Despite this, CET is an out-dated perspective of economics as it was established during the industrial revolution when the majority of the capital class were the owners of factories or mines that produced and traded physical goods (Chang, 2014). These principles do not reflect the current economic landscape, which is predominantly influenced by the stock-

---

₂ The definition used here for 'productivity' is somewhat simplified. Amongst economists, there are disagreements about the nature of the term. For more information see Tangen (2004).

market. Further to this, history has shown that the lynch-pin of CET does not stand up to the realities of modern economic markets: they do not act rationally (Ariely, 2008). For instance, the last two global financial disasters were caused by risky financial decisions, which ultimately placed a large proportion of the population across the world in economic hardship (Chang, 2014; Ariely, 2008). Poorly thought financial decisions made by those in positions of power often result in many people forced into poverty through means such as large-scale redundancies in order to 'balance the books' (Kasper, 1967). In light of this, it seems the concept of 'self-selecting' into poverty is inappropriate as those who are at risk of economic hardship are not the individuals driving major financial decisions that could potentially displace large numbers of people. It is, therefore, an out-dated assumption that individuals are completely autonomous in their ability to avoid economic hardship. In conclusion, CET is an inappropriate theory to explain poverty for the current study.

Neo-classical economic theory

N-CET is considered to be the dominant school of thought in today's economy (Chang, 2014). While both orthodox theories assume individuals at all levels of the economic market engage in rational decision-making, N-CET argues that marginal utility has a crucial role in executing economic decisions (Chang, 2014). Marginal utility is the additional satisfaction gained from consuming one unit of a good or service and is thought to be the key factor in explaining an individual's consumer habits (Kauder, 1953). Traditionally, economists suggested that the more resources an individual has, marginal utility decreases (Stigler, 1950a; Stigler 1950b). For instance, an individual with a weekly income of £100.00 will perceive a higher utility for an increase of £50.00 per week compared to an individual who has a weekly income of £1000.00. For the individual with lower earnings, this increase in money represents half their income and would allow them to participate in society more comfortably compared to their original income. Meanwhile, for the individual with higher earnings, the additional money represents 5% of their weekly income, and so the additional money will not provide as much satisfaction for someone who can participate in society comfortably.

However, Karelis (2007) argues that marginal utility works differently if products or services provide a feeling of relief rather than satisfaction or pleasure. Relief-based marginal utility is often attributed to debt, a frequent contributor to poverty in the present society (Hartfree & Collard, 2014). For example, if an individual has multiple debts, paying off the first debt is unlikely to make as much of a difference to their living conditions compared to paying off the last debt. Therefore the individual is likely to attribute a higher added value to the last debt compared to the first debt. Applying marginal utility in this way has major implications in explaining the spending habits of individuals experiencing poverty (Karelis, 2007). As there is such a high value in alleviating the negative feelings associated with experiencing poverty, when their income is first awarded, individuals often spend it to participate fully within society, and halt both the psychological and physical discomfort of experiencing

poverty (Karelis, 2007). Although the "rational" solution would be to spend within the means of the budget in order to slowly lift oneself out of poverty, this would provide no short-term feelings of relief from their current situation. While N-CET acknowledges that rational decision making is rare within the field of economics, the application of marginal utility is usually as its original form. The improper application of marginal utility has led to a discourse claiming that individuals experiencing poverty or in large amounts of debt need to "save more" or "better handle" their finances without understanding the value such individuals place on the complete relief of financial problems (Karelis, 2007; Günther & Maier, 2014). With this in mind, using the default assumption that marginal utility decreases as resources accumulate may not be appropriate for the present study's context, where the "norm" is a lack of resources. An appropriate theory for the present study would acknowledge that norms can vary within an economic landscape.

Criticisms of Orthodox Economic Theories

By rejecting the need for government to have an active role in fiscal and monetary policies to regulate the economy, orthodox theories place responsibility for economic hardship at the individual level (Davis & Sanchez-Martinez, 2014). Although N-CET appears more sympathetic towards those experiencing poverty compared to CET; there is still an assumption by both theories that affluent and influential individuals are making rational economic decisions. The discourse around rational decision making remains dominant within economics despite substantial evidence suggesting that regardless of socio-economic status, individuals do not make rational financial decisions (Ariely, 2008). Furthermore, throughout the history of capitalism, where there have been cycles of regulated and deregulated economic markets, the intervention of the government is associated with better performing and more stable economies (Chang, 2014). Nevertheless, the general response to economic uncertainty is to fall back on the "trickle-down effect" of reducing the barriers for the production and trading of goods and resources for richer contenders in order to stimulate investment in new ventures that would benefit the rest of the economy (Smith, 1976). However, reducing constraints for one area of the economy requires increasing constraints for another, and this usually affects the most vulnerable population within society. It is from a moral stance that the researcher dismisses orthodox economic theories as a basis for the present study as the researcher argues that individuals are not solely responsible for the poverty they find themselves in.

2.3.2 Heterodox theories of poverty

The term heterodox means to not conform to the accepted belief or standards (Stevenson, 2010). In this case, the theories explored in this section reject the notion that rational decision making allows the economy to self-regulate (David & Sanchez-Martinez, 2014). The degree to which these theories reject the classical and neo-classical principles vary. For instance, the Keynesian theory argues that pure rationality is impossible within economics, and therefore the government should make alterations to shield the vulnerable

from risky macroeconomic decisions (Chang, 2014). Meanwhile, Marxism theory is more sceptical of the intentions of those in positions of economic power and proposes that the government should have complete control of the economic space of the state (Marx, Fernbach & Mandel 1867). However, Social Capital Theory (SCT) shifts the focus from macroeconomics[3] to the social influence individuals have in order accrue and trade resources (Coleman, 1988). The present section of the literature review will determine the suitability of each heterodox theory for the current study.

Keynesian economic theory

Keynes (1937) developed this theory of economics in response to The Great Depression and proposed that the primary driver for economic growth is aggregate demand. Aggregate demand is the spending total of all the different sectors that form an economy (Case, Fair & Oster, 2017). Advocates of the Keynesian theory argue that without government intervention to maintain steady economic growth, sharp increases in aggregate demand can lead to economic "bubbles" which will eventually lead to recessions (Kindleberger, 1978). The government's role is, therefore, to balance the spending and investment across the different sectors (through both monetary and fiscal policy) to encourage society to reach full employment (Chang, 2014). Poverty is, therefore, a reflection of market failure (Davis & Sanchez-Martinez, 2014). If aggregate demand is low, the different market sectors further scale back their production to compensate for the lack of demand of their goods and services, which results in wage reduction and higher levels of unemployment, two major contributors to poverty.

On the surface, the Keynesian economic theory seems to be an appropriate concept to explain poverty, as it recognises the involvement of government decisions and its effect on the depth and breadth of poverty. However, as a macroeconomic theory, it is difficult to apply to the context of the present study, which focuses on the experiences of poverty at the neighbourhood and individual level (Chang, 2014). The application of this macroeconomic theory to a specific context such as schools serving impoverished catchment areas is that the financial situation of the school is a less pertinent issue compared to the other problems faced by schools in such neighbourhoods (see section 4.4). Furthermore, the Keynesian economic theory can explain unemployment rates and low wages, but it is less able to explain the geographic clustering of poverty often documented within neighbourhoods across villages, towns and cities (Chang, 2014).

Marxist theory

Although the formation of Marxism is argued to be a retaliation to capitalistic and economic ideologies such as CET, Karl Marx and Fredriech Engels, the founding fathers of the

---

[3] Macroeconomics is concerned with the monetary and fiscal processes that occur at the state and international level.

movement, viewed the socialist alternative as an improvement on the achievements of capitalism (Chang, 2014). In Marxist theory, there are two social classes: the proletariat, and the bourgeoisie (Marx, Engles, Moore & McLellan, 1992). The proletariat is the labour force whose only "property" is the skills they provide in exchange for wages; whereas the bourgeoisie owns the businesses that utilise the labour force (the proletariat) that is responsible for a generation the profits proletariat (Marx et al., 1992). While Marxist theory acknowledges that competition between companies and the different sectors of the economy drives growth, it also places pressure on the bourgeoisie to maintain an increase in profits (Case, Fair & Oster, 2017). Marxism argued that the cost of labour was kept lower than its actual value, forcing the proletariat into poverty while the bourgeoisie was able to remain in the economic competition (Marx, Engles, Moore & McLellan, 1992). Marx (1867) hypothesised that this class divide would increase until the proletariat rebel and dismantle the capitalistic economic model in order to build a socialist model where a central system (owned by all employees) would distribute wealth equitably.

While Marxist theory inspired the development of communist frameworks in countries such as China and Russia, the western capitalistic model has made more substantial strides in economic and social issues that concern the public (Chang, 2014). The failure of most communist frameworks is partly due to underestimating capitalism's ability to adapt and reform itself — for instance, other socialist thinkers such as Bernstein and Harvey (1961) criticised Marxist theory for suggesting revolution was the only means of reducing the class divide. Instead, there were suggestions put forward that a revisionist approach would be a superior approach, whereby the government slowly introduced reform (Bernstein and Harvey, 1961). Evidence of such practice exists today in many western countries, with legislation such as the living wage being introduced, trialled and then revised based on feedback from government and the public (Chang, 2014). With this in mind, the present study's context reflects a capitalist model that looks very different from Karl Marx's era. While it may not always serve the interests of the vulnerable, revisions by government sometimes appear to make positive impacts. Therefore, a theory that recognises this interplay would be more appropriate for the present study.

Social capital theory
The overarching principle of SCT is that the relationships a person has within their social network allow them to access to resources (Coleman, 1988). Factors such as the number and quality of connections an individual has within a social network influences the kinds of resources they can access, as well as the quantity (Putnam, 2000). According to SCT, the social landscape an individual operates in can contribute to their experience of poverty in some ways. For instance, if an individual has poor social-emotional skills, they may find it hard to establish new relationships with people who can improve their economic standing (Putnam, 2000). Further to this, SCT argues that social connections are transactional so an individual wishing to establish a new connection must provide services, skills or goods which

benefit the other party (Coleman, 1988). Therefore a contributing factor to poverty through a lack of lucrative connections is due to a lack of skills or resources to offer others. A lack of useful connections is a common issue amongst individuals who are trying to access the job market after experiencing unemployment for a substantial period, but a key driver in various "gaps" reported above and below the poverty line (e.g. Janlert, Winefield & Hammarström, 2014; Kim & von dem Knesebeck, 2016).

Alternatively, the broader social network may perpetuate an individual's experience of poverty through the encouragement of maladaptive behaviours. For instance, unlawful activity such as violent behaviour or vandalism may lead fewer job prospects such as a criminal record would deter employers because of the negative connotations associated with it (Pager, 2003). Meanwhile, engaging in substance use will not only harm employability, but it will further reduce resources in order to fund addictions (See Henkel, 2011 for a review). While these points do not necessarily reflect the majority of the population experiencing poverty, peers can still influence poor decisions regarding the expenditure or investment of income, as well as lifestyle choices which may hinder an individual's ability to stay above the poverty line (Putnam, 1995). The aforementioned routes to impoverishment proposed by SCT help to explain the substantial evidence base for the geographical clustering of poverty (see section 4.4.1), and the associated risk this has for the health, attainment, and wellbeing of those experiencing poverty (see sections 2.4, 1.5 and 3.2.3). It would, therefore, seem that this particular theory applies to the context of the present study for two reasons. Firstly, SCT provides an explanatory function towards the link between maladaptive behaviours often reported amongst individuals experiencing poverty, as well as providing some insight as to why reversing these behaviours may be difficult. Secondly, of the three theories, SCT is applicable at multiple levels of society, from explaining nationwide "gaps" between the affluent and non-affluent, to individual experiences of poverty (Brehm & Rahn, 1997). Therefore, SCT was selected to form part of the theoretical framework of the present study, as it was the most applicable of the theories examined within this section of the literature review.

2.4 The Impact of poverty on children

This section explores and critically evaluates the literature surrounding the impact and outcomes associated with experiencing poverty during a child's development. Research suggests that experiencing socio-economic hardship has negative effects on many facets of a child's life, both proximally and distally, with seminal pieces reviewing the extensive literature such as Duncan and Brooks-Gunn (1997) and McLoyd (1998). However, research should acknowledge that child poverty does not occur in isolation. As suggested by social capital theory, an individual's social network will likely comprise of other people with similar ability to access resources (Bourdieu, 1972). Children facing socioeconomic hardship tend to live in impoverished communities, with schools that serve these areas. Therefore, it is

vital, when reviewing the literature, to acknowledge that these complex interactions that could influence underprivileged children's outcomes, not just on an individual level, but at higher levels too.

2.4.1. Factors influencing childhood poverty outcomes

The key purpose for critically reviewing this area of literature is not only to establish current understanding of how poverty impacts on children throughout their lives, but also to explore the extent to which these differential gains occur. There is a wealth of research that has been conducted to gain insight in both the proximal and distal implications associated with experiencing economic hardship, particularly the educational, health, and behavioural outcomes. Systematic reviews and meta-analyses tend to partition sections according to outcomes (for examples see: Duncan & Brooks-Gunn, 1997; McLoyd, 1998). Previous synthesises of poverty literature that utilised this structure tend to address the question "Does impoverishment impact on child outcomes?". With the wealth of empirical evidence available, this focus seems outdated. Instead, researchers investigating poverty should instead ask "what factors moderate the extent to which childhood poverty impacts on outcomes?". Using a similar structure and themes used in Sirrin's (2005) meta-analysis, the researcher has identified key factors that may moderate the magnitude of the association between economic hardship and the outcomes of interest. Therefore, this section will focus on three factors that may influence childhood outcomes. These are persistence of poverty, the developmental stage at which poverty occurs, and the methodology used to measure poverty.

Persistence of poverty and outcomes for children

Although children experiencing transitory poverty tend to make smaller gains in terms of cognitive, health, and behaviour outcomes compared to never-poor children (e.g. Dubow & Ippolito, 1994; Garret, Ng'andu & Ferron., 1994; Aber, Bennett, Conley & Li, 1997; Pagani et al., 1999); researchers have consistently found that the length of time a child spends in impoverished conditions is detrimental to their outcomes (Duncan, Magnuson, Kalil & Ziol-Guest, 2012; Duncan & Brooks-Gunn, 1997; Korenman, Miller & Sjaastad., 1995; Smith, Brooks-Gunn & Klebanov., 1997). The above finding, in part, has led to the adoption of the cumulative effect model as a possible explanation regarding how poverty (amongst other risk factors) moderates development and its associated outcomes. This model argues that the longer a child is exposed to poverty (and the risks associated with it) the more adverse the consequences are likely to be (Deutsch, 1973). For instance, the disparity of cognitive functioning between children experiencing persistent poverty and their more affluent counterparts is likely to mediate the academic attainment the prospective groups can attain (see Duncan et al., 2012; McLoyd, 1998; Duncan & Brooks-Gunn, 1997).

This achievement gap is a crucial trend in the DfE report on a yearly basis. The latest figures that are available from the Department for Education (DfE, 2017) report that at the end of

Key Stage 2 there was a 21% difference between FSM eligible children and their non-eligible counterparts in meeting the expected standard for reading, writing and mathematics. However, the achievement gap is larger by the end of Key Stage 4, where there is a disparity of 27.6% between the two groups achieving 5 A*-C GCSE grades (DfE, 2016b). However, direct comparisons of cross-sectional data provided annually should be approached with caution; pupils further along their education trajectory have had their experiences shaped by the policies and curriculums that are different to the present state of education which may contribute to this increase. Despite changes in policy, political agendas, and curriculums, longitudinal studies using nationally representative data consistently find differential gains in educational, health, and behavioural outcomes between children experiencing poverty and their more affluent peers, both nationally (e.g. Wickham et al., 2017; McKenna, Law & Pearce, 2017; von Stumm, 2017) and internationally (e.g. Duncan et al., 2012; Cantillon,Chzhen, Handa & Nolan., 2017; Gregg, Propper & Washbrook, 2007).

It is thought that allostatic load could be responsible for mediating the relationship between poverty and associated outcomes (Duncan et al., 2012; Evans, Chen, Miller & Seeman., 2012). Allostatic load is the wearing down of coping mechanisms when exposed to chronic stress. Allostatic load may be related to poverty because it is a stressful experience, which over long periods of time puts coping mechanisms under strain (Evans & Kim, 2012; Nusslock & Miller, 2016). Emerging research has found that this wear and tear has long-term effects on the structural components associated with the cognitive (e.g. Tine 2014; Farah et al., 2006), health (e.g. Apouey & Geoffard, 2013; Pearce, Lewis & Law, 2013), and behavioural (e.g. Bernard, Zwerling & Dozier, 2015, Mazza et al., 2016; Lansford et al., 2018) outcomes of children. If children spend long periods in adversity due to low income, already strained biological components, particularly the structural atrophy or hypertrophy of key areas of the brain could, therefore, increase the likelihood of developing mental health disorders such as later in life (Rogosch, Dackis & Cicchetti, 2011).

Poverty in the early years and its impact on children
There is a substantial amount of empirical evidence suggesting that early childhood is a critical developmental phase (for an overview see Shonkoff & Phillips, 2000), and that experiencing adversity or neglect during the first few years of life has a large impact on children's outcomes (Duncan, Brooks-Gunn & Klebanov, 1997; McLoyd, 1998; Votruba-Drzal, 2006). For instance, Duncan and colleagues (1997) found that children who experienced economic hardship within the first 5 years of life had more substantial detriments to their academic outcomes compared to children experiencing poverty during middle-childhood or adolescence. Between birth and age 5, children make the most rapid advances in their development, particularly in the areas of cognition and wellbeing (Shonkoff & Phillips, 2000). It is thought that experiencing the chronic physiological stress associated with living in poverty during early childhood puts an individual at risk of having a higher allostatic load later in life (Evans & Kim, 2012; Evans & Schamberg, 2009; Evans et al.,

2012). The association is because the physiological systems that respond to the stressful demands of the environment become less responsive and recover less effectively over time (McEwen & Gianaros, 2011). Exposure to stress is particularly problematic in early childhood because allostatic load causes lasting neurobiological changes to areas of the brain that are not yet fully developed (Kim et al., 2013), which may explain why children who are born into poverty have worse outcomes compared to peers who experience poverty at a later period of childhood.

It would also seem that early childhood is the period in which children are at the highest risk of experiencing poverty (Bronfenbrenner et al., 1996). With not only the highest amount of neural development occurring in the first five years of life, but such rapid physical growth, parents with already few resources will find it difficult to provide the nutritious meals, clothing, and environmental stimulation (e.g. through toys and books) required to give children the opportunity to thrive. Not only will the lack of resources directly impact a child's development, but the stresses of raising a child during economic hardship can affect the relationship parents have with their child, with each other, as well as their own mental health (Finegood et al., 2017; Shaw & Shelby, 2014; Sosu & Schmidt, 2017). Parents with a lower socio-economic status are more likely to engage in a punitive approach to rearing, engage in conflict in front of their children, or neglect their children's needs when experiencing mental health issues (Cooper & Stewart, 2017; Conger, Conger & Martin, 2010). All of these instances have been associated with higher levels of problematic externalising and internalising behaviours in children compared to their peers (Bøe et al., 2014; Sosu & Schmidt, 2017; Mazza et al., 2016), and may explain why there is an over-representation of single mothers in samples from studies investigating the interaction between impoverished family environment and children's future wellbeing. It is unsurprising that single-parent households will have access to fewer resources, thus straining the psychological well-being more for single parents compared to two-parent households (Lange et al., 2017).

Research consistently indicates that children born from mothers who were experiencing poverty during pregnancy are more likely to be born prematurely and underweight (Shenkin, Starr & Deary, 2004; Parker, Schoendorf & Kiely, 1994). Such children are more likely to have poorer health, cognitive, and wellbeing outcomes in childhood and adolescence (Duncan & Brooks-Gun, 1997). If a mother is unable to provide herself with the adequate nutrition, resources and services to keep her and baby healthy during pregnancy, it is unsurprising that these already scarce resources become further stretched after the birth. The consequences of indirectly experiencing poverty during gestation has further implications in terms of the allostatic load model, in that such children are born with coping mechanisms that are already likely to be under stress due to inadequate input during prenatal development, which is likely to further weaken with the persistence of poverty (Conger, Conger & Martin, 2011). While it is important to acknowledge that poverty is not a

one-size-fits-all experience for children, both the length and the time point in which they experience economic hardship is are both key factors that affect the developmental milestones and outcomes for children. The methodological qualities of the studies investigating poverty outcomes may also affect the magnitude for these observed differences.

Methodological decisions moderating the outcomes of impoverished children

While there are many methodological decisions that affect the outcome of interest, there are three particular aspects that are of interest in the present literature review. The first is the choice of proxy utilised as a measure of socioeconomic status. Sirrin (2005) concluded from his meta-analysis on child poverty and academic attainment, the choice of proxy for poverty affected the extent to which the moderator could be considered to impact on the effect sizes that were reported for academic attainment. Indirect measures of poverty yield smaller effect sizes (e.g. parental education: k_4=30, ES=.30; parental income: k= 15, ES= 0.29) compared to proxies based on the needs perspective (e.g. home resources: k= 4, ES= .51) which are considered to be more direct methods of measuring poverty. Although the literature is lacking similar reviews with regards to proxies of poverty moderating other childhood outcomes, relatively recent studies reflect a similar trend for both approaches to measuring poverty (Gregg et al., 2007; Violato et al., 2010). While direct measures of poverty show larger effect sizes compared to indirect proxies, within education research, indirect measures are often easier to obtain (Hobb & Vignoles, 2007). Therefore, interpretation of the many studies investigating poverty that used indirect proxies may not yield impressive effect sizes. It is likely due to a methodological compromise rather than poverty lacking impact on the outcome variable.

The second methodological decision that may moderate the impact poverty has on child outcomes is the design of the study. Although RCTs surrounding income are rare due to their obvious ethical implications, studies that have used this design report higher effect sizes for children experiencing poverty compared to studies utilising a quasi-experimental design (e.g. Cancian, Yan & Slack, 2013; Duncan, Morris & Rodrigues, 2011). Furthermore, observational studies yield lower effect sizes compared to both forms of experimental design, particularly within the fields of health, behavioural, social, emotional, and cognitive outcomes (Cooper & Stewart, 2017). For instance, Gennetian and Miller (2002) analysed data from a RCT and found that children in households that had received financial aid were less likely to display problem behaviours (ES= -.16), whereas quasi-experimental studies such as Hamad and Rehkopf (2015) and Milligan and Stabile (2011) reported effect sizes for behaviour between .07 and .10. This is still larger than those reported by observational methods (e.g. Dearing et al., 2006: .02-.03; Votruba-Drzal, 2006: .01; Zachrisson & Dearing,

---

4 refers to the number of effect sizes, not the number of studies. This is because the studies featured in the meta-analysis tended to utilise multiple indicators within their methodology. See Sirrin (2005) for an in-depth discussion regarding this topic.

2015: .02). Although differences in study designs appear to contribute to small differences in effect sizes, the outcomes often measured in child poverty literature usually require a large amount of input to enact a small change. Therefore, study designs should still be considered when critically reviewing the extent to which poverty impacts the development of children.

The final methodological decision that may moderate the impact poverty has on child outcomes is the proxy used to represent the outcome variable. For the different facets of a child's development, there are many ways behavioural, wellbeing, cognitive outcomes can be measured and the choice of the measure will affect the effect size produced. For instance, studies interested in poverty's influence on child health outcomes often measure the birth weight or the height of the child participant; these often yield differing effect sizes. Both Fernald, Gertler and Neufeld (2008) and Milligan and Stabile (2011) conducted quasi-experiments and used height for age as the outcome variable to represent children's physical development. They reported effect sizes of .24 and .04 respectively. Meanwhile, Mocan, Raschke and Unel (2015) and Chung, Ha and Kim (2016) also conducted quasi-experiments but used birth weight as their outcome variable and reported smaller effect sizes (.01 and .02). However, the majority of research regarding the impact poverty may have on child outcomes uses psychometric measures to represent the construct of interest; this is particularly the case for behavioural, cognitive and social-emotional outcomes.

The implication of this is that the reliability of the measure has an impact on the effect size that is generated (Coe, 2002). In an ideal situation, researchers would pick the most reliable measure for their chosen outcome; however, the reality is that other factors such as time, money, or a lack of understanding of psychometric properties mean researchers must compromise when choosing from a wealth of measures. For instance, several studies have used the Behavioural Problems Index (BPI) to assess the extent to which poverty influences behavioural outcomes (Gennetian & Miller, 2002; Milligan & Stabile, 2011; Hamad & Rehkopf, 2015; Blau, 1999; Votruba-Drzal, 2006). However, BPI was a shortened version of a checklist developed by Achenbach and Edelbrock (1981). Both measures have very little information regarding their psychometric composition available, thus suggesting that the reliability of the measure may not be up to standard. The lack of credible measure may be a factor as to why Blau (1999) and Votruba-Drzal (2006) reported low effect sizes (.01 and .02) from their observational studies. While only a couple of examples have been provided, they are demonstrative of the importance of the careful consideration that a researcher must give to the operationalisation of the outcome variable as the decision has implications when determining the extent to which poverty influences childhood outcomes.

Childhood poverty and the present study

It no longer seems appropriate to discuss whether poverty influences childhood outcomes, as a wealth of evidence indicates that in almost every facet of a child's development can be

hindered through experiencing impoverishment. Instead, the present study explores the extent to which poverty influences childhood outcomes. There is no straight-forward answer as factors such as when the poverty started, the length of time spent in impoverished conditions, and the methodological considerations of the study further complicate the debate. However, it is vital for research to continue to investigate poverty from the multiple perspectives, developmental time points, and methodological choices available in order to continue building a truly holistic understanding of how poverty impacts children throughout their lives and beyond, in the hope such knowledge could be used to diminish the inequality gap.

With this in mind, the present study focuses on the behavioural outcomes of children, as this was the primary outcome of the tested intervention: The Good Behaviour Game (GBG: see section 3.4). In order to justify this focus, the literature review will provide a more refined focus on the moderating effects poverty has on behavioural outcomes in later sections (sections 3.2.3 and 3.2.4). However, the present section of the literature review has provided an opportunity to explore the methodological issues surrounding the conducting of studies within the field of poverty. These points have informed the methodological decisions the researcher has made for the present study and are further discussed in the methodological section (section 5.6.1).

2.4.2 School-level poverty moderating child outcomes

In this section of the literature review, the role of the schools is explored to determine the extent to which this environment influences children's outcomes. The roles the broader environment plays a vital role within an individual's development (Bronfenbrenner, 1993). As it is common for the majority of children in the UK to attend a school within their neighbourhood, there may be some features of schools serving impoverished areas that may further impact on the childhood outcomes of individuals experiencing poverty.

The school environment may play a role in moderating the outcomes of children experiencing poverty, with a positive climate serving as a protective factor, and a negative climate providing additional risk (Hopson & Lee, 2011). The school climate is multi-dimensional construct concerned with the environmental quality of a setting (Tagiuri, 1968). The school population and the wider setting the school resides in influence the school climate (Coleman et al., 1966). Although the definition of school climate provided by Tagiuri (1968) is vague, a recent review by Wang and Degol (2016) supports the definition by suggesting the term encompasses the day to day school experience. The school climate may comprise of dimensions such as the quality of teaching, school community relationships, organisational capacity, and the structural features of the school site (Wang & Degol, 206; Tagiuri, 1968). Schools with a higher proportion of their population experiencing poverty are more likely to have climates that are less supportive of the development of both academic and behavioural outcomes (Gottfredson et al., 2005;

Stewart, 2003; McEvoy & Welker, 2000). For instance, a key climate factor that influences childhood development is the quality of teaching individuals experience when they are at school (Clotfelter et al., 2007).

The literature suggests that the number of years teaching is often a good indicator of teacher quality, with teachers who have been practising for under three years being considered to be of lower quality, as they are still developing their skillset (Clotfelter et al., 2007). Schools with a population that consist of a high concentration of children experiencing poverty has higher numbers of teachers who are relatively new to the teaching profession (Clotfelter et al., 2007; Goldhaber, Lavery & Theobald, 2015). This is a pertinent issue within the UK education context where there is a shortage of teachers, and an increased awareness of a number of schools, particularly ones that serve disadvantaged communities which appear to be using a "recruit-burnout-replace model" (Sims & Allen, 2018, p.28) with the high proportion of newly qualified teachers (NQTs) that are employed by the school. The NQT year is the most stressful year for teachers, with evidence suggesting teacher stress further impacts on the quality of teaching (Sims & Allen, 2018). For instance, children attending schools with higher proportions of impoverished children were less likely to receive positive interactions from their teachers (Pianta et al., 2002), which could compound the pupils' SEL and wellbeing outcomes as teachers are often considered to be role models to students.

Another factor that influences the degree to which a school has a favourable climate is the organisational capacity of the school. School organisational capacity is the collective power to improve pupil outcomes (Young & King, 2002). Dimensions of organisational capacity include: the number of staff serving the pupil population, the hierarchical structure the staff are assembled within, development and implementation of policy, as well as the allocation of the resources and time to achieve successful implementation of school policy (Young & King, 2002). Schools serving impoverished neighbourhoods often have a lower organisational capacity compared to more affluent schools (Stosich, 2016; Young & King, 2002). The above association is because schools within impoverished communities are often serving a larger than average proportion of pupils with additional needs such as special educational needs and disabilities (SEND) compared to more affluent schools (Anderson, Hollinger & Contay, 1992; Lupton, 2005). Although such schools in the UK gain access to more funding than their counterparts in order to compensate for this, there is debate as to whether giving additional income to schools with low organisational capacity is the most effective way to improve childhood outcomes (Wang & Degol, 2016). This may be because the organisational structures that are in place may not be able to effectively utilise and allocate resources to allow staff to implement policy and teaching strategies effectively.

Schools serving impoverished neighbourhoods and the present study

The broader context of a school is an integral factor that contributes to a child's development, but it is often over-looked within educational research. Meanwhile, the majority of studies that do focus on the wider context are based within the US in predominantly urban minority communities (e.g. Evans, 2006; Duncan & Brooks-Gunn, 1997; McLoyd, 1998). While such studies provide valuable insight, the extent to which these apply to the UK context is questionable, as there are many differences in policy, culture, and the composition of neighbourhoods between the two western countries. Therefore it is essential to begin similar research within the UK context in order to determine the impact school-level, and community-level factors influence the outcomes of children experiencing poverty. With this in mind, the present study includes an equal measure of poverty among the student population a school caters to as part of establishing whether preventive interventions provide differential gains for children experiencing poverty. There is also further examination of school-level factors within the literature, as these are also thought to influence the implementation outcomes of preventive interventions (see sections 4.3.4 & 4.4). By acknowledging that there are broader influences on a child's development, research is able to provide a more detailed and accurate understanding of how poverty affects children.

2.5 Chapter summary

First, a range of definitions describing the nature of poverty was examined and critically evaluated, with the researcher deciding that relative poverty is the most appropriate definition for the present study. Then, the literature review critically evaluated orthodox and heterodox economic theories. These theories attempted to explain how specific individuals/groups of people find themselves in economic hardship. In this case, the researcher adopted the social capital theory for the present study. Subsequently, there was an examination of the literature surrounding the impact of poverty on children's outcomes. There was a particular focus on the persistence of poverty, poverty in early childhood, as well as methodological differences between studies moderate the academic, behavioural, and health outcomes of impoverished children. Finally, the literature review focused on the school-level factors associated with childhood poverty.

Chapter 3
Using preventive interventions to reduce disruptive behaviour in the classroom

3.1 Introduction to the chapter

First, the concept of disruptive behaviour (DB) is defined. Then there is an examination of developmental trajectories associated with DB, and a review of the contribution of differential risk factors such as gender and poverty. Second, the literature review introduces preventive interventions as a means to improve outcomes at varying risk levels for the disruptive developmental trajectory. The researcher pays particular attention to intervention development, as well as an exploration of the literature on the effectiveness of universal preventive interventions. Third, the GBG, the intervention utilised in the present study, is described in detail. Fourth, a review of the evidence on the impact of the GBG is presented, with an exploration of differentiated intervention effects for the risk-groups as mentioned above. The chapter concludes by applying inferences taken from the literature examined throughout to the present study.

3.2 Disruptive behaviour in the classroom

DB in the classroom has been an identified as an ongoing problem within schools, and a significant concern for parents and teachers (Kaplan, Gheen & Midgley, 2002; Wheldall & Merrett, 1988; Stephenson, Linfoot & Martin, 2000). With mounting pressures for teachers to cover the requirements of the curriculum and maximise the academic progression of children, this threatens valuable teaching time (Ofsted, 2014). However, policy and clinical research are also interested in the DB of children in early and middle childhood (Tremblay, 2010). The interest in DB amongst different stakeholders is because a large body of research has identified that children who display DB within this time frame are more likely to display conduct problems later on in life (for a review, see Tremblay, 2010). Although this appears to be a small subset of the population, in later life individuals with conduct problems are at further risk of mental health problems, committing violent crimes, and engaging in substance use (Shaw & Shelleby, 2014). The identification of precursor behaviours related to disruption in the classroom could alter the developmental trajectories for children at risk of early-onset conduct problems (Tremblay, 2010; Tremblay et al., 2013). Early detection of DB would not only lead to more positive outcomes for the individuals themselves, but also for society.

3.2.1 Defining disruptive behaviour

Researchers often conflate DB with terminology such as "aggression", "externalising problems" and "conduct problems". Within the literature, these constructs are often used

interchangeably without consideration of the consequences this places on the latter stages of the research process, as well as the interpretation and application of findings. Although these terms share overlapping characteristics in the form of shared behaviours, researchers should not consider these terms as synonyms. In this section, DB is defined and differentiated in two respects. First, it is placed on a continuum alongside other problem behaviours in order to resolve the conflation of terms such as "conduct problems". Then, constructs that comprise DB are justified, with this section presenting an outline of their role in moderating the progression along the problem behaviour continuum.

The position of disruptive behaviour on a continuum of externalising problems

DB has fallen victim to the fallacy of discarded differentia, despite appeals for a more objective term to be formulated (Wheldall & Merrett, 1988). The term "disruption" has two definitions listed in the Oxford Dictionary of English. The first is concerned with interrupting an event, while the second definition refers to destruction (Stevenson, 2010). It would seem that both definitions represent polar ends of a spectrum of behaviours that range from mild inconvenience to causing damage, and is a likely contributor to conflation with terms such as "conduct problems" and "externalising problems". For instance, Kaplan, Gheen and Migley (2002) suggest that DB comprises of low-level behaviours such as teasing peers, talking out of turn, or wandering around the classroom, as well as more extreme behaviours such as vandalism or assault. It appears that authors either fail to state which of the two forms of disruption they are referring to, or are settling for a definition that is too broad for empirical research. These two issues indicate that the research community must further refine the term DB in order to clarify the present understanding of problem behaviours to inform future research.

The researcher proposes that externalising problems fall on a continuum. Externalising problems can be defined as behaviours directed outside of the self, in response to an individual's lack of emotional regulation (Krueger, Markon, Patrick & Iacono, 2005). On the lower end of the continuum would sit DB, which represents behaviours that cause an interruption. Meanwhile, the more extreme behaviours would be considered as the term 'conduct problems'. This proposed distinction is not an arbitrary one, as teachers report that "low-level behaviours" are of particular concern in the classroom, and is likely due to the frequency children often display these behaviours as well as interrupting the learning that is taking place in the class (Wheldall & Merrett, 1988). Such behaviours can fall into three sub-categories: poor attention, infringement of classroom rules, and off-task behaviour (Fields, 1986). Therefore, for the present study, DB is defined as behaviours that cause interruptions to one's own, or other's learning in the classroom.

Constructs associated with disruptive behaviour and their role on the continuum

The research community frequently utilises two constructs within the existing DB literature. The most popular is aggression, this construct comprises of behaviours that have the

intention to harm an individual who does not want to receive it (Baron & Richardson, 1994). Aggression is a divisive construct, which has lead to disagreement over its composition (Card, Stucky, Sawalani & Little, 2008). Much of the literature refers to physical aggression, which involves behaviours that inflict bodily harm to another person. However, more recently, evidence suggests that there are other forms of aggression that rely on more covert means to hurt others; this is known as relational aggression (Bjorkvisit, 2017). Relational aggression is often called social or indirect aggression, and the intention to cause harm is achieved by harming an individual's social status (Card et al., 2008). Behaviours that are used to achieve this involve either manipulating and damaging relationships that are important to the individual, or excluding the individual (Card et al., 2008). However, there are also many instances where children perform disruptive acts without any intention of causing harm.

Such behaviours fall under the construct of impulsivity; this is a range of behaviours that suggest little consideration for consequences and are often inappropriate to the individual's situational context (VandenBos, 2007). Much like aggression, there is a lack of consensus on the different forms of impulsivity (Gvion & Apter, 2011). However, researchers agree that inattention and a lack of self-regulation are subsets of impulsivity (Evenden, 1999). Inattention is when an individual has a low ability to sustain their attention on a particular task, while a lack of self-regulation refers to an individual's inability to inhibit emotions, thoughts or actions in situations where their display would seem unsuitable (Martin et al., 1994; McClelland & Cameron, 2012). The researcher also argues that aggressive and impulsive behaviours fit on a continuum. Aggressive behaviours range from pushing and teasing to the deliberate destruction of property and violence, while impulsive behaviours range from fidgeting and interrupting conversations to extreme mood swings and substance abuse (Tremblay 2010; Evenden, 1999). The more extreme examples are associated with conduct problems and require higher levels of aggression and impulsivity to enact such behaviours. Therefore the present study will only include lower-level aggressive and impulsive behaviours within the definition of DB to reflect the inconvenience these actions cause to teachers and other children in the classroom (Wheldall & Merrett, 1988).

3.2.2 Developmental trajectories and disruptive behaviour

In this section, the literature review will examine the developmental trajectories of children with different levels of DB. As already established, the extant literature conflates it with several terms. As definitions of DB are not consistent, this makes inferences about children's disruption levels throughout their development challenging to establish, as findings may not be comparable. However, with a clear framework regarding the composition of DB for the present study set out in the previous subsection, existing literature relating to low-level aggression and impulsivity can be utilised to explore normative and high displays of the two DB constructs throughout a child's development.

Aggression

The developmental trajectories of children's aggression levels are dependent on how it measured, as the literature suggests different trends (Côté, Vaillancourt, Barker, Nagin & Tremblay, 2007). Studies measuring physical aggression suggest there is a spike in the frequency of aggressive behaviours ranges from 2-4 years (NICHD Early Child Care Research Network, 2004; Nagin & Tremblay, 1999). As infants lack the motor skills and physical strength required to enact the extreme behaviours associated with conduct problems, it is likely that this age range would be regarded as highly disruptive (Tremblay, 2010). For the majority of children, physical aggression levels decrease, particularly after starting school (NICHD Early Child Care Research Network, 2004; Nagin & Tremblay, 1999; Brame, Nagin & Tremblay, 2001). However, studies consistently find a small percentage of children that continue to display high frequencies of aggressive behaviour (NICHD Early Child Care Research Network, 2004).

For instance, Brame and colleagues (2001) identified 20% per cent of their sample (*N*= 926) as displaying high levels of aggression at six years old, but at age 15, this percentage had fallen to 5%. However, the authors reported that there was little evidence of stability in physical aggression levels, except for children identified as displaying high levels of aggression throughout the longitudinal study (3%). Although representing a small proportion of the population, with increases in strength, cognitive, and motor skills, highly aggressive adolescents are more likely to engage in behaviours associated with conduct problems. Figure 3.1 depicts a steady increase in physical fighting and violence as children move through adolescence. The graph indicates there is an association between the frequency of physically aggressive behaviours and the severity of the behaviour. However, intra-individual studies focusing on the increase in the severity of physical aggression concerning age group norms from early childhood onwards would be required to establish this (Tremblay, 2010).

Figure 3.1
Cumulative onset curves for minor aggression, physical fighting and violence in the oldest sample of the Pittsburgh Youth Study (from Loeber & Stouthamer-Loeber, 1998).
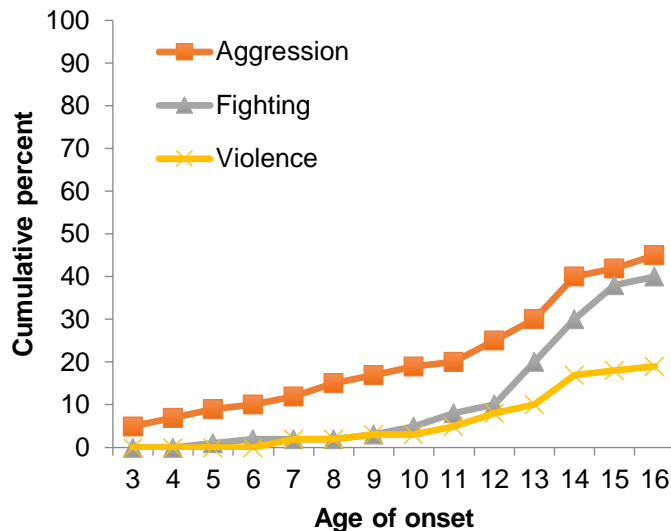
However, developmental trajectories for relational aggression contrast with physical aggression; this suggests they are two separate constructs. Research indicates that relational aggression increases from the onset of school while physical aggression levels decrease, with higher incidences of the former form of aggression established throughout the primary school/elementary school years (Björkqvist & Niemela, 1992). Although fewer studies are measuring normative and atypical levels of relational aggression, as the broader research community slowly accepts this form of aggression, this should soon change. Vaillancourt, Miller, Fagbemi, Côté and Tremblay (2007) found two distinct trajectories in the sample of children (*N*= 1401) followed from age four until they were ten years old. The first cluster identified comprised of 35% of the sample and increased their use of relational aggression with age, while the remaining sample continued to display their relational aggression at a lower frequency. Björkqvist, Lagerspetz & Kaukiainen (1992) argue that in normative development, there is an association between physical aggression and early developmental stages in order to fulfil their needs, as their verbal expression is more limited compared to later developmental stages.

Thus, as children mature cognitively, their use of physical aggression reduces and their use of relational aggression increases, as it requires more complex cognitive skills to achieve harm (Tremblay 2010; Björkqvist, 2017). However, there are very few studies that examine the development of both forms of aggression, particularly at the intra-individual level (Côté et al., 2007). Côté and colleagues (2007) identified eight aggression profiles from measuring both forms of the construct. As demonstrated by table 3.1 there are very few children who displayed high levels of one form of aggression and low forms of the other. Interestingly, the majority of children decreased their displays of physical aggression, but their relational aggression levels remained low. These findings question the model proposed by Björkqvist as the predicted transfer was not evident. The evidence suggests that highly aggressive children were more likely to display high levels of relational aggression at later time points.

Although more empirical evidence would be required to investigate intra-individual aggression levels, present findings indicate that there are a small proportion of children whose aggressive displays increase with age for both types of aggression.

Table 3.1
The proportion of children in developmental profiles for both physical and relational aggression. Adapted from Côté and colleagues (2007).

Group	Proportion	
	%	*N*
Low PA-low IA[1]	5.0	75
Low PA-rising IA	0.4	7
Low desisting PA-low IA	32.4	409
Low desisting PA-rising IA	3.8	53
Moderate desisting PA-low IA	29.7	324
Moderate desisting PA-rising IA	14.2	16
High PA- Low IA	1.0	10
High PA- High IA	13.5	139

[1]IA= indirect aggression and is a synonym for relational aggression.

Impulsivity

Similarly, research indicates that impulsive behaviours are related to inattention and lack of self-regulation peak during early childhood and decrease with age. Eisenberg, Spinard and Eggum (2010) argue that in early childhood individuals have low levels of effortful control; the executive abilities concerned with: inhibiting inappropriate behaviour; shifting and focusing attention as needed, and performing actions individuals usually avoid. Although the majority of evidence measures impulsivity as a general construct, there does appear to be emerging evidence that there is variability in trajectories concerning inattention and self-regulation (Romer et al., 2011; Larsson, Dilshad, Lichtenstein and Barker, 2011; Lee et al., 2017). For instance, Galéra and colleagues (2011) followed infants aged five months until eight years old (*N*=2057). They reported that hyperactivity[5] decreased slightly with age, but inattention symptoms continued to increase. For both subordinate categories of impulsivity, three trajectories were established: high (hyperactivity: 16.1%; inattention: 13.0%),

[5] As described in section 3.2.1, hyperactivity would fall under the self-regulation sub-category of impulsivity.

moderate (hyperactivity: 52.7%; inattention: 58.2%), and low (hyperactivity: 31.2%; inattention: 28.8%). Despite the lack of studies investigating the developmental trajectories of impulsivity compared to aggressive trajectories, Best and Miller (2010) concluded in their review of executive functioning in childhood and adolescence that in normative populations foundational components such as inhibiting and shifting increase significantly with age. The above conclusion suggests that children will generally become less disruptive as they mature and develop greater executive skills.

However, research has identified a similar trend for self-regulation and inattention as the two forms of aggression. For example, a small percentage of individuals remain highly impulsive at adolescence, and these individuals are likely to display both a lack of self-regulation and inattention (Romer et al., 2011; Larsson, Dilshad, Lichtenstein and Barker, 2011; Lee et al., 2017). Larsson and colleagues (2011) followed twin pairs (N= 2900) from age eight until early adulthood. The authors found two trajectories for hyperactivity[2] (low: 91%, N= 2186; high: 9%, N= 219) and inattention (low: 86%, N= 2067; high: 14%, N= 338). The adolescents who were identified on both high trajectories (6%, N= 156) were more likely to display externalising and internalising problems at ages 19-20. Although further intra-individual research is required to attempt to replicate these findings, this is indicative that high impulsivity trajectories remain stable after adolescence. Therefore there is a need for intervention during the earlier years of development when impulsivity trajectories are malleable.

3.2.3 Risk factors associated with disruptive behaviour

As established, small proportions of the child population will display high levels of aggression and impulsivity throughout their development and will contribute to persistent DB in the classroom. However, research has identified contexts and characteristics that put particular groups of children at higher risk of displaying persistent DB (Tremblay, 2010; Shaw & Shelleby, 2014). While many risk factors are associated with a higher than average frequency of DB in childhood, the two with the most significant proportions of children deemed to be at risk are gender (i.e. being male) and experiencing poverty. Before the researcher presents the evidence for the above risk factors, a short summary of other risk factors for disruptive behaviour is presented.

Prenatal, perinatal and infancy risk factors associated with disruptive behaviour

As discussed at the start of this chapter, researchers have been inconsistent in their application of the definition of DB (see section 3.2.1). The majority of researchers have used diagnostic labels such as externalising behaviour disorders when determining risk factors within this area rather than low-level behaviours which constitute as part of the definition provided by the present study (for examples see 2010). A primary focus of such studies have looked at critical stages of development, particularly the pre-natal, perinatal and infancy stages, when determining risk factors for DB risk factors (Latimer et al., 2011). Pre-

natal risk factors associated with the DB include maternal cigarette and alcohol use during pregnancy (For further detail see a comprehensive meta-analysis by Ruisch et al., 2017). Meanwhile, Latimer and colleagues (2011) reported that low birthweight, adoption and family separation, as well as harsh parenting styles were both perinatal and infancy stage risk factors associated with children developing externalising behaviour disorders in later childhood. Some of these risk factors, particularly parenting styles and maternal substance use are discussed at various points of the literature review as part of risk factors associated with poverty (see the appropriate section below). Other risk factors such as low birth weight also have strong evidence bases that indicate poverty has a large part to play in this relationship with the developmental trajectory associated with DB (see McLoyd, 1998; Duncan et al., 1994). With so many risk factors interlinked with each other, the present study intends to focus on two broad contributors to a DB developmental trajectory: Gender (i.e. being male) and experiencing poverty.

Gender

A substantial amount of empirical evidence indicates that males are more likely to be disruptive throughout development compared to females (see Beaman, Wheldall & Kemp, 2007 for a review). The majority of evidence exploring gender differences in aggression and impulsivity levels consistently report that males display higher levels of these traits (see the following for a review: Weinstein & Dannon, 2015; Archer, 2004). As with many gender differences, research has sought biological explanations of aggression. For instance, research often cites the sex hormone testosterone as a critical moderator (Bjorkqvist, 2017). The basis for the popularisation of the sex hormone being responsible for aggression disparities is due to the assumption that males in the majority of animal species are more aggressive than females (Bjorkqvist, 2017). However, researchers have also identified that the role of socialisation and stereotyped expectations of males could explain the gender gap in disruptive behaviour (Underwood, 2003). Males and females are raised to conform to a gender role based on their sex (Rivers & Barnett, 2011). For males, there are explicit and implicit expectations from family and society to behave and express attitudes that are masculine (DiPietro, 1981). Aggression and impulsivity are often encouraged in males as it suggests strength and competitiveness, both stereotyped as masculine traits; whereas females are encouraged to be sensitive and value maintaining relationships as they mature (Osterman et al., 1998). Therefore gendered parenting and teaching practices may be responsible for maintaining or further escalating disruptive behaviours in the classroom.

One could argue that these gendered expectations have directly influenced the research agendas of those interested in disruptive behaviour. As discussed earlier, aggression was predominately conceptualised as a set of physical behaviours, whereas this is now considered to be a dated viewpoint (see section 3.2.2). The focus on physical aggression has meant the majority of evidence of collected has suggested males are more disruptive (Bjorkquist, 1994; Tremblay 2010; Underwood, Galen & Paquette, 2001). While evidence

indicates that males are more likely to display direct forms of aggression compared to females, for the majority of cultures such displays often occur in exceptional circumstances (Tremblay, 2010; Bjorkquist, 1994). The disparity of aggressive incidents is because there are high levels of risk involved, whereby the aggressor could harm himself during aggressive acts towards others (Tremblay, 2010). As females are, on average, physically weaker than males, they will engage in relational aggression in order to attack others without putting themselves in physical danger (Tremblay, 2010). Although such behaviours have been known to cause problems within schools and classrooms, the emerging research into the impact of females' tendencies to use indirect aggression appears to be primarily qualitative (e.g. Owens, Shute & Slee, 2000).

Although evidence indicates that females are more likely to engage in relational aggression with the increase of age, Table 3.1 demonstrates that individuals who exhibit high levels of physical aggression are more likely to also display high levels of relational aggression (Côté et al., 2007). However, Côté and colleagues reported that there was no significant difference between the proportion of males (12.53%, *N*= 74) and females (10.99%, *N*= 65) in high physical-high relational aggression profile (Odds Ratio= 1.16). Despite this, males are still more likely to be reported by their teachers as being highly disruptive in the classroom compared to females (Wheldall & Merrett, 1998, Kaplan et al., 2002; Stephenson et al., 2000). This gender disparity supports the proposed role gendered socialisation plays in perceptions of DB, as males are anticipated to display the masculine behaviours teachers identify as being disruptive, e.g. challenging the teacher's authority, distracting others from being on-task, or placing classmates in harm (Wheldall & Merrett, 1988).

Poverty

Research consistently reports an association between childhood poverty and disruptive behaviour (Shaw & Shelleby, 2014; McLoyd, 1998). As examined in Chapter 1 (see section 2.4), deprivation is associated with a myriad of proximal and distal negative outcomes across the different spheres of development. With increasing attention to child poverty across various facets of society, including government policy, researchers have begun to test whether the association between poverty and disruptive behaviour is causal. Cooper & Stewart (2017) conducted a systematic review of studies that consisted of RCTs and QEDs to determine a causal link between poverty and a range of developmental outcomes, including behaviour. The authors reported that the majority of studies identified (12 of 16) found that an increase in household income did cause a positive effect on behavioural development. Interestingly, the additional income caused more significant improvements in households at the lowest end of the income distribution. However, there is uncertainty surrounding whether the increase in income directly improves behaviour or if an indirect path is responsible.

Two frameworks that represent both sides of the debate have been developed to explain poverty's role in mediating behavioural trajectories. The family stress model of economic hardship (FSMEH: Elder, 1974) argues that economic strain indirectly influences childhood development through the increased levels of stress placed on households as they struggle to maintain an adequate standard of living (see Figure 3.2). As a result, parents that experience the cumulative effect of the stressor develop poor psychological functioning, which can lead to harmful parenting practices. Shaw and Shelleby (2014) adapted the original model to reflect more recent findings which acknowledged the reciprocal nature between parenting and parent psychological resources[6] (Figure 3.2). Meanwhile, the investment model (Mayer, 1997) argues that higher income families can invest more resources in their children, thus increasing their wellbeing. Families with lower incomes are less able to invest such resources, such as high-quality childcare, safe homes and neighbourhoods, which are associated with child positive wellbeing and behavioural outcomes (McLoyd, 1998; Duncan, Brooks-Gunn & Klebanov, 1994). Thus, deprived children are less likely to develop positive social skills and coping mechanisms that are thought to be protective factors from conduct problems in later life (Shaw & Shelleby, 2014).

Figure 3.2
The family stress model applied to early-starting conduct problems. Taken from Shaw and Shelleby (2014).
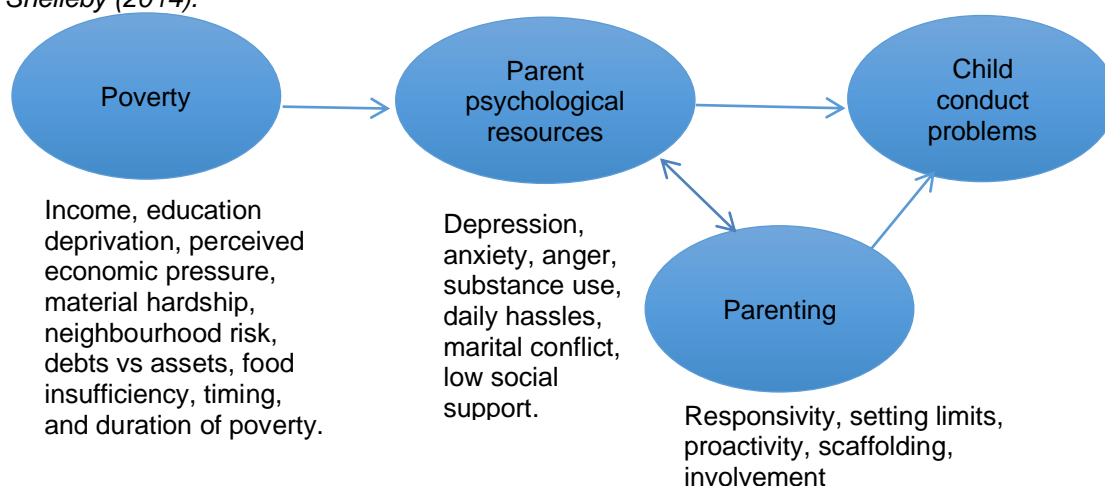
[6], I.e. dealing with misbehaviour can be considered to be a stressor in its own right.

However, it appears the investment perspective has a stronger basis for explaining cognitive, academic, and career outcomes for deprived children as opposed to behavioural outcomes (Shaw & Shelleby, 2014). For instance, Yeung, Linver and Brooks-Gunn (2002) found that investment in stimulating resources mediated the relationship between household income and childhood achievement, while parent psychological functioning and parenting practices mediated childhood behavioural outcomes. Meanwhile, there is a larger body of evidence that supports the FSMEH model with parental psychological functioning and parenting practices mediating poor behavioural outcomes for deprived children (see Shaw & Shelleby, 2014 for an overview). For instance, Neppl, Senia & Donnellan (2016) conducted a longitudinal study and reported that economic hardship was associated with parental distress. The researchers also reported an association between economic hardship, punitive rearing practices and child problem behaviour. Thus, the FSMEH model provides additional understanding to the complexities surrounding poverty and its effects on children discussed in chapter 1 (section 2.4), particularly in conjunction with SCT theory (section 2.3.2) as part of the framework for the present study. Not only is the quantity and quality of social ties in part causing economic hardship within households, but poverty is also damaging the condition of existing relationships within the household. Therefore, poverty has an impact on the child's development and affecting their ability to behave in a way that allows them to improve their social capital in the future.

3.3 Preventive Interventions

Interventions are "purposively implemented change strategies" (Fraser & Galinsky, 2010, p. 459), and can take the format of formalised programmes developed to tackle a specific issue or an unofficial strategy that has been undertaken to achieve the desired outcome. Preventive interventions aim to inhibit the progression of maladaptive behaviours by encouraging children to develop strategies that promote more positive outcomes (Stallard & Buck, 2013; Coie et al., 1993). They are thought to be superior to treatment interventions, which are reactive as these interventions are implemented once health and social problems occur, as they are cheaper to implement and thought to alleviate the burden on society in the long term by reducing service use for a range of public services (Eisenberg & Neighbours, 2009). Schools have increasingly become a site for the implementation of preventive interventions as they play a critical role in a child's development. They are often the primary setting in which problems or difficulties are identified as well as remediated (Greenberg, 2010). In this section of the literature review, the forms of preventive

interventions, as well as their development and dissemination, are explored. The section will provide a foundation for a critical analysis of the impact of preventive interventions in reducing DB, particularly among children experiencing poverty.

3.3.1 Types of preventive interventions

Preventive interventions come under a taxonomy based on their target population; Gordon (1983) proposed three levels: universal, targeted, and indicated interventions. However, with more interventions being designed and implemented, their classification is becoming increasingly more complex (Foxcroft, 2014). Foxcroft (2014) developed a matrix that divides the interventions based on their form, as well as their function (see table 3.2). In this section of the review, attention is paid to the developmental strand of the preventive intervention taxonomy as such programmes aim to prevent maladaptive developmental trajectories described previously in the chapter (see section 3.2.2). Indicated (IPI) and selective preventive interventions (SPI) will be discussed before universal preventive interventions (UPI) as although they are relevant for aiding the reader's understanding, the programme utilised in the present study is a UPI.

Table 3.2
Classifying prevention: form and function in a prevention matrix adapted from Foxcroft (2014).

	Universal interventions	Selective interventions	Indicated interventions
Environmental interventions: *relating to public policy*	Making undesirable behaviours illegal; taxing unhealthy commodities; gun controls prohibiting firearms without permission from relevant authorities.	Reducing alcohol retail outlet density in high-risk neighbourhoods; improved street lighting and CCTV in town centres; age restrictions placed on harmful products.	Legislation preventing violent individuals purchasing firearms; incarcerating individuals who will harm the broader population.
Developmental interventions: *shape the socialisation of individuals as they grow*	Parenting programmes allowing for modelling and practice of positive interactions; classroom behaviour management programmes promoting prosocial	Home visits for at-risk new mothers; parenting programmes for at-risk families in a known area of deprivation.	Multi-systemic therapy for individuals with severe antisocial behaviours; counselling for individuals displaying a particular personality characteristic.

Indicated	behaviour; social skills programmes that teach young people skills to cope with social influences.		
Informational interventions: *improving knowledge and awareness*	Mass media campaigns raising awareness of risk behaviour; school-based awareness curricula providing information about substance use.	Informational interventions targeting a subgroup with a specific characteristic; targeting school-based knowledge curricula to a subgroup with a specific characteristic.	Normative feedback interventions for individuals displaying a behavioural risk factor; information messaging for high sensation seekers.

preventive interventions

IPIs [7] are designed to target individuals within the population who show symptoms that suggest they are at high risk of developing a clinical disorder (Gullotta, 2015). Although IPIs are often difficult to distinguish from treatment, programs that fall into this category prevent significant difficulties experienced by individuals considered "sub-clinical" to advance into symptomology requiring therapeutic or medical services (Durlak & Wells, 1998). Screening is essential for children within a school or class to identify candidates that are in need of the intervention (Durlak & Wells, 1998). At present, this public health model of screening is rare, with the majority of schools and other public services utilising the "refer-test-place model" (Cash & Nealis, 2004). In this model, individuals are referred to a specialist, such as an educational psychologist, who provides access to interventions to a small number of individuals who are at the highest level of risk (Kleiver & Cash, 2005). The model is inefficient and costly and leaves large proportions of children and young adults at risk to a variety of mental health and wellbeing difficulties (Humphrey & Wigelsworth, 2016; Dowdy, Ritchey & Kamphaus, 2010).

Selective preventive interventions

SPIs target subgroups of the population that have been identified by substantive research as being at higher risk than the average (Gordon, 1983). SPIs focus on developing the knowledge, and skills individuals within the targeted group may be lacking compared to the rest of the population, usually due to their life circumstances (Weissberg, Kumpfer & Seligman, 2003). These interventions are thought to be most successful if the at-risk subgroups are easily reachable and identifiable (Gullotta, 2015). For SPIs that focus on development, an educational setting is ideal as at-risk children can be easily reached and identified through the vast amounts of information school collect from their pupils (e.g. FSM, EAL, and attainment levels). In theory, children can be identified quickly and access a relevant SPI if available.

---

[7] This is often referred to as "secondary prevention" in earlier literature. See Gordon (1983) for an explanation for the change in terms.

Universal preventive interventions

UPIs target the whole population, which makes them less stigmatising than IPIs and SPIs (Greenberg & Abenavoli, 2017). Such programmes within school settings provide the opportunity to develop general 'soft skills' such as self-regulation and social skills that have been identified to prevent multiple deleterious outcomes that share common risk factors (Greenberg, 2010). Although whole population programmes are unlikely to increase positive outcomes by a large amount at the individual level, the cumulative effect will benefit the whole population (Rose, 1992). Therefore, the majority of individuals must change their behaviours through the participation in a UPI in order for a smaller section of the population to benefit (Greenberg & Abenavoli, 2017). A popular clinical example of this is the vaccination programme. Although the majority of individuals are unlikely to develop the target disease, higher vaccination uptake decreases the transmission of the pathogen, which in turn further protects individuals who are at a higher risk of having compromised immunity (NHS, 2016).

The academic community sometimes call non-medical UPIs behavioural vaccines. Embry (2002) referred to the GBG, the intervention utilised for the present study, as a behavioural vaccine. According to Embry, such programmes expose the population to a weakened version of a behavioural risk, much like a pathogen in a medical vaccine. He also argues that behavioural vaccines may perform in a similar way to vaccines utilised in computer science. In this case, interventions teach children a set of protective behaviours that guards against any behavioural "attacks" in the future. With high levels of compliance to a set of behaviour changes, UPIs such as the GBG could prevent the onset of mental health disorders, engagement in substance use, criminal behaviour, as well as other behaviours that could harm the afflicted individual but also wider society.

3.3.2 Designing and evaluating preventive interventions

Before exploring whether UPIs are an effective way of reducing disruptive behaviour in children, particularly those experiencing poverty, it is essential to understand the processes used to develop interventions. By understanding how developers design the inner workings of interventions in order to promote behaviour change, the literature review can scrutinise the evidence base presented in later sections, as well as the findings of the present study. Fraser and Galinsky (2010) proposed a 5-step model in which an intervention is developed, refined and tested. Table 3.2 summarises the model.

Table 3.2

Steps in Intervention research, adapted from Fraser & Galinsky (2010).

Stage	Objectives
Development of problem and program theories	Develop problem theory of risk, promotive and protective factors. Develop a program theory of malleable mediators. Identify intervention level, setting, and agent(s). Develop a theory of change and logic model.

Specify the programme structure and processes	Develop the first draft and submit for expert review. Specify essential program elements and fidelity criteria. Pilot program and measures (i.e. outcome and fidelity measures). Expand content to address training and implementation.
Refine and confirm in efficacy tests	Maintain high control and test intervention components. Estimate effect sizes and test for moderation and mediation. Develop rules for adaptation based on moderation and mediation.
Test effectiveness in practice settings	Test intervention under scale conditions in multiple sites. Estimate effects under the intention to treat conditions. Estimate effects on efficacy subsets.
Disseminate programme findings and materials	Publish findings. Publish program materials. Develop training materials and certification.

Development of problem and program theories

First, the literature should be consulted to identify a suitable problem theory to explain the identified issue (Fraser & Galinsky, 2010). A suitable problem theory identifies the risk, promotive and protective factors associated with the stated issue that has been established by substantive empirical evidence (MacKinnon, 2011). MacKinnon (2008) argues two components drive the theoretical underpinnings developed in this stage: the first is the action theory, which specifies and determines how the programme will affect malleable mediators. Meanwhile, the conceptual theory emphasises how mediators relate to outcome variables. Explicit formations of these theories will then allow developers to define the critical features of the intervention (Gottfredson et al., 2015). Once decided, the developers will produce precise action strategies in the form of a logic model or theory of change. A logic model describes the processes for creating change by synthesising the main program elements into a succinct diagram demonstrating how the program is theorised to work (Forman, 2015). A theory of change can be considered to be a more nuanced form of a logic model used for complex interventions that intend to achieve longer-term outcomes (Weiss, 1995). However, the key difference between the two methods of theorising interventions is that theories of change are not static. One of the basic principles of the theory of change is that causality is more complicated than that assumed in logic modelling (Humphrey et al., 2016). Nevertheless, further refinement of action strategies should lead to the clarification of when expected outcomes should be observed based on the understanding of the "developmental epidemiology of the targeted behaviour" (Gottfredson et al., 2015 p.899).

Specify program structure and processes

In this stage, the bulk of intervention design is undertaken and guided by the logic model together with the programme theory in order to produce a manual and essential content

(Fraser & Galinsky, 2010). The manual provides an overview of the intervention's structure, while the essential content addresses the core risks mechanisms the intervention is underpinned by (Humphrey et al., 2016). Both the manual and essential content inform the fidelity criteria (explored further in section 4.2.2). Stakeholders such as potential implementers and experts in the field should provide feedback on the content and materials in order to improve the likelihood of the intervention succeeding during the testing phases (Fraser & Galinksy, 2010).

After multiple reviews and revision of the core intervention elements, pilot testing in the form of a feasibility study can take place. Feasibility studies test the intervention under optimal conditions and are concerned with implementation as opposed to intended outcomes (Onken, Blaine & Battjes, 1997). Such studies are usually conducted on a small scale to assess whether the implementation is feasible and whether the intervention has social validity (Humphrey et al., 2016). Social validity is concerned with the value attributed to a programme by its implementers or recipients (Wolf, 1978). The construct has been identified as a key factor that influences implementation levels in further stages of intervention research. The intervention theories driving the logic model or theory of change developed in the first stage should also be assessed during the feasibility study. Connell and Klem (2000) suggest four criteria to determine intervention theory quality, they are:

- Plausible: Stakeholders agree with the logic behind the theory.
- Doable: The required resources are available to implement the intervention.
- Testable: There are credible methods to test the theory.
- Meaningful: Stakeholders value the outcomes, which are perceived to be worth the effort required by the intervention to achieve them.

Evidence gathered during the pilot stage may require developers to revise the programme structure, materials and theory in order for the intervention to show promise in producing the expected change in a real-world setting.

Refinement and confirmation in efficacy tests

After stakeholders approve the modifications from the feasibility studies, researchers will conduct efficacy studies. The primary purpose of an efficacy study is to determine whether the intervention produces the desired change in optimal conditions (Singal, Higgins & Waljee, 2014). Efficacy trials typically take the form of a randomised control trial[8] (RCT) where there is a control condition to compare to in order to determine if the intervention is producing a causal effect on the desired outcome (Torgerson & Torgerson, 2013). The usual practice of all implementers should be assessed before randomisation together with post-

---

[8] See section 5.4 for further detail about randomisation practices generally, and those utilised in the present study.

test in order to determine programme differentiation[9] and whether the implementers in the control condition attempted to compensate for the lack of intervention by changing their usual practices (Humphrey et al., 2016). Meanwhile, adequately powered efficacy studies can determine effect size estimates for the intervention, as well as moderation, mediation, and subgroup analyses (Fraser & Galinsky, 2010). Subgroup analyses allow insight into how the intervention works for different groups of individuals.

Testing effectiveness in practice settings

In the next stage, effectiveness studies, which also take the form of an RCT, are conducted[10]. The purpose of an effectiveness study is to estimate intervention effects when a program is implemented in real-world conditions once developer support is withdrawn (Singal, Higgins & Waljee, 2014). The implementation of interventions increases in variability during this stage, which typically leads to diminished intervention effects observed previously in efficacy trials (Wigelsworth et al., 2016). One of way assessing this is by documenting the contextual factors that moderate implementation, which is likely to vary across the multiple sites that have been randomised to the intervention condition (Humphrey et al., 2016; Fraser & Galinsky, 2010).

Disseminating program findings and materials

After multiple cycles through the previous stages, the intervention is ready for dissemination. However, the reach of evidence-based interventions (EBI) needs to be improved as many schools are either not implementing preventive interventions or choose to implement interventions with have not undergone rigorous testing (Spoth et al., 2013). This translation from research to practice means public health benefits often take approximately 20 years to manifest (Balas & Boren, 2000). Translational research is an area of prevention science which investigates the mechanisms which allow interventions to integrate with the public domain (Spoth et al., 2013). This area of research faces several barriers that hinder the progression of the field (Glasgow, Litchenstein & Marcus, 2003). One key issue is the limited funding received, as the treatment for already existing issues often receive investment over preventing them in the first place (Catalano et al., 2012). The lack of funding is likely due to the priorities of policymakers who have not considered the wider benefits to society by investing in strengthening and building infrastructure to facilitate dissemination (Woolf, 2008).

Although the model developed by Fraser and Galinsky (2010) provides useful insight into the development and testing of interventions, the extent to which the model is reflective of current practice within prevention science is questionable. For instance, the efficacy and effectiveness trial stages are not always discrete phases; hybrids of the two trial forms are

---

[9] See section 4.2.2 for further detail regarding programme differentiation.
[10] See section 5.4.1 for a discussion regarding the use of efficacy and effectiveness trials for intervention testing, as well as the implications these have on the present study.

increasingly adopted over running two separate RCTs (see section 5.4.1). However, alternative frameworks such as the interactive systems framework (Wandersman et al., 2008) or the Institute of Medicine model (Institute of Medicine; IOM: 1994) do not mirror the present intervention development processes either. However, should models reflect current practice? Gottfredson and colleagues (2015) argue that proposed models can be developed to "purposefully set higher to encourage growth in the field" (p.893). Therefore, models such as the one proposed by Fraser and Galinsky (2010) should be seen as guidance to address some of the issues surrounding the development, testing and dissemination of preventive interventions. The present study draws from this by contributing to the evidence base for The Good Behaviour Game through conducting subgroup analyses for a key at-risk group: children experiencing poverty.

3.3.3 The impact of preventive interventions in decreasing disruptive behaviour

The present section of the literature review evaluates the extent to which prevention interventions are effective in reducing disruptive behaviour within the classroom. The focus will be on UPIs rather than the other forms of intervention discussed previously (see section 3.3.1) to reflect the intervention utilised in the present study. First, the main effects of UPIs on disruptive behaviour will be detailed and later the differential gains for children experiencing poverty will be explored.

Universal preventive interventions and disruptive behaviour

There appear to be two forms of UPI that have been designed to address disruptive behaviour: social-emotional learning (SEL) interventions and behaviour management interventions (Sabey, Charlton, Pyle, Lignugaris-Kraft & Ross, 2017). Social-emotional (SEL) interventions have been designed to facilitate the mastery of social and emotional skills that lead to positive outcomes such as stronger emotional regulation and improved relationships with others (Greenberg et al., 2003). Improvements in these areas reduce disruption levels in the classrooms, as "soft skills" address the underlying issues that contribute to disruptive behaviour. Meanwhile, behaviour management interventions have been designed to specifically reduce disruptive behaviour using principles such as reinforcement, punishment, and prompting (Sabey et al., 2017). It is difficult to compare and contrast the impact of behaviour management and SEL programmes with one another as there are fewer behaviour management interventions compared to the vast selection of SEL interventions available. However, empirical evidence indicates that the effect sizes between the different forms of programmes are similar (Wilson & Lipsey, 2007). Therefore, this section will review both types of interventions together.

Recent meta-analyses that have reviewed the testing of UPIs within the classroom have concluded that they are an effective method for reducing DB (e.g. Durlak et al., 2011; Skald, 2012; DuBois, 2002; Korpershoek, Harms, de Boer, van Kuijik & Doolaard, 2016). For

instance, Durlak and colleagues (2011) reported that universal UPIs produced small but significant effect sizes in both decreasing disruptive behaviour and increasing prosocial behaviour. However, they also reported that the intervention effects from follow up studies almost halved. Taylor, Oberle, Durlak and Weissberg (2017) also investigated post-intervention follow up effects in their meta-analysis and found that although participants had improved SEL skills and less emotional distress, there were no significant findings for positive social behaviours or conduct problems. Wigelsworth and colleagues (2016) argue that discrepancies between findings are due to a range of factors such as: implementation failure (see section 4.2), a lack of focus on the real-world application of the intervention (Flay et al., 2005), developer involvement during testing (Eisner, 2009), and a lack of cultural transferability (Castro, Barrera & Martinez, 2004). Within their meta-analysis, Wigelsworth and colleagues (2016) found evidence to support this argument, particularly for efficacy studies, developer involvement and cultural transferability. This evidence suggests that although UPIs have the potential to improve behavioural outcomes within the classroom, there are many gaps in knowledge that need to be addressed in order to sustain intervention effects consistently (Wigelsworth et al., 2016).

Reductions in disruptive behaviour for impoverished children

One such gap in knowledge is the differential gains of at-risk subgroups when exposed to interventions. Although UPIs are designed to produce a cumulative effect that benefits a population, recent research has reported mixed findings with reports of both increased gains for vulnerable groups (Deković et al., 2011; Jones, Brown, & Aber, 2011; Rutter, 1985), as well as increased inequalities (Ceci & Papiero, 2005; Frolich & Potvin, 2008). The "Matthew effect" is one argument as to why UPIs are thought to fail at-risk groups (Ceci & Papiero, 2005). This effect is where the initial advantage is amplified to widen pre-existing gaps (Walberg & Tsia, 1983). Cunha and Heckman (2007) formalised this notion into a skills-building economic model of human development, where investments and child endowments interact to produce an individual's stock of accumulated human capital. Essentially, children who possess foundational skills are in a better position to develop more complicated skills in the future (Cunha & Heckman, 2007). The application of the Matthew Effect may explain why some UPIs are more successful at addressing skill deficit because they incrementally build from simpler skills to more complex ones (Bailey, Duncan, Odgers & Yu, 2017). This application may explain the mixed results for vulnerable groups. The present study focuses on children experiencing poverty as a risk group because they represent a large proportion of the UK population. As explored in earlier (see section 2.4), children experiencing poverty are at a higher risk of a range of negative outcomes, including developing disruptive behaviours (see section 3.2 for a synthesis of the literature). With a moral obligation (and policy priority) to improve outcomes for such children, it is important to review whether UPIs are helping or hindering these attempts to improve behavioural outcomes.

As there is a moral obligation for research to benefit the disadvantaged as much as possible throughout the research process, many trials of interventions operate in areas of socio-economic disadvantage (e.g. Jones, Brown & Aber, 2011). For example, The Conduct Problems Prevention Research Group (2010) conducted an RCT to test the effectiveness of the social-emotional learning intervention Fast Track Promoting Alternative Thinking Strategies (Fast Track PATHS) in three high-risk locations across the USA and identified intervention effects for children identified as highly disruptive. By using sizeable samples in deprived areas, it is tempting to conclude that large groups of vulnerable children will benefit from universal preventive interventions. However, there are also instances of experimental studies trialled in highly deprived areas which reported non-significant intervention effects, including established, evidence-based programmes. For example, Little, and colleagues' (2013) RCT conducted in Birmingham to test PATHS reported null results. Other interventions have also failed to establish positive effects on outcomes (Cho, Hallfors, & Sanchez, 2005; Kam, Greenberg, & Walls, 2003; Wigelsworth, Humphrey, & Lendrum, 2013). Although this raises the issue around the variability of intervention effects in deprived areas (See Chapter 3), these studies do not provide an insight into the differential gains children experiencing poverty would make *compared to the rest of the population*.

Comparison of the differential gains between children experiencing poverty and their more affluent peers requires sub-group analyses within adequately powered RCTs which have large enough samples of both groups of interest. As this requires large numbers of participants representing both characteristics, there are very few studies that directly compare and contrast the disruptive behaviour outcomes of both groups. Though with large RCTs becoming more commonplace in the field of education to test UPIs, such subgroup analyses should become more commonplace, tentative findings suggest there may be differential gains for impoverished children (e.g. Second Step, Holsen, Iversen & Smith, 2009). However, the lack of studies in this particular area strengthens the need for the present study to begin addressing these gaps in knowledge.

3.4 The Good Behaviour Game

The Good Behaviour Game (GBG) was originally developed in the USA after Barrish, Saunders and Wolf (1969) observed a teacher playing a classroom behaviour management game and reported the findings to the broader research community. The intervention has since been further developed and tested across the world which has led to multiple variations of the intervention utilised in different cultural contexts (e.g. The Netherlands: Leflot, van Lier, Onghena, Coplin, 2013), or different situational contexts (e.g. the cafeteria, McCurdy, Lannie & Barnabus, 2009). The literature review dedicates a focused section examining the different forms of GBG that have been developed throughout the intervention's history - this can be found in section 4.5.1. Meanwhile, this portion of the literature review focuses on the version of the GBG produced by the American Institutes for

Research (AIR: Ford, Keegan, Poduska, Kellam, & Littman, 2014) as this was the version utilised in the present study.

3.4.1 Key components of the Good Behaviour Game

The GBG is a universal preventative intervention that relies on group membership and the following of four classroom rules to reduce disruptive behaviour in the classroom. Several key components function during a typical game that change over time as the class becomes more adept at following the intervention. These are the team composition, the rules, dosage, and the rewards offered.

Teams

The teacher assigns children to teams consisting of up to seven members. These are balanced by gender, behaviour, and academic ability. Teachers are expected to review and change team composition at least once a term. The rotation of teams is so children learn to work together with all classmates rather than encourage competition. Each team has a team leader, whose primary function is to participate in the post-game celebration (see section 3.4.2). Teachers should appoint new team leaders with every review of teams.

The four rules

Each team must follow the rules (which are reviewed at the start of every game) while completing a task set by the teacher. The four rules are:

1. We will work quietly[11]
2. We will be polite to others
3. We will get out of our seats with permission
4. We will follow directions

These rules are based on universal, western expectations of what "good behaviour" constitutes and are meant to be visible throughout the classroom. Teachers are required to display a large poster at the front of the class, in the view of all pupils. Meanwhile, smaller posters are stuck onto the desks at which the children work. This is so individuals can remind themselves of the rules at any point during the game.

Dosage

The teacher must tell the children how long they will be playing the game for as part of the preamble. The teacher will use a timer which must be visible to the children, so they can be aware throughout the game how long they have left to complete the task, as well as adhering to the rules. When children are first introduced to GBG, the length of time spent playing the game is short, and the frequency of games is restricted to three times per week. The low

---

[11] This rule is a minor misnomer. Children are taught about the importance of working at an *appropriate* noise level for the activity that has been set. Optional "voice levels" can be introduced to the children and range from silence to loud conversation.

dosage allows the class to become familiar with the intervention, as well as demonstrating that the teams can win the game. As the children follow the rules and win the game more frequently, teachers are meant to increase the game length based on the progress of the children and begin playing the game daily. The intervention allows the children to develop the skills in order master classroom appropriate behaviour within a safe space.

Rewards

When first introduced to GBG, each team that wins the game receives immediate gratification with a tangible reward (e.g. stickers). Over time, the delay in gratification should increase, eventually leading to teachers delivering rewards at the end of the school week. The nature of the reward should also become more abstract (e.g. free time) in order to encourage the internalisation of the behaviours and strategies developed through participating in the GBG.

3.4.2 Playing the Good Behaviour Game

A typical GBG session has three distinct parts. First, there is the preamble, where the class revisit the game rules, and the teacher explains the activity they wish the class to complete during the game. Then the GBG is played for an agreed amount of time. At the end of the winning teams celebrate their achievement. There is a fourth component: the probe which is where the teacher assesses the extent to which the "good behaviour" is generalising outside of the intervention.

Preamble

Before the game is played the teacher discusses the meaning of the rules with their pupils, the teacher and children generate examples of behaviours that constitute as following the rules, as well as examples of not following the rules. There is strict guidance within the manual that prohibits children imitating examples of rule breaking, as developers argue this might confuse pupils and inadvertently encourages infractions during the game (Ford et al., 2014). The class review strategies or parameters where rule breaks often occur. For instance, if children drop equipment on the floor, they may not get an infraction for rule three, as this may be an instance where the teacher has permitted at the start of the game. Strategies and parameters are therefore likely to vary both between and within classes, as these are often dependent on a wide range of factors such as the activity set for the game, as well as the needs of the class. At this point, the teacher also explains the activity they wish the children to complete within a determined amount of time. The activity is usually a relevant piece of work to the subject area in which the game is played. The developers designed the intervention to not compete with teaching time, and in theory can be integrated into any teaching opportunity rather than allocating a specific time during the school day to administer the intervention, like many curricula-based interventions.

During the game

The teacher then announces the start of the game and sets the time. The teacher then observes and records rule breaking. The implementation manual prohibits the teacher from interacting with the children. If a child tries to speak to the teacher, the manual advises the teacher to ignore the request and respond by giving an infraction. The teacher's role is to observe and record rule breaking only and is achieved by following a strict protocol called "check, comment, redirect" (CCR):

1. The teacher identifies the team (not the individual) and states which rule was broken.
2. The teacher notes the infraction on the "team assignments" poster that should be visible to the children. The infraction is recorded on a large poster that depicts a table in which each team receives a tally mark in a box corresponding to each rule-break.
3. The teacher should discretely show the child(ren) that caused the infraction which rule was broken by pointing on the individual poster displayed on the desk. It is paramount that the rule-breaker(s) remain anonymised throughout CCR.
4. Finally, the teacher praises the rest of the team and other teams for obeying the rules.

After the game

At the end of the game, each team that received four or fewer infractions is considered to have won the game. The teacher rewards winning teams with a prize. It is important to note that the teams are not competing with each other, so if one team has fewer infractions than another, this does not make them a 'better team'. However, the teacher does *not* draw teams that have not won the game. In conjunction with this, the team leader distributes GBG booklets where children record their progress based on the number of games they have won by placing a sticker or stamp on a monthly chart. The team leader also places a star on the "weekly winners" poster. If a team wins at least once during the week, team members gain access to a weekly prize.

Probes

In order to assess the extent to which the class are generalising their learnt behaviours, teachers are encouraged to conduct a probe once per week. The teacher covertly observes and records infractions during an ordinary lesson. The children are not meant to be aware that the teacher is monitoring them, so there is no preamble, prize giving, or following the CCR protocol.

3.4.3 Theories underpinning the Good Behaviour Game

Reinforcement is a cornerstone concept within GBG and is integrated at multiple levels of the intervention by drawing on both behaviourist and social approaches to reduce disruptive

behaviour and promote adaptive classroom behaviours in the classroom. The GBG utilises three theories in order to promote behaviour change; these are operant conditioning, social learning theory (SLT) and social field theory (SFT).

Operant conditioning

Operant conditioning focuses on promoting desired behaviours through reinforcement and discouraging undesirable behaviours through punishment (Skinner, 1948). Positive reinforcement is where a reward is given to the individual if a desirable behaviour is displayed, while negative reinforcement is strengthening a behavioural response by removing an unpleasant stimulus when the desired behaviour is observed (Aeschleman, Rosen & Williams, 2002). Meanwhile, punishment can involve either removing or presenting a stimulus in order to weaken a behavioural response (Skinner, 1948).

The GBG draws heavily on positive reinforcement at the end of a session. Children receive stamps or stickers in their booklets as well as receiving a reward (both at the end of the game, and the end of the week). The celebration process where the team leader places a star on the wall chart could also be considered a reward. However, the manual specifies that punishment should not be used during the GBG (AIR, 2010). A teacher achieves this by not drawing attention to losing teams or deviating from the CCR protocol teachers are expected to follow when rule-breaking. However, the marking of infractions on the scoreboard could be considered to be a punishment. The tally is a stimulus that is meant to weaken behaviours that are associated with rule breaking. Thus, pupils may interpret reinforcers differently to the developers' intentions.

Social learning theory

Social learning theory (SLT; Bandura 1977) argues that children learn through observational learning. By watching a variety of social contexts, children develop behaviours that are reflective of their surroundings. Children are more likely imitate behaviours if the models (the individual the child observes) are perceived to be similar to the child (Bandura, 1977). The influence of models suggests that demographic characteristics such as gender and race can be influential in determining whether behaviours will become internalised. Social feedback is also a deciding factor as to whether the behaviour will be repeated, with positive feedback encouraging future iterations of the behaviour. Children can learn a new behaviour through either their direct experiences or vicariously

The GBG applies the principles of SLT through the use of teams. By ensuring demographic characteristics such as gender are balanced, children are more likely to identify with some of their teammates. Balancing the team characteristics is meant to increase the chances of encoding positive behaviours, through social reinforcement given from the teacher (via CCR protocol) and feedback from their teammates. There are also opportunities for vicarious reinforcement by watching pupils receive infractions for rule breaking, as well as losing

teams watching winning teams engaging in celebratory behaviour at the end of the game. However, the principles of SLT can also be applied to socially undesirable behaviours too, with children identifying with peers who are perceived to be getting reinforcement from by engaging in maladaptive behaviours in the classroom. Again, this is dependent on whether the children interpret the reinforcement as the developers intended. For instance, children who have severe behavioural difficulties may feel validated if they sabotage the game and prevent the rest of the team from winning[12] (Kellam et al., 1994). Although that specific child may receive a considerable amount of negative attention, just the fact they are receiving any attention at all could be perceived as positive reinforcement.

Social field theory

Social field theory (SFT; Kellam et al., 1975) is composed of two strands that influence the likelihood children will conform to socially desirable behaviour throughout their lifetime. The first is social adaptation status, which is a child's ability to meet the demands of a social situation. Each stage of life consists of multiple social fields, and school is considered to be an integral one for children. The school is considered to be important as it is predictive of future progress in adulthood. Children must meet task demands that are defined by the social setting (e.g. being prosocial or attentive in class) in order to be successful within a social setting. Natural raters, such as the teacher, rate how well individuals respond to task demands. Rating is often an implicit process often achieved through regular classroom behaviour management practices or parenting practices. The second strand is related to the child's psychological well-being. How successful an individual has "adapted" within a social field based on their ability to meet task demands, can be predictive of later adaptation attempts (Cicchetti & Schneider-Rosen, 1984). It is thought that psychological wellbeing is reciprocally related to social adaptation (Kellam et al., 2008). Children who are not as successful at adapting to social task demands are more likely to be at risk of feeling negative emotions such as depression or aggression (Kellam et al., 2008). This leads to further maladaptation to social task demands.

The GBG provides a platform in which children are explicitly taught the social task demands through reinforcing the behaviours required to meet these demands. The intervention achieves this through the rules, and the CCR protocol. In order to prevent negative impacts on psychological well-being when children are not meeting social task demands, teachers avoid using punitive measure or negative language when delivering an infraction is, or when a team loses the game. However, SFT does not provide enough detail relating to how

---

[12]. However, the manual does suggest to place individuals who frequently sabotage the game for their peers in a "team of one" until the teacher feels the child can re-join a team without the running the risk of further attempts of sabotage.

psychological well-being is reciprocally related to a child's social adaptation status, and there is very little if any robust empirical evidence to support this particular claim. Kellam and colleagues (1975) developed SFT almost exclusively to develop a logic model for GBG.

3.5 Is the Good Behaviour Game a successful preventive intervention?

With an understanding of the intervention implemented in the present study established, the empirical evidence about the effectiveness of GBG can be examined. The remainder of this chapter is presented in two parts. As the GBG is a UPI, the majority of studies have tested for a main 'intent to treat' effect when assessing outcomes, so these are synthesised first. The final section scrutinises the evidence on differential outcomes for risk groups, particularly children experiencing poverty.

3.5.1 Evidence for the main intervention effect

Generally, findings have provided empirical evidence supporting the hypothesis that the GBG reduces disruptive behaviour in the classroom. Two meta-analyses have examined the overall impact of the intervention. The first, conducted by Bowman-Perrott and colleagues (2016) included 21 studies that utilised a single case design and reported a large effect size (d= 1.99) across different cultural contexts. Flower and colleagues (2014), who also reported positive results, including a moderate effect size (d= .5) from 22 single case design studies as well as a high immediate treatment effect (20.38%) using hierarchical linear modelling (HLM). This suggests that the GBG would decrease DB soon after being introduced to the classroom or school. However, researchers have also included the GBG within systematic reviews and meta-analyses (e.g. Korpershoek, Harms, de Boer, van Kuijik & Doolaard, 2016; Whear et al., 2013) that have investigated a wider range of behaviour management strategies and practices. Findings from these studies have been more conservative. For instance, Korpershoek and colleagues (2016) reported an average effect of .24 for behavioural outcomes. Such disparities in these reports are likely to be due to differences in the inclusion criteria utilised in meta-analyses. Korpershoek and colleagues included only matched quasi-experimental or RCT designs, which led to the authors including only four studies specifically testing the GBG within their analyses. Further caution should be taken to interpreting the single case design meta-analyses that exclusively focus on the GBG, particularly the one conducted by Bowman-Perrott and colleagues (2016). The researchers acknowledge that their reported effect sizes may be inflated as these were converted from TauU to Cohen's d. This suggests that the reported effect size for reductions in low-level DB (d= .81) is less credible, and therefore the more conservative effect sizes reported by Korpershoek and colleagues (2016) are more reflective of the impact the GBG has on children's behavioural outcomes.

The impact of the Good Behaviour Game in other cultures

Even though the majority of evidence supporting the GBG's impact on low-level disruption is from the US, the intervention is one of the few universal prevention programs that researchers have tested across a multitude of cultures (Gu, Lai & Ye, 2011). Nolan, Houlilhan, Wanzek, and Jenson (2014) reviewed the international empirical evidence in European countries such as: Belgium (Leflot, van Lier, Onghena & Colpin, 2013), Spain (Ruiz-Olivares, Pino & Herruzo, 2010), and Canada (Dion, Roux, Landry, Fuchs, Wehby, & Dupéré, 2011). Although Nolan and colleagues provide a positive commentary for the intervention's success with Europe, interpretations of the empirical evidence must be cautious. Caution is required because many of the international tests of the GBG have involved adaptations of the intervention to fit the cultural context in which it operates. For example, the most tested iteration was developed in The Netherlands by van der Sar and Goudswarrd (2001), with adaptations that went beyond linguistic translation (see section 4.5.1 for a description of *Taakspel*). This degree of adaptation raises the question as to whether findings can be comparable to the original intervention.

Researchers have tested the GBG has also in two developing countries: Sudan (Saigh & Umar, 1983) and Belize (Nolan, Filter & Houlihan, 2014). These single case design studies did not utilise an adapted version of the intervention. Both studies reported decreases in disruptive behaviour from baseline (30.7% and approximately 7% respectively). However, as the two studies did not utilise a control group, the extent to which the GBG was responsible for the effects observed remains unanswered. With little robust and directly comparable evidence regarding the impact of the GBG outside of the US, there is an apparent gap in the literature for an adequately powered RCT in a different cultural context.

3.5.2 Differential effects of the Good Behaviour Game for at-risk groups

Some of the most robust findings of the GBG to date have been studies utilising data from an RCT that was conducted in Baltimore, USA from 1984 to 1986 (Kellam et al., 2014). Although the immediate findings did not report significant main effects, the authors tracked children from first and second grades (ages 6-8) until young adulthood. Various studies have reported differential effects regarding a range of social, behavioural, and well-being outcomes (e.g. Kellam et al., 2008; Petras et al., 2008; Poduska et al., 2008; Wilcox et al., 2008; Hendricks Brown et al., 2008). As there has been considerable interest in the moderating effects of gender on intervention outcomes, the present section reviews this first. Then, studies examining the moderating effects of poverty in relation to the GBG outcomes are examined in order to justify the need to investigate further the differential gains of children experiencing poverty in the study of UPIs.

Gender

The majority of evidence indicates that males benefit more from the GBG than females (Kellam et al., 2014). As discussed in section 3.2.3, females typically exhibit lower rates of DB throughout their development compared to their male counterparts due to a range of biological and socialisation factors. The above evidence suggests that on average, females may not benefit from the GBG as much as males because they are less likely to exhibit direct forms of DB (i.e. aggressive or impulsive behaviours) throughout their development (Kellam et al., 2011). Meanwhile, evidence indicates that reductions in conduct problems in early adulthood were only significant for males who were identified at the start of the trial having persistently high levels of DB (Poduska et al., 2007), with no significant long-term intervention effects for males of disruptive behaviour in the medium and low profile (Kellam et al., 2014; Poduska et al., 2007). Although these findings question the "universal" effect the GBG is meant to produce, the children in need of a change in behavioural trajectory appear to benefit the most from the intervention. Therefore, it could be a cheaper alternative to more expensive and time-consuming targeted interventions for at-risk males.

However, contradictory evidence suggests that the differential gains for males may not be as impressive as first thought (Humphrey et al., 2018). For instance, Petras, Masyn and Ialongo (2011) utilised data from a second iteration of the Baltimore trial which contained two experimental conditions alongside a control condition. The first was a classroom condition (CC) where the GBG was delivered alongside an academic intervention. The second condition was a household-based condition (FC) where a family-based intervention was delivered. The authors reported a marginal intervention effect (p=.07) on the transition from a highly aggressive/disruptive profile to a lower profile for males between grades 1-3 and grades 6-12 for both intervention conditions. Although the authors justified altering their significance threshold, the condition containing the GBG was associated with a smaller increase in the probability males would transition from a high aggression profile to a lower one compared to the FC condition (*CC*= 23.8%; *FC*= 34.4%). Although the GBG was not the only intervention in the CC condition, it is unlikely that the academic intervention played a significant role in the reduction of aggressive behaviours. The above evidence suggests that family-based interventions may be a suitable alternative to classroom-based interventions as the provision of behavioural expectations would consistent across multiple contexts that have a major influence in a child's development.

Children experiencing poverty

To date, there are very few studies that have specifically investigated the differential effects of the GBG among children experiencing poverty. However, with areas of socio-economic disadvantage experiencing a higher demand on resources in order to meet the needs of large numbers of children (see section 4.4.2), it is unsurprising that researchers test many UPIs in deprived neighbourhoods. Many of the smaller scale studies conducted in the US

had high percentages of participants receiving reduced or free lunches at school[13] (e.g. Lannie & McCurdy, 2007; Mitchell et al., 2015). For instance, Kleinman and Saigh (2011) reported that there was an 88% reduced/free lunch uptake within their sample (*N*= 26) and calculated substantial reductions in talking (58%), and moderate decreases in aggression (17%) and out of seat behaviour (25%).

Similarly, in the aforementioned seminal RCT conducted in Baltimore, a city known to be one of the poorest in the US, approximately half of participants received free/reduced-price lunches (Kellam et al., 2014; United States Census Bureau, 2017). As described earlier, many positive distal outcomes were found for the most disruptive males exposed to the GBG. By using sizeable samples in deprived areas, it is tempting to conclude that large groups of impoverished children will benefit differentially from the GBG, but without direct comparison to children who live in more affluent conditions, such a conclusion is not robust.

Although not the main aim of the study, Kellam and colleagues (1998) did investigate whether the GBG interacted with classroom-level[14] and individual-level poverty, while the researchers reported null effects for child-level poverty, they detected an effect at the classroom-level. Kellam and colleagues reported that classroom poverty influenced the extent to which males were at risk of being highly disruptive in the GBG classrooms as it did in control classrooms. This meant that males in high poverty classrooms were more likely to be rated as highly disruptive and aggressive in middle school. With classroom level poverty remaining a strong predictor of DB despite the GBG's implementation, contextual influences associated with school poverty may provide a moderating effect. It may be that factors such as the disparity between the supply and demand of school resources may be moderating the intervention's implementation (see section 4.4). Alternatively, a more targeted approach in reducing DB may be more appropriate for children experiencing poverty.

3.5.3 Application to the present study

Although the majority of studies testing UPIs, including the GBG, are conducted in areas of impoverishment, there are no studies to date that have directly tested whether poverty moderates intervention effects on disruptive behaviour. Although Kellam and colleagues (1994) did explore the role of poverty at the individual level and classroom level with regards to the intervention effects of the GBG, the analytical strategy did not account for the multi-level structure of the data set. With a greater understanding that contextual factors may compete or interact with UPIs to manipulate the intervention effect, and sophisticated statistical methods becoming more widely available since Kellam and colleagues' study, there is a clear need to examine whether poverty moderates the GBG's outcomes in more

---

[13] The equivalent proxy for poverty in the UK is FSM.
[14] The percentage of children receiving free or reduced-price lunch

detail. By comparing and contrasting individuals and contexts with one another, the extent to which poverty influences intervention outcomes can be established. As children invariably attend a local school, those serving impoverished neighbourhoods will have a higher population of children experiencing poverty, each with their own unique experience that has influenced their development. Therefore, a suitable analytical strategy – such as multi-level modelling - that can account for the nested nature of the data set, allows us to test the moderating effect of poverty on the GBG's effectiveness at multiple levels.

Furthermore, the reader may have noticed that gender was a risk-factor that was often explored alongside poverty throughout the chapter. The researcher thought this was necessary as gender is a variable that is controlled for as part of the analytical strategy detailed in the methodology (see section 5.8). This is because, as established in this section, "being male" is a key risk factor that is likely to explain a high proportion of the variance when modelling for contributors to disruptive behaviour outcomes. Therefore, to establish accurately the extent to which poverty contributes to the outcomes of the present study, it was decided that the inclusion of gender in this chapter, and as a controlled variable in the analysis, was necessary.

3.6 Chapter Summary

The chapter started by defining DB, which was then used as a basis to explore the developmental trajectory associated with disruption. Then, the literature review explored the role risk factors such as gender and poverty influenced DB. Subsequently, the present study introduced preventive interventions as a possible means to improve outcomes for individuals at risk of a disruptive developmental trajectory. Detail related to intervention development was explored, this was followed by a review of the effectiveness of universal preventive interventions. Then the GBG, the intervention utilised in the present study and was described in detail. Finally, a review of the evidence on the effectiveness of the GBG was presented, with an exploration of differential intervention effects for the aforementioned risk-groups. The chapter finished by applying inferences taken from the literature examined throughout the chapter to the present study.

Chapter 4
The role of poverty in the implementation of preventive interventions

4.1 Chapter introduction

First, the term "implementation" is defined and then further explored regarding its subordinate dimensions. Then, the literature review scrutinises the extent to which these dimensions moderate the outcomes of preventive interventions. Next, the researcher considers the factors that may influence the implementation of preventive interventions. Following this, social disorganisation theory is introduced as part of the theoretical framework of the present study, to justify the role poverty may have in the implementation of preventive interventions such as the GBG. Penultimately, the researcher highlights the lack of empirical evidence surrounding the delivery of the GBG. Finally, the chapter concludes with the rationale, contributions to knowledge and proposed research questions for the present study.

4.2 The role of implementation in achieving intervention outcomes

Although UPIs have been developed to address a multitude of issues that place individuals at risk of negative outcomes, there are disparities between the strength of the intervention effect during the efficacy and effectiveness stages of development (see section 3.3.2). The high levels of success achieved with developer involvement are rarely replicated when schools deliver an intervention with reduced or withdrawn assistance (Wigelsworth et al., 2016). Implementation is considered to be a critical factor that moderates the extent to which UPIs are successful in achieving the desired outcomes (Durlak and DuPre, 2008). In this section of the literature review, critical examination of the term implementation will facilitate a definition that will be applied to the present study.

4.2.1 Defining implementation

Throughout the progression of the field of prevention science, there have been multiple definitions of implementation proposed. Rogers (2003) developed one of the first: "put[ting] an innovation into use" (p.164). However, as the field further developed its understanding of implementation, more refined definitions have emerged. For instance, Lendrum and Humphrey (2012, p.635) proposed implementation is "… the process by which an intervention is put into practice". Similarly, Forman (2015, p.10) proposed implementation is "the set of activities that compose the process of putting a practice or programme into place in an organisation". However, Lendrum and Humphrey's (2012) definition will be utilised in the present study as it indicates that implementation is a continuous process that occurs

throughout the duration a school or teacher decide to put into effect, whereas Forman's (2015) definition suggests implementation is a process that occurs during the introduction of an intervention. This distinction is important as the present study follows participants over two years to observe the possibility of changes to implementation practices over time.

Macro perspective vs micro perspective of implementation

Implementation research consistently reports that teachers do not always deliver interventions as designed by the developers; this suggests that low implementation can occur despite the active use of the programme (Dusenbury, Brannigan, Falco & Hansen, 2003). Implementers of interventions are often "active modifiers" rather than "passive acceptors" of the delivery guidelines proposed by the developers (Rogers, 2003, p.180). Implementation research has identified several facets that can be manipulated to change the intervention's success in achieving desired outcomes. These are fidelity, programme reach, programme differentiation, dosage, adaptations, and participant responsiveness (see section 4.2.2 below). The macro-view of implementation views the term as a broad construct that is based on the extent to which the delivery of an intervention meets the developer's guidance (Humphrey et al., 2016). If the implementer delivers a programme at the incorrect dosage to the target audience then according to the macro-view, this would be considered poor implementation. Meanwhile, the micro-view argues that implementation is a superordinate construct that is comprised of separate facets listed above (O'Donnell, 2008). With multiple factors to consider when measuring implementation according to the micro-view, it is difficult to determine what a good standard of implementation looks like in practice.

Each perspective has implications for the design of the present study, particularly about methodological decisions relating to the measurement of implementation. For instance, by adopting the macro-view of implementation, researchers would utilise a single measure to assess the delivery of an intervention. Developers produced the GBG to be highly specific about the intervention's delivery, and provide a checklist in which the coaches and teachers can measure implementation levels (Ford et al., 2014). However, the present study has adopted a definition of implementation that uses the term "processes". The term "processes" suggests that implementation is comprised of multiple facets that work independently of each other (although they may interact), to moderate the extent to which delivery achieves the desired outcomes. Therefore, for the present study, the micro-view of implementation is adopted.

4.2.2 Aspects of implementation

As the micro-view has been accepted to form part of the present study's theoretical framework, this section defines and explores each subordinate implementation construct. These are fidelity, quality of delivery, dosage, programme reach, programme differentiation, adaptations, and participant responsiveness.

Fidelity

Fidelity can be considered as the extent the intervention was delivered as intended by the developer (Dusenbury et al., 2003). The literature often refers to the term as "adherence" or "integrity" (Durlak & DuPre, 2008). As described in the previous chapter, many programmes require implementers to follow a manual to achieve the intervention outcomes as dictated by the logic models or theories of change developed during the intervention's design stage (see section 3.3.2). If implementers alter the structures and sequences of a programme, the intended changes in behaviour may not occur. The high levels of developer involvement during the pilot and testing stages may account for higher successes rates of interventions compared to later stages of implementation testing and development (Elliot & Mihalic, 2004).

Quality of delivery

Dane and Schneider (1998) defined quality of delivery as "how well" an implementer delivers the different components of a programme. However, this definition is vague and ambiguous to operationalise for the present study. Meanwhile, O'Donnell (2008) suggests quality is concerned with the *processes* associated with delivery, rather than the *structure* of delivery (i.e. fidelity). The method in which an implementer delivers a programme is crucial as there are often interactive elements that aid pupils in developing adaptive behaviours necessary to achieve intervention outcomes. To achieve a high quality of delivery, the implementer must take on an active role, rather than relying solely on a 'script' (Dusenbury et al., 2003; Jennings & Greenberg, 2009).

Dosage

In its purest form, the dosage is the amount of intervention delivered (Durlak & Dupre, 2008). However, dosage can be measured regarding session duration and frequency of delivery (Nation et al., 2003). Both are thought to have a negative impact on intervention outcomes if either is considered low (Nation et al., 2003). Short intervals of time dedicated to a programme session may mean the critical content is not delivered, thus jeopardising intervention outcomes. On the other hand, infrequent sessions prevent pupils from practising and reinforcing key behaviours or skills that are thought to be a buffer from the negative outcomes the intervention is meant to prevent. Although developers appear to set arbitrary dosage requirements on the principle that high levels of exposure will lead to significant improvements, there has been little research to determine at what point the intervention effects reach saturation (Voils et al., 2012).

Programme reach

Programme reach is concerned with the extent the intended recipients receive the intervention (Berkel, Mauricio, Schoenfelder & Sandler, 2011; Moore et al., 2015). In the case of UPIs, all pupils in the specified population (e.g. class or school) should receive the intervention (see section 3.3.1). However, the reach of the intervention is often out of the remit of the implementer, with one in ten pupils considered to be persistently absent

throughout the academic year (DfE, 2018). However, there may be factors within the school's control that influences programme reach, such as withdrawing recipients to complete other tasks (Humphrey et al., 2016).

Programme differentiation

Programme differentiation refers to the distinctiveness of the intervention theory and procedures from other practices or interventions within the setting (Naylor et al., 2015; Durlak & Dupre, 2008). Identifying how UPIs are different to established practices allows insight into which components are responsible for any observed intervention effects (Dusenbury et al., 2003). Programme differentiation may also enhance our understanding of the role of familiarity with intervention concepts as either a barrier or facilitator for implementation (see section 4.3).

Adaptations

An adaptation to an intervention is when the programme procedure has changes made in the form of either: additional content, removal of content, or the modification of existing content (Naylor et al., 2015; Durlak & Dupre, 2008). The use of adaptations is a contentious issue within implementation science. Many developers argue that changes to the intervention structure will render the programme ineffective (Dusenbury et al., 2003; O'Donnell, 2008). However, others argue that adaptation may serve a positive function, such as increasing a sense of ownership of the intervention (Blakey et al., 1987; Castro, Barrea & Martinez, 2004).

Nevertheless, it is generally accepted that structural changes, such as omissions of critical components, are detrimental to the effectiveness of an intervention, whereas surface level changes (e.g. changing names, pictures or vocabulary) may improve engagement and understanding of the intervention components for recipients (Ferrer-Wreder, Sundell & Mansoory, 2012). Critical component analysis during the development of interventions are rare, so there is little understanding of what constitutes a 'key component' in the majority of interventions (Humphrey et al., 2016). Until implementation and process evaluation procedures are considered standard practice in trials, the exact role of adaptations in moderating intervention effects will remain elusive.

Participant responsiveness

Participant responsiveness is the extent to which the intended recipients engage with the intervention (Carroll et al., 2007). This facet of implementation could be considered to be an extension of the programme reach, as exposure to the intervention does not guarantee improvements in the desired outcomes. If children do not engage with the intervention components, they are less likely to learn the adaptive skills and behaviours that are thought to prevent (as dictated by the logic model or theory of change) negative outcomes in the future (Carroll et al., 2007).

4.2.3 Theoretical model of implementation

Before considering the extent to which the aforementioned factors of implementation moderate intervention outcomes, it is essential to consider *how* these dimensions influence programme outcomes. Despite the wealth of empirical evidence that associates implementation variability with variable intervention success, up until recently there has been no theoretical framework to assimilate the trends in implementation (Berkel, Mauricio, Schoenfelder & Sandler, 2011). However, Berkel and colleagues (2011) have proposed a model, which suggests that the different aspects of implementation interact to produce variability in intervention delivery (see figure 4.1). Specifically, the dimensions associated with the facilitator (i.e. fidelity, quality, and adaptation) influence the extent to which the participants respond to the programme, and this in turn moderates intervention outcomes. Although the model has yet to be tested, mainly as there is a lack of empirical evidence measuring participant responsiveness (see section 4.2.4), programmes implemented with low fidelity or low levels of participant responsiveness are less likely to successfully achieve their outcomes (Hansen et al., 1991). Therefore, the present study measures multiple implementation dimensions to reflect the complex interactions that may occur between the facets during intervention delivery.

Figure 4.1
An integrated theoretical model of programme implementation, adapted from (Berkel et al., 2011).
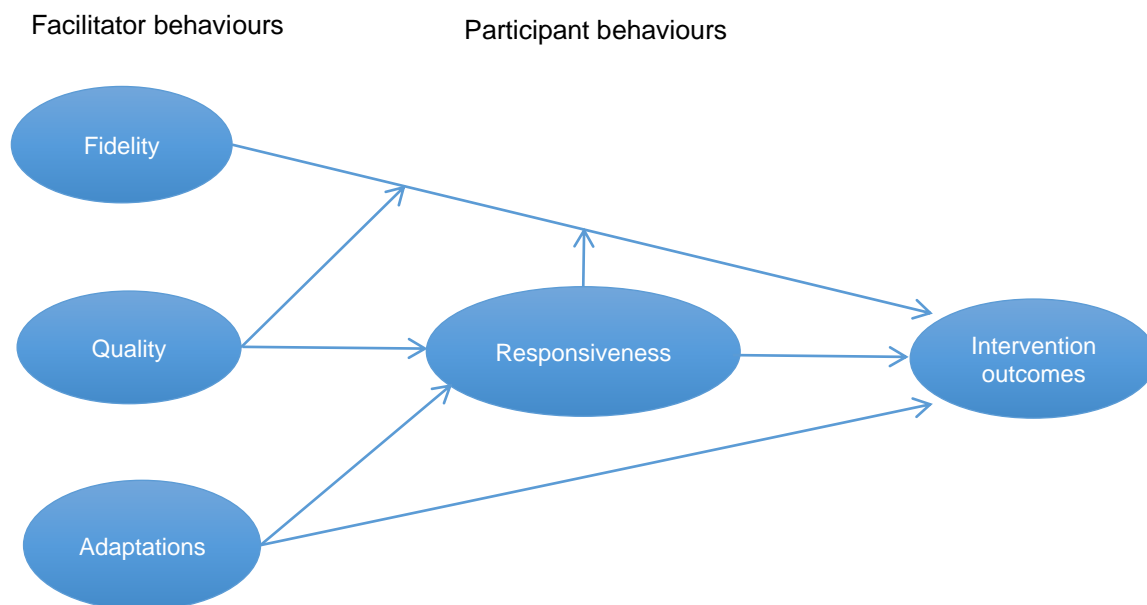
4.2.4 Measuring implementation variability

This section of the literature review describes the methods available to researchers who wish to conduct process evaluation. By providing detail of these methods, the following sections of the literature review will be framed in a context that aids scrutiny of the available empirical evidence. While researchers generally accept that relying on one method will be

of little value to understanding the relationship between implementation and intervention outcomes, there is no one combination of methods which capture all available data (Borglin, 2015). Although there are frameworks where both qualitative and quantitative methods could be used (e.g. Borglin, 2015), as the present study takes a post-positivist stance (see section 5.3.2), only quantitative methods are covered in this section.

Assessing the different aspects of implementation

Quantitative methods for assessing the different aspects of implementation provide an overview of implementation variability, both across a sample and over a specific period (Moore et al., 2015). Although there are standardised measures for assessing implementation dimensions (e.g. Bishop et al., 2014; Dariotis, Bumbarger, Duncan & Greenberg, 2008), these do not account for the individual specifications of delivery for different programmes (Humphrey et al., 2016; Dusenbury et al., 2003). However, bespoke measures, such as the one utilised for the present study, should undergo the same rigorous testing through piloting and reporting of psychometric properties before they are used more widely (Humphrey et al., 2016). Another consideration is the consequences of possible researcher effects influencing the implementers' behaviour. For instance, Smith, Schneider and Smith's (2004) review of school-wide anti-bullying interventions found systematic monitoring produced larger effect sizes compared to studies where implementation was not measured. As a result of anticipating observation, implementers may ensure their delivery (or reporting of it) is in keeping with developer expectations (Hansen, Pankratz & Bishop, 2014).

Table 4.1 presents examples of quantitative assessment for each dimension of implementation. Generally, collecting numerical data for implementation can be achieved through a structured observation or survey. Researchers design such instruments with items that specifically measure implementation according to whether implementers meet the requirements for each dimension (see section 4.2.2 for indicators of good implementation practice according to each aspect of implementation). However, the limitations of quantitative methods make such methods inappropriate for certain aspects of implementation (e.g. adaptations: Nastasi et al., 2007). In such instances, qualitative methods would be more appropriate to gather data (Moore et al., 2015). Even so, quantitative methods are used to measure the implementation variability of the dimensions utilised in the present study (i.e. fidelity, quality of delivery, and dosage).

Table 4.1

Example quantitative methods for collecting quantitative data on implementation dimensions (adapted from Humphrey et al., 2016).

Aspect of implementation	Example method
Fidelity	Rating the extent to which the implementer followed the manual/session protocol.
Quality	Rating dimensions such as implementer interest, preparedness, and clarity of responsiveness during intervention delivery.
Dosage	Recording the number of sessions delivered or time spent delivering the intervention.
Participant responsiveness	Rating the extent to which recipients engage with the intervention during delivery.
Reach	Recording the number of intended recipients of the intervention is present during the programme.
Adaptation	Recording the number of adaptations made during an intervention session.
Programme differentiation	Comparing usual practice data before and at the end of the trial.

Sources of quantitative data to assess implementation

As multiple stakeholders are involved in the trialling of a preventive intervention, there is an opportunity to utilise a diverse set of resources to collect implementation data. Table 4.2 summarises the advantages and disadvantages of four typical sources of implementation data. Researchers argue that including both external sources (i.e. researcher or technical support) and self-report implementation is the most acceptable combination of sources to generate credible findings (Hansen et al., 2014). However, factors such as time and cost are critical determinants of such methodological decisions in implementation research, often making self-report data an attractive option (Lillehoj, Griffin & Spoth, 2004). The attraction is because asking teachers to rate their implementation levels often guards against data burden and is inexpensive compared to sending researchers to intervention sites (Humphrey et al., 2016). Despite this, the present study utilises data from structured observations carried out by trained researchers, as it is considered the "gold standard" of implementation assessment (Hansen et al., 2014). Section 5.6.3 discusses the methodological decisions and justifications employed by the present study in more depth.

Table 4.2
A summary of the advantages and disadvantages to utilising different sources for data collection (adapted from Humphrey et al., 2016).

Source	Advantages	Disadvantages
Researchers	High in rigour and objectivity. Stronger associations with intervention outcomes than other sources.	Time-consuming and costly. Intensive observation may include interaction with implementation.
Implementer	Low cost and quick to complete. Provides a summative account for a specified period of implementation.	Subject to biases such as impression management or demand characteristics. Weaker associations with intervention outcomes than other sources.
External support	Minimal data burden on schools. Increased external validity.	Reduced evaluator control (e.g. focus, consent, inter-rater reliability). Data may be biased by the relationship between external support and implementers.
Participants[15]	A strong indicator of participant responsiveness.	Also subject to biases such as impression management or demand characteristics. Lack of knowledge of fidelity requirements.

4.2.5 Implementation variability as a moderator of intervention effects

This section of the chapter reviews the evidence that supports the extent to which the aforementioned aspects of implementation moderate intervention outcomes. Several reviews reporting implementation variability have been produced since the need for addressing intervention delivery has been raised (e.g. Durlak & Dupre, 2008; Dane & Schneider, 1998; Dusenbury et al., 2003, O'Donnell, 2008). However, empirical evidence for the different aspects of implementation is imbalanced, as researchers place a heavier focus on fidelity and dosage compared to other constructs. With this in mind, fidelity and dosage each receive dedicated sections which synthesise the literature, whereas the remaining aspects (quality of delivery, programme differentiation, programme reach, adaptations to interventions, and participant responsiveness) are discussed as a single section.

Fidelity

Researchers generally agree that lower levels of fidelity will negatively impact intervention outcomes (Durlak & Dupre, 2008; Dane & Schneider, 1998; O'Donnell, 2008). For instance, Haataja and colleagues (2014) reported that higher fidelity to the *KiVa* anti-bullying intervention sessions was associated with reduced levels of victimisation in classrooms.

---

[15] It is rare that participants are asked about implementation aspects other than participant responsiveness, and that is usually achieved through qualitative means.

Meanwhile, Pettigrew and colleagues (2014) found that students who received poorly implemented sessions of the *Keepin' it REAL* programme had a lower success in reduced substance use compared to high adherence to the intervention. Dusenbury and colleagues (2003) argue that low levels of fidelity are often due to omissions of practices or intervention content during delivery. However, not all these omissions are intentional, as a variety of factors influence the extent to which an intervention is delivered with fidelity (see section 4.3). A lack of fidelity across implementers has led to calls for further support for teachers through channels such as extra training or supervision (Dane & Schneider, 1998; Adelman & Taylor, 2000; Greenberg, 2010). With the recognition that technical assistance may encourage higher fidelity levels, an evidence base is beginning to build up to support this notion (e.g. Kretlow & Bartholomew, 2010; Stormont et al., 2015).

For instance, Pas, Bradshaw and Cash (2014) reviewed the coaching models for several interventions which aimed to either improve children's behavioural, SEL, or academic outcomes and concluded that practices such as reflection and guided practice were effective in improving fidelity levels. With more interventions providing technical assistance, usually in the form of coaching, developer involvement no longer completely ceases once the scale-up stage is reached. Technical assistance is distinct from training as it is an individualised and on-going relationship between the developer and the implementer (Chinman et al., 2005). By providing regular contact through the provision of observation and feedback, technical assistance has been identified as a promising way of addressing the disparity of fidelity levels between efficacy and effectiveness stages (Tibbits, Bumbarger, Kyler & Perkins, 2010; Lee et al., 2008). As researchers consistently identify fidelity as a crucial aspect of implementation within the existing evidence base, the present study includes this construct.

Dosage

Like fidelity, the consensus in the literature is that low levels of dosage will negatively impact intervention outcomes (Durlak & DuPre, 2008; Dane & Schneider, 1998; O'Donnell, 2008). For instance, Rosenblatt and Elias (2008) used self-reported dosage from *Talking with TJ*, an SEL intervention, to group classes by dosage levels (i.e. high, medium, and low). The researchers reported that classes with high levels of dosage had smaller decreases in their grade point average during the transition to middle school compared to their peers in lower dosage classrooms. Similarly, Ferrer-Wreder and colleagues (2010) reported that classes with higher levels of dosage of drug use prevention interventions (*Life Skills Training* and *Time Wise*) were more likely to refuse drugs. Dosage is one of the most frequently assessed aspects of implementation as it is thought to be relatively easy to measure (Durlak & DuPre, 2008). However, this is somewhat of a fallacy, as it is often conceptualised as unidimensional when this is not the case (see: 3.2.2). This assumption has led to bias within the literature surrounding dosage levels, with the *frequency* of intervention delivery being measured over other dimensions (Voils et al., 2012). The bias is particularly apparent within

systematic reviews, with researchers limiting their inclusion criteria to studies that measure the frequency of programme sessions (e.g. Dane & Schneider, 1998). However, there is an increased understanding that adopting a unidimensional understanding of dosage is reductionist and is likely to account for some inconsistencies in the literature (e.g. Malvin, Moskowitz, Schaeffer & Schaps, 1984). For example, Nation and colleagues (2003) utilised a multidimensional definition of dosage (i.e. frequency and length) as part of their inclusion criteria for their review and were able to provide more insight into the application of interventions within multiple contexts as a result of this.

Multidimensional approaches to measuring dosage have been developed and are increasingly advocated in process evaluations (Voils et al., 2012; Warren, Fey & Yoder, 2007; Polanin & Espelage, 2015). One method is cumulative intervention intensity (CII), which calculates the overall exposure an individual has to an intervention by multiplying the length of an individual 'dose' by the frequency with which a teacher dispenses a programme and the overall delivery period (Warren et al., 2007). However, Polanin and Espelage (2015) appear to be the only researchers to utilise CII within a UPI evaluation (Second Step). They reported an *inverse* correlation with CII scores and SEL outcomes and argued that this might be due to teachers spending more time on the Second Step lessons because their class had greater SEL deficits. The inverse correlation suggests the intervention was likely a poor fit for these classrooms and that an SPI would have been more appropriate. However, until practices such as CII are more widespread in prevention research, the assumption that "more is better" will continue to prevail (Warren et al., 2007). With this in mind, the present study includes, with the aim of incorporating a multi-dimensional approach (see section 5.6.3).

Remaining aspects of implementation

The remaining aspects of implementation have received relatively little attention, and coverage for each dimension is variable (Domitrovich & Greenberg, 2000). For instance, Durlak and DuPre (2008) identified only 10% of papers in their seminal review that had studied quality, while programme differentiation and participant responsiveness were not included due to a lack of examples in the literature. This lack of empirical evidence for some aspects of implementation is a concern as they may be of equal importance to fidelity or dosage in determining the successful delivery of an intervention (Durlak, 2015). Indeed, this proposition is borne out in the few studies that have focused on the less 'popular' aspects of implementation (i.e. quality, programme differentiation, programme reach, adaptations to interventions, and participant responsiveness) (Dane & Schneider, 1998; Durlak & DuPre, 2008; Domitrovich, et al., 2010). Quality of delivery is often the aspect measured in addition to fidelity and dosage, and an increasing evidence base suggests that it is an essential factor to consider when establishing the successful implementation of a programme (Humphrey, Barlow & Lendrum, 2018; Kam, Greenberg & Walls, 2003; Spoth et al., 2007).

At the same time, aspects of implementation such as programme differentiation and participant responsiveness have received very little attention, and the sparse evidence available concerns mainly indicated prevention interventions. For instance, Hogue, Liddle, Singer and Leckrone (2005) conducted a process evaluation of multidimensional family prevention intervention for drug use and included programme differentiation within their study. Researchers compared the intervention to multidimensional family therapy and cognitive therapy. They concluded there was little differentiation between the family therapy and the intervention. Meanwhile, Hansen, Graham, Wolkenstein and Rohrbach (1991) utilised the proxies of class enthusiasm and student participation as measures of participant responsiveness when evaluating two drug prevention interventions (Resistance Training and Normative Education). They reported that the dimension of implementation was a significant moderator for programme outcomes. However, researchers slowly recognise participant responsiveness as an aspect of implementation which requires more research attention, particularly as it may indicate the extent to which intervention recipients have internalised the concepts of a programme (Mihalic, 2004).

With the above in mind, the present study measures quality in order to further expand the literature base and understanding of implementation as a whole (Elliot & Mihalic, 2004).

4.3 Factors affecting implementation

With implementation variability established as a moderator of intervention outcomes, the question must be asked: "what influences implementation variability"? The literature indicates that a multitude of factors are likely to be responsible (Durlak & DuPre, 2008; Domitrovich et al., 2008). These are theorised to influence implementation at different levels, much like ecological systems theory (Bronfenbrenner, 1992; Domitrovich et al., 2008). This section of the literature review describes four levels of factors that affect implementation. These are the intervention level, the implementer level, the classroom level, and the school level.

4.3.1 Intervention level factors

Factors related to this level are concerned with the properties of the intervention (e.g. the materials or processes required to implement the programme). During the development process of a programme, the developers must design a manual that contains enough detail to allow the implementer to deliver all aspects of the intervention as they intended (see section 3.3.2). Unsurprisingly, vaguely described interventions have been delivered at lower levels of implementation fidelity (Carroll et al., 2007; Dusenbury et al., 2003; Mihalic, 2004). Precise programme instructions enhance fidelity by addressing any uncertainty or confusion surrounding implementation. A factor that may influence this is the complexity of the intervention (Carroll et al., 2007). Many preventive interventions comprise multiple strands which either address different outcomes (e.g. Head Start) or address the same outcome

based on the level of risk (e.g. PBIS). Although these programmes are thought to be superior to single-stranded interventions because of their multi-functional approach, in the longer term, their implementation is often unsustainable (Dusenbury et al., 2003; Greenlaugh et al., 2004). The lack of long-term implementation is because the additional delivery requirements for each additional intervention component may exceed the organisational capacity of the school (see section 4.4.2). Therefore, while it may be tempting for schools to adopt complex interventions which aim to address multiple issues, the empirical evidence suggests multi-component interventions are no more effective than single-strand programmes (Dusenbury et al., 2003).

Even though providing precise instructions to optimise fidelity is the dominant stance in implementation science, a conflicting perspective argues that interventions require a degree of flexibility, i.e. implementers should be able to make adaptations to the programme in order to fit the needs of children and school (Bernal & Sáez-Santiago, 2006; Castro et al., 2004). This balance of fidelity and adaptability has often been debated, particularly concerning the cultural transferability of UPIs (Castro et al., 2004; Ferrer-Wreder et al., 2012; Castro & Yasui, 2017). Although "culturally blind" preventive programmes are less likely to be sustainably implemented and adopted, adaptions to core intervention components are likely to inhibit outcome effects (Kumpfer, Alvardo, Smith & Bellamy, 2002). This inhibition of intervention effects is due to the logic model/theory of change that drives behaviour modification no longer reflecting the problem theory identified during the development stage of the programme's design (see section 3.3.2).

However, unless a critical component analysis has taken place, it is not known which elements are responsible for the intervention's effects (Mihalic, 2004). While this premise is used by those who place fidelity higher than local need as an argument against intervention adaptation (e.g. Mihalic, 2004), tracking the successes and failures in programme alterations may be a feasible alternative to the costly and time-consuming critical component analysis. Nevertheless, a degree of compromise must be made when transferring UPIs to a different culture (Castro et al., 2004). The translation of interventions to different cultures has led to several models being developed to illustrate how developers alter their intervention in order to fit the context in which they will operate (see Ferrer-Wreder et al., 2012 for an overview of proposed models). It is generally accepted that surface level adaptations to materials (such as Anglicising the GBG manual in the case of the present study) are positive changes accepted by both implementers and developers, while deeper adaptations are still controversial (Resnicow et al., 2000). Even so, the fidelity-adaptation debate about cultural transferability is pertinent to the present study, as the GBG was developed in the US. Although surface level adaptations are minimal, it is unknown whether the core components of the intervention are suitable for UK classrooms.

4.3.2 Implementer level factors

Implementer factors include their professional and psychological characteristics. Domitrovich and colleagues (2008) emphasise that the majority of literature on the influence of implementer factors is theoretical (e.g. Han & Weiss, 2005; Jennings & Greenberg, 2008) rather than empirical. However, the evidence base is beginning to grow. For instance, there has been speculation that teachers at the start of their career may implement preventive interventions at higher levels as they have yet to build their own instructional and behaviour management strategies, whereas more experienced teachers would be resistant to change their practices because they are confident in their abilities (Coburn, 2004). Although the degree of support for this proposition varies (Ransford et al., 2009; Stein et al., 2008; Harn, Parisi & Stoolmiller, 2010). Many studies that investigate this area rely on self-report data, which is prone to inflated scores due to factors such as self-selection bias or demand characteristics (see section 4.2.4 and 5.6.3). Therefore, researchers should pursue more objective approaches to assessing implementer characteristics and their influence in order to further understanding in this area.

Meanwhile, the psychological characteristics of implementers are thought to be a significant contribution to implementation variability (Domitrovich et al., 2008). Kessler (1999) argues such factors are of importance when implementing preventive programmes that aim to reduce the frequency and intensity of maladaptive behaviours associated with SEL, behavioural, and mental health outcomes. Psychological characteristics are an essential factor because implementers' must have an awareness of their own needs and competencies, as well as their audience, in order to successfully deliver interventions that target these "soft skills" (Kessler, 1999). Factors such as stress levels and self-efficacy have been identified as contributors to implementation variability (Ross, Romer & Horner, 2012; Reinke, Lewis-Palmer & Merrell, 2008; Wehby et al., 2011). Teacher stress is a pertinent issue within the UK educational system, as teachers face increasing levels of accountability both within and outside the school context (Kidger et al., 2016). Research indicates that multiple stressors are associated with an increased risk of teacher burn out and staff turnover (Brown, Davis & Johnson, 2002; Skaalvik & Skaalvik, 2016, Kyriacou, 1987). Such stressors range from poor pupil behaviour to a lack of emotional and organisational support (Brown, Davis & Johnson, 2002). Therefore, teachers situated in a school community that has a robust organisational capacity can deliver interventions at higher levels of implementation due to more support, resources and time (Oullette et al., 2018). The extent to which organisational capacity influences implementation levels is discussed further in section 4.3.4.

Finally, the perceptions and attitudes of an implementer towards the intervention are factors that can serve as either barriers or facilitators of implementation (Domitrovich et al., 2008; Rohrbach, Graham & Hansen, 1993). More specifically, these perceptions relate to the

extent teachers see the intervention as addressing the needs of their class, as well as whether the programme will achieve the desired outcomes (Durlak & DuPre, 2008). Teacher "buy in" to the intervention is thought to be a key determinant of this (Castro et al., 2004). If teachers are not motivated to implement the intervention, particularly if they are not convinced of its social validity (see section 3.3.2), cessation is likely (Domitrovich et al., 2008). Positive perceptions of the intervention, such as viewing it as a convenient means to address a current problem, or better than current practice, is thought to facilitate implementation (Elias, Zins, Gracyzk & Weissberg, 2003; Ringwalt et al., 2003). Therefore, in the stages of pre-implementation, such as during training, it is key for developers to demonstrate the intervention's fit or adapt the intervention to the local needs, in order to ensure "buy-in" (Durlak & DuPre, 2008).

4.3.3 Classroom level factors

Factors that influence implementation at the classroom level are often related to the demographic, organisational and climate characteristics of the class. However, these factors are the least studied of all the levels (Wanless & Domitrovich, 2015). Classroom factors may influence processes specific to the programme. For instance, the organisation of the class timetable is likely to influence the frequency and duration with which a teacher can deliver the intervention (Ozer et al., 1997). Classroom-level organisational factors are an essential issue in the UK context, as teachers are under pressure to deliver content that meets curriculum criteria, leaving little flexibility for additional initiatives (Lendrum & Humphrey, 2012). Further compounding this is the interaction between the form of the intervention and the day-to-day schedule of the class. For instance, many UPIs follow programme lesson plans and require the teacher to make an executive decision regarding *when* implementation occurs, and whether it is a priority amidst other competing demands (Miller et al., 2011; Evans, 2012; Fabiano et al., 2014). Forman and colleagues (2009) reported that developers of interventions felt that teachers often considered UPIs a low priority, while Baldacchino (2017) found that while they often saw value in the issues a given programme aimed to address, agendas set by the school or state often did not allow the dedication required for successful implementation.

Class climate and composition could also facilitate or hinder the implementation of a UPI (Domitrovich et al., 2008). For instance, DB is a commonly reported issue reported by teachers and a significant contributor to lost teaching time (see: section 3.2). These losses in time may not only contribute towards lower dosage levels but could also indicate lower engagement with the intervention. However, empirical evidence is lacking in the pupils' impact on implementation, and therefore this proposition remains speculative (Wanless & Domtrovich, 2015). Intervention fit is an essential factor when implementing a UPI within a classroom setting (Domitrovich et al., 2008; Durlak & DuPre, 2008), mainly as pupils must be able to comprehend the instructions or concepts in order to internalise the behaviour changes required to produce the intended intervention outcomes. If there is a poor fit

between the intervention and the pupils, the implementer may adapt it to better suit class needs or cease implementation. Even so, until process evaluations of interventions extend to gathering data to establish an evidence base on classroom climate, the influence of pupils on the implementation of UPIs will remain unknown.

4.3.4 School factors

Finally, the broader context of the school is a crucial factor that influences the implementation of an intervention as it is often the unseen aspects of school factors such as policies and wider practice that have an influence on the implementer and the recipients of the intervention (Durlak & DuPre, 2008). A growing body of empirical evidence indicates that schools with maladaptive organisational and staff relationships, as well as poor leadership from the headteacher, have poorer implementation levels (Gregory, Henry, Schoeny & The Metropolitan Area Child Study Research Group, 2007; Kam et al., 2003; Thapa, Cohen, Guffey & Higgins-D'Alesssandro, 2013). Higher levels of commitment from leadership are therefore required to foster and maintain substantive changes in the everyday practices of the teachers, such as the implementation of a preventive intervention (Kam et al., 2003). Head teachers often show their commitment to new initiatives through the allocation of resources, time and support (Adelman & Taylor, 2003; Schien, 1992). If there is disinterest with the intervention at higher levels of the school organisational structure, teachers are more likely to decrease their time and commitment to the programme, particularly if the head teacher wishes other priorities to take precedence (Baldacchino, 2017).

However, it is often the case that schools lack the organisational capacity to sustainably implement new initiatives rather than a lack of interest (Thaker et al., 2008). Botvin (2004) argues that it is factors such as budgeting difficulties, high student or staff turnover, a large school population, and increased pressures on academic achievement, which leave schools struggling to meet the demands of meeting implementation requirements set by developers. Many schools that have poor organisational capacity are situated within communities experiencing poverty (Gottfredson, Gottfredson, Payne & Gottfredson, 2005; Welsh, Greene & Jenkins, 1999), which means their pupil population are at a higher risk of a multitude of negative outcomes (see section 2.4 for child-level and community level outcomes associated with poverty). With higher levels of demand for resources, time and support, schools are often unable to meet the needs of their population. The level of demand impacts day-to-day functioning as well as the academic attitudes and achievement of their pupils (Gottfredson, 2001; Ozer, 2006). Thus, it appears that the schools with the greatest need for preventive interventions are in the weakest position to successfully implement them (Gregory et al., 2007). The disparity in organisational capacity could lead to a "Matthew Effect" (see section 3.3.3) at the catchment level, with schools in more affluent areas having favourable climates that allow them to implement UPIs more successfully (Gregory et al.,

2007). As there is little empirical evidence for the role of school-level poverty in intervention implementation, this is an additional exploratory focus in the present study.

4.4 Poverty as a factor affecting implementation

As established in the first chapter, poverty negatively influences a variety of outcomes at multiple levels of society (see section 2.4). Researchers have theorised that its influence may moderate the extent to which an intervention is successfully implemented within schools (see section 4.3.4). This is thought to be due to a lack of organisational capacity as schools serving impoverished communities are often overstretched regarding resources, support and time due to serving a pupil population with higher levels of need (Gottfredson, Jones & Gore, 2002). This section of the literature review presents social disorganisation theory (SDT) as a part of the theoretical framework of the present study in order to explain poverty's role in affecting implementation variability between schools. Subsequently, the researcher examines the effect implementation of preventive interventions has in schools serving impoverished communities in order to determine the suitability of SDT in the present study.

4.4.1 Social disorganisation theory

SDT (Shaw, McKay & Hayner, 1942) is a seminal sociological theory that explains crime rate variability in neighbourhoods. SDT proposes that structural factors such as low economic status, ethnic heterogeneity, family disruption, urbanisation and residential mobility, lead to disruption within the social structure of the community (Sampson & Groves, 1989; Sampson, Raudnebush & Earls, 1997). This disruption is because areas with fractured social networks prevent the collective supervision of residential issues, such as forming common values that maintain social control (Kornhauser, 1978). The structure of a community is therefore composed of the prevalence and interdependence of informal and formal social networks; both forms of social networks are required to allow collective problem solving for local issues to occur (Putnam, 2000). A lack of formal social ties (e.g. membership to political organisations) can lead community members to feel disenfranchised and demotivated to address local issues, or engage in criminal behaviours, thus resulting in further deterioration of positive social structures (Putnam, 2000). Meanwhile, a lack of kinship between community members (e.g. casual conversation between neighbours) leads to residents feeling isolated from the wider community and showing a reluctance to interact with or offer support to other members within the neighbourhood (Taylor, 1996).

The role of social capital in maintaining social order

Section 2.3.2 theorised that low levels of social capital are responsible for persistent. SDT can be considered an extension of SCT, as the alignment of both formal and informal ties with community values foster higher levels of social capital. This alignment in values, in turn,

continues the further generation of social networks and organisations to continue addressing local issues and thus improving the living standards of the residential area (Kay, 2005). Research consistently demonstrates across multiple social spheres that residents who reside in more affluent areas are more active in contributing towards their community and broader society (examples include: political involvement: Bartle, Birch, & Skirmuntt, 2017; informal gatherings: Putnam, 2000; civic engagement: Haezewindt, 2003).

Generalised reciprocity is thought to be the foundation that socially organised communities utilise in order to maintain social order (Schuller, Baron & Field, 2000). Generalised reciprocity is when an individual exchanges labour, money, or other resources to another without the expectation of an immediate return (Coleman, 1988). The return may not be in the same form or from the same person, but in a community with strong social networks and higher levels of capital, if an individual is in need, other members are more likely to provide aid or assistance (Putman, 2000). The juxtaposition to this is a community where individuals do not partake in such exchanges in order to maintain the common good within the neighbourhood. In such communities, there are higher levels of deprivation and delinquency as the social networks within the residential area either do not serve a common goal of maintaining a certain standard of living and safety or there is not enough investment of social capital to address the local problems (Putnam, 2000).

However, the raising of community standards is not just the responsibility of the citizens that inhabit the neighbourhood; often there are state and private services that also operate within the geographical area which either directly (e.g. schools, clinics, and charities) or indirectly (e.g. businesses providing jobs or trade) meet local needs. Links between informal and formal social networks must exist in order for the community to develop in a direction that satisfies the common good (Coleman, 1988). If organisations, institutes and state-provided services are not meeting the expectations of local people, then communities with high levels of social organisation members can invest their social capital to hold these establishments accountable (Putman, 2000).

Meanwhile, many services and organisations that have intentions to improve a neighbourhood operate within communities that are difficult to engage. The lack of engagement is because community members have a lack of ties to each other and are not interested in working with organisations to solve local problems (Putman, 2000; Sampson & Groves, 1989; Pichler & Wallace, 2007). The principal reasons for this are high residential mobility and low generalised reciprocity. Socially disorganised communities have high residential mobility, which means organisations often are unable to spend time addressing individual concerns as residents move in and out of the area (Shaw & McKay, 1942). With a high turnover of residents, social connections are harder to maintain as community members do not stay in the area long enough to form social bonds where enough trust is

established for reciprocal exchanges to occur (Dempsey, Bramley, Power & Brown, 2011). This lack of social capital within a community puts further pressure on organisations to solve local problems the residents are detached from, leading to services such as charities, schools and health centres becoming overwhelmed.

4.4.2 Applying social disorganisation theory to the school context

Although Shaw and colleagues (1942) developed SDT within the field of criminology, researchers have applied the theory to educational settings (e.g. Akiba, 2010; Bradshaw, Slawyer & O'Brennan, 2009). Bevans, Bradshaw, Miech and Leaf (2007) suggest that factors such as teacher turnover, pupil mobility, social cohesion, and the scale of poverty in the school population can influence the extent to which a school can meet the needs of its pupils. A school with low organisational capacity is also less able to exert social control over the population; this leads to high levels of DB within classrooms, school violence, the incidence of bullying, and truancy (Barnes et al., 2006; Bradshaw et a., 2009; Akiba, 2010). With the structural components of a school compromising the extent to which it can fulfil its objectives, the collective efficacy of the school population is also reduced (Akiba, 2010). Collective efficacy is the shared belief that through mutual action, a group can achieve positive outcomes (Bandura, 1977). Low collective efficacy amongst teachers can lead to further deterioration of the social structures of the school as teachers may feel demoralised and disempowered to bring about positive change (Goddard, Hoy & Woolfolk Hoy, 2000). This cycle of lowered collective efficacy and depleted organisational capacity is likely to increase teacher stress levels, which may explain high levels of burn out within the teaching population and the high turnover of staff in schools serving impoverished communities (Klassen, Usher & Bong, 2010; Stephanou, Gkavras & Doulkeridou, 2013).

The role of poverty in school disorganisation

Although the role of poverty, in the form of social capital, has been explored in relation to the general application of SDT (see section 4.4.1), it is vital to establish the role of impoverishment and SDT within the school context before applying the framework to the implementation of interventions. The need to establish the above is because the school can be considered to be a social entity with its organisational structures, but also, educational settings form part of the social fabric of a wider community, to which SDT is usually applied. If a school serves an impoverished neighbourhood, it is likely to be operating within a location that has low social capital and high disorganisation (Shaw & McKay, 1942; Sampson & Grove, 1989). As established in previous chapters, researchers have reported an association between poverty and with poor living conditions, this includes high levels of crime, poor quality services or facilities, and unsafe housing (see section 2.4.2). Furthermore, at the individual level, large bodies of evidence indicate that poverty is associated with adverse social, academic and behavioural outcomes, mainly if it is persistent and occurs in early childhood (see section 2.4.1).

Schools serving deprived neighbourhoods often educate a pupil population that on average have higher levels of need compared to the rest of the population (see section 2.4.2). Though schools in England receive extra funds per child experiencing poverty (e.g. Pupil Premium), utilising these resources to develop organised social structures through fostering effective teaching practices and policies is challenging (Lupton, 2005). Staff within schools serving disadvantaged areas engage in "daily firefighting" whereby the volume of immediate crises that require attention detract from teachers' time and energy for reflection and planning (Lupton, 2005). Meanwhile, children from disorganised neighbourhoods are at a higher risk of reinforcing disruptive and delinquent behaviours through socialising with each other at school, thus further reinforcing community disorder and reduced collective efficacy on the school's ability to reinstate order (Bonnell et al., 2013). For instance, Hopson and Lee (2011) reported that perceptions of the school climate moderated the extent to which poverty influenced behavioural outcomes of pupils in schools. These findings suggest that improving the organisational capacity of schools requires efforts to improve the organisational capacity of the surrounding residential area to relieve high levels of social disorder within educational settings (Lupton, 2005).

School-level poverty, organisational capacity, and the implementation of preventive interventions

One solution to address the high level of need faced by schools with compromised organisational capacity is to implement a preventive intervention. However, as touched on previously, disorganised schools are often paradoxically the settings in the greatest need of the improved outcomes offered through an intervention but are the least able to implement it successfully (see section 4.3.4). Although preventive interventions are designed to address issues associated with a social disorder, such as DB, these programmes further deplete the already low resources that staff invest in order to maintain the school's already weak social structure. In addition to this, schools develop collective efficacy through formalised processes such as policy and teacher practice (Akiba, 2010). If teachers perceive the current procedures as ineffective, they are less motivated to implement the practices promoted by the higher tiers of the school organisational structures (Akiba, 2010). It, therefore, would be reasonable to extend a lack of collective efficacy to the administrator's choice of preventive intervention.

4.4.3 Evidence for organisational capacity moderating intervention implementation

Although the empirical evidence is somewhat limited, there are mixed findings for the role of school organisational capacity in moderating the implementation of preventive interventions. For instance, some studies report that teacher-teacher relations were a significant moderator of implementation (e.g. Beets et al., 2008; Greggory et al., 2007) while others reported null findings (e.g. Low et al., 2013; Malloy et al., 2015). While there are mixed findings for other dimensions of organisational capacity, such as administration

support (see section 4.3.4), there are other features that have received no empirical attention (e.g. adequate resources). One possible explanation for null and negative findings surrounding strong organisational capacity and intervention implementation is that individuals assume that high levels of implementation would be automatic, and so invested less time and resources than required for the successful delivery of a preventive programme (Payne & Eckert, 2010). Another explanation could be that negative school climates could be considered a challenge, and thus increasing an implementers' motivation to succeed in implementing the intervention (Klimes-Dougan et al., 2009). The above explanation suggests that the implementer's efficacy in bringing about positive change within the school context may be an interactive factor in the extent to which an individual implements an intervention despite the disorganisation of the broader educational setting. However, it is clear that more research into the role of organisational capacity has in preventive intervention delivery is required.

The organisational capacity of impoverished schools and intervention implementation

Empirical evidence for both supporting and refuting the role of school organisational capacity has generally been sampled from educational settings serving deprived neighbourhoods. For example, Malloy and colleagues (2015) drew their sample of schools from deprived neighbourhoods known to have high residential mobility when investigating the role of organisational capacity in intervention implementation. This piece of research does not provide insight into whether there are different levels of implementation between schools serving affluent residential areas and schools serving high proportions of children experiencing poverty. Achieving such a comparison would require sub-group analyses within an adequately powered RCT that has large enough numbers of schools from both impoverished and affluent areas. To date, there are no such studies, as the number of schools is often unfeasible to recruit and train in delivering the intervention during effectiveness trials. However, with large RCTs becoming more commonplace within the field of education, exploratory analyses may provide insight into poverty as a factor affecting implementation. This lack of research, therefore, strengthens the need for the present study to begin addressing this gap within the literature.

4.5 Empirical evidence on the implementation of the Good Behaviour Game

Although there is a substantial evidence base for the GBG spanning over fifty years, there has not been much investigation on the implementation of the programme. Due to the lack of evidence, the present section includes the literature concerning both implementation variability and the factors associated with implementation. As observed within the wider literature about preventive interventions, empirical evidence on the different aspects of implementation influencing the behavioural outcomes of the GBG are not equally

represented. Due to the success of the intervention, the GBG has been adapted to fit different contexts. Therefore this dimension is reviewed first, with a dedicated section. The remaining aspects of implementation are then be reviewed together due to the thin empirical base. There has been more interest in the factors affecting the implementation of the GBG. However empirical evidence appears to be focused on the coaching model attached to the intervention.

4.5.1 Adaptations of the Good Behaviour Game

AIR own the license for the version of the GBG utilised in the present study (see section 3.4). However, two prominent versions have been produced by other developers/researchers outside of the institute in order to address local needs. Embry (2002) developed *PAX GBG* to ensure teachers had greater autonomy over the rules set, and that children did not perceive the recording of infractions as a negative experience for the children. Meanwhile, *Taakspel* is a Dutch version of the GBG which has been tested in the Netherlands (Van Lier, Van der Sar, Muthé & Crijnen, 2004) and Belgium (Leflot, Van Lier, Onghena & Coplin, 2013). The most prominent difference between *Taakspel* and the American GBGs is the way in the teacher monitors infractions, where cards are used in place of an infractions chart (see Table 4.3). This modification is thought to ensure the child(ren) responsible for the rule-breaking are not identified to the rest of the team or class (van Lier, 2002). Meanwhile, the GBG differs from the adapted versions by having established rules devised by the developers (see section 3.4.1), whereas teachers and children are encouraged to develop rules to suit the needs of their class. See table 4.3 for a summary of the implementation differences between the three versions of the GBG.

Table 4.3
Contrasting fidelity requirements for the GBG, PAX GBG and Taakspel.

Component	PAX GBG	AIR GBG	Taakspel
Rules	Children are invited to choose the rules at the start of the academic year.	The rules are predetermined as part of the intervention.	Teacher and class agree on the rules at the start of the academic year.
Dosage	Three games per day.	Increasing from three games a week, to daily.	Three games per week.
Winning	Four or fewer infractions.	Four or fewer infractions.	At least one team card is left on the table.
Rewards	Are always intrinsic rewards and activity-based.	A stamp in the GBG booklet, a team star on the weekly winners' chart, plus a reward chosen by the implementer.	A mixture of intrinsic and extrinsic rewards. Compliments are received for the daily prize, while weekly and monthly rewards are chosen by the implementer.
Language	Uses the terms "Spleem" and "Paxis" in order to prevent negative	The GBG implementer has a script to follow; this script neutrally directs behaviour when a rule is broken. It is not a punishment.	Teachers do not verbally indicate that an infraction has taken place. Instead, teams have a card removed from their team table.

emotions when
rules are broken.

Further variations of the Good Behaviour Game

With the original version of GBG developed approximately 50 years ago, it is unsurprising that many variations have been developed to suit different cultures and contexts. Modified versions have been applied in throughout different stages of schooling such as preschool (e.g. Sweizy, Matson & Box, 1993) and high school (e.g. Lynne, 2015), as well as outside of the classroom (e.g. after-school clubs, Hynes, Phillips Smith & Perkins., 2009). A version of GBG has been tested within a clinical population of children (e.g. Breeman et al., 2015), as well as being combined with other UPIs such as Say-Do-Report (Ruiz-Olivares et al., 2010) and PATHS (Domitrovich et al., 2010).

4.5.2 Evidence for the remaining aspects of implementation moderating behavioural outcomes

Despite the considerable evidence base for the GBG, the literature base details very few instances where implementation data has been collected on the intervention. When researchers have reported implementation data, the primary use has been descriptive. For example, Dion and colleagues (2011) reported that alongside the French version of Peer-Assisted Learning Strategies (PALS; Mathes, 1998), the observed fidelity of GBG was 92%. Meanwhile, Ialongo, Poduska, Wherthamer & Kellam (2001) reported an average implementation of GBG of 59.9%. The lower levels of implementation reported by Ialongo and colleagues may be due to the inclusion of dimensions such as intervention preparedness. However, Ialongo and colleagues (1999) is the only study to date to use implementation data for explanatory purposes. The researchers found that higher levels of fidelity were associated with a greater impact on behavioural and academic outcomes. It is important to note that the teachers implemented GBG alongside an academic intervention, and therefore caution should be taken when interpreting the extent to which it affected the academic outcomes. The present study is, therefore, an opportunity to further extend this work by assessing multiple aspects of implementation in relation to behavioural outcomes (specifically fidelity, quality and dosage). By contributing to the evidence base in this way, the present study works towards to the standards expected of empirical evidence on implementation of preventive interventions by experts in the field (Gottfredson et al., 2015).

4.5.3 Factors affecting the implementation of the Good Behaviour Game

Although there has been not much investigation into whether implementation variability moderates the behavioural outcomes of the GBG, there *has* been an interest in identifying the factors associated with implementation variability. The dominant reason for this is that the programme has a coaching model, as evidence indicates that coaching improves

implementation levels such as fidelity, dosage, and quality (Rohrbach, Gunning, Sun & Sussman, 2010; Abry, Rimm-Kaufman, Larsen & Brewer, 2013). For instance, Becker and colleagues (2013) reported that regular contact with coaches improved teachers' quality of implementation of the GBG. Wheby and colleagues (2012) suggest that that the coach may also indirectly improve implementation levels of the GBG by mitigating the effects of teacher burnout. The support of a coach may serve a similar function to the support provided by the administration or co-workers (Wheby et al., 2012), allowing the teacher to reflect and problem solve in an environment which usually is unable to accommodate such practices. Therefore, preventive interventions with a built-in supportive component may be suitable for schools with low organisational capacity as coaches could provide the extra social capital required to invest in sustainably implementing new practices.

Domitrovich and colleagues (2015) conducted a more holistic study investigating teacher and school level factors affecting the implementation of PAX GBG. Teachers reported emotional exhaustion and acceptance of the intervention were significantly related to dosage levels. This meant that teachers are experiencing higher levels of emotional exhaustion if they played fewer games, while teachers who felt the intervention complemented their teaching style were more likely to play GBG more frequently. Domitrovich and colleagues also included school poverty in their analyses but reported that it was not significantly related to implementation. However, the researchers also acknowledge that the sample was underpowered and drawn from an urban district. This methodological issue meant that there was little variability in the poverty levels of the schools within the sample, so researchers could not draw comparisons between disadvantaged schools and those serving more affluent communities. This strengthens the need for the present study in order to determine how factors affecting implementation operate across different schooling contexts.

4.6 Rationale

Although empirical evidence suggests that GBG is an effective UPI for reducing DB in the classroom, there remain significant questions about the role of poverty in moderating its implementation and intervention effects. Despite many tests of GBG internationally, no studies have robustly examined the role of school and individual level poverty.

Research into interactions between poverty at different levels of the school structure are essential to better understand variability in intervention outcomes. Equally important is the interaction between poverty and implementation variability and the influence this relationship has on programme outcomes. As socio-economic deprivation is located in clusters, schools serving these areas will have higher proportions of children experiencing poverty. Therefore, such school are likely to have classroom environments which, in theory, impact on the delivery of interventions. As poverty is nested (i.e. individuals experiencing

poverty attend schools with higher FSM uptake), the present study utilises appropriate statistical techniques to infer interactions between the variables of interest. This method will further our understanding from the very little research that has been conducted in this area. By gaining insight into the impact both child-level poverty and school-level poverty can have for both the implementation and outcomes of a preventive intervention, the suitability of UPIs such as the GBG can be determined for schools serving impoverished neighbourhoods, as well as the children experiencing poverty that participate in such programmes.

4.6.1 Contributions to knowledge

The present study intends to advance the knowledge within the area of prevention science through the achievement of the following objectives:

- Develop an understanding of poverty as a moderator of the behavioural outcomes in the GBG.
- Establish the use of cluster analysis as an appropriate means of developing implementation profiles for teachers delivering UPIs.
- Determine whether implementation profiles of the GBG vary as a function of school-level poverty.
- Establish the extent to which school-level poverty, teacher implementation profile, and individual-level poverty interact to influence the GBG's effect on DB.

4.6.2 Research questions

Based on the synthesis of the literature covered in the chapters so far, the researcher of the present study proposes two research questions.

Research question one

Does poverty moderate the effects of the GBG on children's DB?
- e) Does the GBG reduce children's DB?
- f) Are there differential effects of the GBG among children who are eligible for FSM?
- g) Does school-level poverty moderate the effects of the GBG on DB?
- h) Do school-level poverty and child-level FSM eligibility interact to moderate the effects of the GBG on DB?

Although there is a substantial amount of literature that suggests that the GBG is effective in reducing DB, there is negligible evidence that has explicitly investigated whether these reductions will be differentiated by a child's socio-economic disadvantage (Flower et al., 2014). However, the present study predicts that a school's FSM uptake will play a significant role in moderating the GBG's effectiveness in reducing DB; in that high FSM uptake will lead to a reduced intervention effect. This is because areas of high deprivation are likely to have larger groups of children displaying maladaptive behaviour that could be a stronger

indicator of status than the GBG, a risk acknowledged by Kellam and colleagues (1998) when proposing SFT.

Research question two

Does poverty moderate the implementation of GBG?

    f)   Are there distinct implementation profiles amongst teachers implementing GBG?

    g)   Does implementation vary as a function of school-level poverty?

    h)   Does implementation variability moderate the effects of the GBG on DB?

    i)   Do school-level poverty and implementation variability interact to moderate the effects of the GBG on disruptive behaviour?

    j)   Do school-level poverty, implementation variability and individual-level poverty interact to moderate the effects of the GBG on DB?

Based on the principles of SDT (Shaw, McKay & Hayner, 1942), it is expected that schools with a high percentage of FSM eligible children are unable to dedicate the adequate time and resources to intervention delivery and will, therefore, be less likely to implement GBG to optimal levels of fidelity and dosage than their counterparts in more affluent schools. Furthermore, the researcher expects to find that children in classrooms with higher levels of implementation will experience significantly greater reductions in DB compared to their counterparts in low implementing classrooms. RQ2 is primarily an exploratory question, as there is very little substantive evidence to base hypotheses on the interactions between poverty (at both the school and individual level) and implementation variability with regards to the moderation of the GBG's intervention effects.

4.7 Chapter summary

The chapter introduced the concept of implementation by defining and exploring the construct's subordinate dimensions. Then, the literature review examined the extent to which the dimensions moderated interventions outcomes. The researcher also considered factors that influence the implementation of these. Afterwards, the chapter introduces SDT as part of the theoretical framework of the present study. It justified the role poverty may have in the implementation of preventive interventions such as the GBG. Following from this the reseacher highlighted the lack of empirical evidence that specifically investigated the implementation variability of the GBG. Finally, the chapter presented, the rationale, contributions to knowledge and proposed research questions.

Chapter 5
Methodology

5.1 Introduction to chapter

The current chapter describes, critiques and justifies the methodology regarding its appropriateness to answer the research questions presented in the previous chapter. The first section presents the context of the current study by highlighting the critical divergences from the GBG evaluation. The second section discusses the epistemological and theoretical stances adopted for the present study. It then goes into detail about how a post-positivist perspective impacts the methodology and methods used, as well as highlighting the strengths and limitations of adopting such a stance. The third section justifies the use of the experimental design for the study. Randomised control trials (RCTs) are still considered a contentious design within education research. Therefore, the researcher discusses the appropriateness of the RCT for the present study in detail. The fourth section offers details about the participants. The section includes the recruitment of schools, the process of random allocation to the control or intervention condition, the optimal and final participant numbers, as well as power and sample size calculations required to answer each research question. The penultimate section discusses the selection and development of instruments used for data collection, as well as a justification for the proxies used for child level and school level poverty. Also presented are details regarding the protocol for data collection, validation, and analysis. Finally, the ethical considerations required for the study are outlined.

5.2 Context of the present study

The present study uses data collected from an evaluation study conducted by a research team at the University of Manchester in collaboration with Mentor UK. The Education Endowment Foundation (EEF) commissioned the project. The project evaluated the GBG, a classroom-based preventive intervention for disruptive behaviour. After the present section describes the evaluation project and how it was conducted, the present study is compared to the evaluation study in order to justify it as an original and distinct piece of research from the evaluation project.

5.2.1 Good Behaviour Game evaluation study

The primary aim of the evaluation was to assess the impact of the GBG on academic and behavioural outcomes for pupils. The second aim was to assess the impact of the GBG on outcomes for teachers, while the third aim was to assess the implementation practices of teachers when delivering the GBG. The first aim was measured using quantitative methods. It assessed the impact of the GBG on both pupils' reading scores and behavioural outcomes

such as concentration, disruptive behaviour, and pro-social behaviour. While the evaluation project measured outcome data on all participating children, the project also looked at the impact the GBG had on the outcomes for children eligible for FSM. Although the present study also focuses on the impact the GBG has on the behavioural outcomes for children experiencing poverty; the evaluation project does not investigate the behavioural outcomes in as much depth.

The second aim also used quantitative methods. It assessed the impact of the GBG on outcomes related to teachers. They were: a) efficacy in classroom management, b) classroom stress, and c) retention. The final aim used both quantitative and qualitative methods to assess the implementation practices of teachers, regarding a) fidelity, b) dosage, c) quality, d) participant responsiveness, e) reach, f) programme differentiation and g) adaptations. The primary and secondary aims were measured using teacher-report surveys as well as through The National Pupil Database (NPD). The final aim was measured using structured observations, as well a collecting qualitative data from 6 case study schools that volunteered to participate in semi-structured interviews. Researchers gathered from teachers implementing the GBG, senior leadership teachers, children, parents, and the coaches provided by MentorUK.

The differences between the present study and the Good Behaviour Game evaluation project

Although data utilised in the present study derived from the GBG evaluation, there are distinct differences regarding aims, methods and analyses. Thus, the present study is an original and valid piece of research that is distinct from the evaluation project. Table 5.1 presents the differences between the two pieces of research.

The present study's aims have a clear focus on poverty and its different forms (i.e. individual-level and school-level) as possible moderators for the main intervention effect. Meanwhile, the GBG evaluation project has prioritised the intervention's effectiveness for all children in two distinct outcomes. Though both studies investigate implementation practice, the evaluation project uses both qualitative and quantitative methods to assess seven constructs of implementation. However, the present study utilises quantitative data to assess teacher implementation in conjunction with school-wide poverty in order to provide apply SDT to educational settings. A final distinction is that the present study utilises a 'person-focused' approach to analysing implementation data (e.g. cluster analysis), which contrasts with the variable-focused approach taken in the evaluation study.

Table 5.1

An overview of how the present study differs from the GBG evaluation project.

	The present study	The GBG evaluation study
Overview	• An individual 3-year project for a doctoral thesis.	• A 2-year project run by a research team.
Aims	• To investigate the roles of individual poverty and school-wide poverty as moderators of behavioural outcomes. • Determine the extent school-wide poverty moderates the extent to which teachers implement the GBG. • Comment on both theoretical and practical implications in order to contribute to literature regarding theory and practice.	• To assess the effectiveness of the GBG on pupils' academic and behavioural outcomes. • To assess the implementation practices of teachers when delivering the GBG. • A larger concern was placed on practical implications as opposed to theoretical ones.
Methods & analysis	• Used cluster analysis to explore teacher implementation profiles using variables measuring fidelity, quality, and dosage. • Used Multilevel modelling (MLM) to run planned interactions between school-level poverty and individual poverty.	• Used a mixed-methods approach to assess the implementation practices of teachers. • Assessing the intervention impact on pupil behavioural and academic outcomes, as well as teacher outcomes relating to stress and retention.
Target audience	• The academic community.	• The EEF, schools interested in preventive interventions, and the academic community.

5.3 Methodology

This section discusses the epistemological standpoint of Objectivism and how it influences both the theoretical perspective and the methodological decisions applied to the study.

5.3.1 Epistemology

Epistemology is the study of knowledge acquisition; it is concerned with the sources, nature, and the limitations of knowledge (Crotty, 1998). There are multiple epistemological stances, each shaping theoretical and methodological decisions in research. While everyone implicitly subscribes to an epistemological perspective, researchers are expected to explicitly adopt and explore a stance that best aligns with his or her ontological view (Hussain, Elyas, & Nasseef, 2013). Ontological stances influence epistemological perspectives (Grix, 2004). Ontology is the study of existence and reality (Stevenson, 2010). While there are many ontological perspectives, the dominant two are ontological materialism and idealism. Ontological materialism argues that reality is independent of the human mind, with the interactions between material objects being responsible for consciousness (Garza & Fisher Smith, 2009).

Epistemological stances such as objectivism have formed from ontological materialism in order to argue that individuals can gain knowledge by observing the independent reality from our senses, also known as empiricism (Grix, 2004). Meanwhile, ontological idealism maintains that reality is not independent of our perception and that consciousness responsible for the creation of the material objects we interact within our everyday lives (Jones, 1893). Epistemological stances such as constructivism have developed from ontological idealism that individuals create knowledge through mental constructions in order to explain sensory experiences (Crotty, 1998). While researchers in psychological and educational research use both epistemologies, Objectivism influences the present study. The researcher believes that by using objective methods to study the world and the people in it, theoretical models can be created to represent and explain reality (Lakoff, 1987). The models and theories developed in an attempt to explain how the nature of the world and society works to contribute to knowledge but also progress and improve society (Burrell & Morgan, 1979).

5.3.2 Theoretical perspective

A theoretical perspective is a set of philosophical principles that inform the types of research that are conducted by a researcher; it further builds on the assumptions laid out by the epistemological stance one subscribes to (Kuhn, 1970). In education, there are four dominant paradigms to which researchers ascribe. The transformative perspective maintains that research should be conducted using a political lens in order to address social injustice (Creswell, 2013). Constructivism feeds and argues that research should use social constructions such as language in order to generate knowledge and answer research questions (Hamilton, 1974).

However, pragmatism rejects the ontological and epistemological dichotomy presented by objectivism and constructionism by arguing that there is an external reality, but it is not static because ones' actions can influence it (Dewey, 1938). Therefore, the perspective favours the acquisition of knowledge by using the best-suited methodology and methods to answer the research question (Crotty, 1998). The final perspective, post-positivism, is driven by objectivism and is the theoretical perspective influencing the methodology and methods of the present study. The researcher has taken this theoretical stance over other stances because post-positivism acknowledges that research and the construction of knowledge occur within a context while striving for objectivity and rigour (Crotty, 1998).

Post-positivism

The reflection of Objectivism within post-positivism can be demonstrated through the perspectives' principle assumption: that there is a definitive reality. While this perspective succeeded in the positivist paradigm, the shift refutes some of the core assumptions of positivism. Positivism was deemed to be the scientific method of studying reality with the research being deemed "pure" and value-free as data revolved around measurable and

observable facts (Friedman, 1999). However, post-positivism argues that scientific reason is similar to common-sense reasoning, in that researchers are not rational or truly objective and therefore observations about the world and any conclusions from such research are fallible (Popper, 1972). A key assumption of the post-positivist perspective maintains that a true understanding of the reality that is independent of the human mind is impossible and that a researcher should scrutinise and acknowledge the limitations of the knowledge and research generated.

Post-positivists, therefore, attempt to "approximate the truth rather than aspiring to grasp it in its totality or essence" (Crotty, 1998, p. 29). By using more objective methods and acknowledging that error is an important factor in research, positivists report probability values and confidence intervals in order to convey the extent to which research can be considered accurate. This concept also relates to another cornerstone of post-positivism: falsifiability. Theories about how the independent reality operates cannot be confirmed, as research can only strive to be objective, the margin for error always allows for new evidence to appear and disprove an existing theoretical model, opening the way to new models that are closer to the absolute truth (Popper, 1963). Falsifiability is, therefore, a continuous and cyclical process as the academic community refute or revise theories based on multiple accounts of research that has some margin of error. The post-positivist perspective's assumptions influence the present study by acknowledging that the analysis of the data collected could only be used to determine the probability that the findings were accurate rather than absolute. This assumption is because the methodology and methods of the present study contained a degree of error and bias and that it is the researcher's responsibility to be critical of the limitations of the methodological decisions as well as their own biases during the conducting and reporting of the research. Finally, the favouring of objectivity over subjectivity restricted the designs and methods that were suitable for a study influenced by post-positivism, and therefore quantitative approach was adopted.

5.4 Experimental design

The preferred design used by post-positivist researchers is experimental design. This design choice is because post-positivists can establish causation by manipulating variables that have been hypothesised to have an effect on a certain outcome and measuring dependent variables using objective means. Experimental designs are regarded to be more reliable and valid compared to other designs. Thus they have obtained a certain status. However, the randomised control trial (RCT) has been considered the "gold standard" across many disciplines as it removes selection bias between groups of participants (Torgerson & Torgerson, 2001; Wells et al., 2012; Farrington, 2003). RCTs herald from the era of enlightenment where positivism was the dominant paradigm in research. They became the standard design utilised in the medical discipline. To this day, RCTs are primarily used to test whether treatments, particularly pharmaceutical drugs, reduce

symptoms. A vital feature of the RCT is that recruited participants have an equal chance to be assigned to either the treatment or control condition. Random assignment reduces any biases during the allocation of participants to conditions or any researcher biases. It also distributes any confounding variables equally between the two conditions (Torgerson and Torgerson, 2001). Though in theory, randomisation should create balanced groups of participants; this is not always the case. In large RCTs, such as the present study, small but significant differences between the groups of participants can affect the small intervention effect (Treasure & MacRae, 1998).

The present study was at risk of randomly assigning schools so that one condition could have had a higher proportion of children eligible for FSM compared to the other. It would be impossible to attribute outcomes to the GBG as children eligible for FSM are a key group at risk of displaying disruptive behaviour. However, to counteract this risk, during the randomisation process a minimisation algorithm can be applied to the participant pool. Minimisation is a process whereby a participants' group allocation is not solely determined by chance; the process reduces the difference in the distribution of suspected factors that may influence the intervention effect (Torgerson & Torgerson, 2013). The result of minimisation is two balanced conditions in the number of participants that display the suspected factor, in the case of the present study, the proportion of FSM eligibility in each condition. While the data from the present study was reliant on the RCT design in the GBG evaluation (see 'Context of the present study'), the researcher adopted the design for their research. The researcher deemed that the use of randomisation and minimisation to balance confounding variables across the conditions would add to the reliability and validity of the present study, two cornerstones of post-positivist research. The RCT was also deemed the most appropriate design to answer the research questions, as their answering requires establishing causation through the manipulation of variables.

5.4.1 The quality of randomised control trials

RCTs can only be given the title of "gold standard" if they comply with high standards of methodological rigour. Like all forms of research, a poorly designed RCT can harbour biases, which will undermine the validity of the findings from a study (Schulz, Altman, & Moher, 2010). Determining the quality of an RCT requires the taking into consideration many facets of the research process, this would include: the design, procedure, analysis, and reporting (Jüni, Altman, & Egger, 2001). However, the quality of RCTs is closely linked to validity, which occurs on two levels: internal validity and external validity. Internal validity is concerned with the extent to which the independent variable, e.g. the intervention, is the cause of any differences between patients in the control and experimental conditions (Howitt & Kramer, 2005). Meanwhile, external validity is related to generalisability. While it is possible for an RCT to have strong internal validity and weak external validity, a study must have internal validity to achieve external validity (Jüni, Altman, & Egger, 2001). While it is often assumed that researchers must strive for both internal and external validity in order to

produce a high-quality RCT, within prevention research, the academic community usually achieves internal and external validity in separate studies. Following post-positive principles brings an understanding that knowing the true values of an RCT's outcome is impossible, however with multiple measurements through multiple studies, researchers can get close to the truth (see theoretical perspective for more on post-positive principles).

Prevention science must determine whether an intervention works before even asking questions about generalizability. To achieve this, researchers test intervention viability using different forms of RCT. Efficacy trials are designed to have high internal validity in order to determine whether the intervention caused a difference between the two conditions. The trial achieves this by testing how well an intervention would work under optimal conditions. Usually, the developers of the intervention would be involved in the trial in order to help implementers achieve appropriate levels of fidelity, quality, and dosage (Dane & Schiender, 1998). If the experiment produces a difference that researchers attribute to the intervention, then it would be tested in an effectiveness trial, i.e. under "real world" conditions. This type of trial would limit the involvement of the developers in assisting in the implementation of the intervention (Flay, 1986).

Nevertheless, there are issues with using efficacy and effectiveness trials as benchmarks for RCT quality. Achieving high internal validity by conducting an efficacy trial in a school context is associated with barriers that may mean "optimal" intervention conditions may never be reached. Traditionally, efficacy studies are used to test medical interventions, where researchers admit their participants into their care (e.g. staying in hospital facilities) in order to ensure the implementation of the intervention meets the developer's requirements (Streiner & Norman, 2009). It is inconvenient on so many levels to apply such a model when testing classroom-based preventive interventions; instead, the developers and testers have to bring the optimal conditions to participants. However, this approach also has its challenges. Developer and researcher involvement is likely to be less than optimal as schools are unlikely to keep to commitments and appointments because of within-school obligations. These issues are unlikely to threaten the quality of the present study, as the researcher has formulated research questions that focus on generalisation, which requires high external validity. However, as explained earlier, internal validity is also vital, as, without it, there cannot be external validity. To ensure the present study is of high quality, methodological decisions such as adopting a cluster randomised design to increase internal validity and careful selection of instruments to measure outcomes to ensure external validity.

5.4.2 Randomised control trials in education research

Despite Interpretivism being the dominant paradigm in education research, the use of RCTs in this field pre-date their use in medicine (Torgerson & Torgerson, 2001; Oakley, 1998). However, in recent years, the use of RCTs have increased in the field of education with

western governments developing policies around the notion of "evidence-based practice". Thus, providing funding opportunities for large-scale RCTs for education-based interventions in order to see if they "work". This provision of funding is particularly applicable to the UK context, as the current government seems to be following the US model by creating initiatives that allow educational settings to access and interpret the findings on the impact of the interventions. For instance, the UK government have developed a "what works network" where different centres are responsible for funding and reporting RCTs in order to improve policy level decision-making (Cabinet Office, 2015).

The Education Endowment Foundation (EEF) is The What Works Centre for education and is responsible for funding interventions that aim to improve educational outcomes for children experiencing poverty. The organisation has funded 145 RCTs since its inception in 2011 (EEF, 2018). Schools can access a toolkit on the EEF website which contains the evaluation reports of the trials that have been completed so far. The EEF rates each evaluation for robustness. Providing such information allows senior leaders in educational settings to assess whether an intervention matches with their school's ethos and whether investing money into bespoke programmes will see a return through the improvement in academic outcomes. With school governing bodies having more freedom than ever over how they can spend the school budget, the use of RCTs in education research is imperative, as there is a moral duty to test whether an intervention is impactful before schools spend money on a bespoke programme.

Interprevitist researchers have heavily criticised the use of RCTs in education; the majority of reasons being irreconcilable differences in epistemological viewpoints on what constitutes as knowledge, and how knowledge "should" be collected and used (Oakley, 2006). For instance, Biesta (2007) argued that RCTs had no place in education research, as they could not measure the symbolic interactions between teachers and pupils that occur in the use of interventions. Meanwhile, other interpretivists maintain that the use of RCTs is part of a movement to restrict education practitioners' choice and autonomy (Ball, 2001; Atkinson 2000; Stronrach, 2004). Finally, some interpretivists consider the use of RCTs to be unethical as some participants are denied access to the intervention if they are randomised to the control condition (Morrison, 2001; Bridges, 1998; Atkinson; 2000). While Biesta (2007) raises a valid point, in that RCTs cannot measure symbolic interactions between the teacher and pupil, the present study is interested in the impact of the GBG in schools with pupil experiencing medium and high poverty, not the relationships between the pupil and the teacher. Though research does not deny that the rapport a teacher has with their pupils is likely to affect behavioural outcomes (Cooper, 2011; Lang 2013; Roache & Lewis; 2011), the random allocation of participants evenly distributes the presence of both positive and less effective teacher-pupil relationships (Torgerson & Torgerson, 2001).

114

Although on the surface of the argument, denying schools access to an intervention may seem unethical; this would only be the case if an intervention were already known to be impactful. RCTs are used to determine whether interventions bring about positive change in a targeted symptom or behaviour, and the most accurate way of measuring this is to compare any differences between participants who receive the intervention to those who do not. Also, it would be unwise to be presumptuous that participants assigned to the intervention condition would be at an unfair advantage. Many trials within the field of education find null results (Torgerson et al., 2016; Hanley et al., 2016; King & Kasim, 2015; Sheard, Chambers, & Elliott, 2015). It is also important to stress that the effectiveness of an intervention cannot be determined from one RCT, an intervention must be successful in a range of RCTs before it can be considered as effective (see 'The quality of randomised control trials). This standard of effectiveness is essential for the present study. Although there have been many testings of the GBG within the US, the educational system and classroom is culture is different to the UK. Therefore, one can not assume that children in the intervention condition will have an advantage over their control group counterparts, as there is no empirical evidence to suggest there would be any behavioural gains implementing the GBG in a UK classroom.

Finally, many of the researchers resistant to the use of RCTs in educational settings fail to acknowledge that the design's origins are rooted in education- the earliest reported study was published in 1928 (Forestlund, Chalmers, & Bjorndal, 2007). Therefore, it is not a 'foreign' design taken from 'harder' disciplines that are invading education research. The research community abandoned the RCT as a large number of null or negative results were produced (MacDonald, 1997). Instead, the research community should regard the rediscovery of RCTs within the field of education as a welcome addition to the discipline. Especially as the design addresses issues and research topics that interpretivist designs are unable to cover.

5.4.3 A cluster-randomised control trial

Traditional RCTs use individual participants as the unit of randomisation. However, in educational settings, randomising individual students is not always feasible. Impractical measures would have to be implemented by the school to prevent intervention effects leaking between children and teachers; even then intervention leakage could still occur (Puffer, Torgerson, & Watson, 2003). This need to prevent contamination between the control and intervention conditions has seen the increased use of the cluster-randomised control trial design (C-RCT) where researchers randomise by groups of participants as opposed to individuals. C-RCTs lend themselves well to educational settings in this respect, as schools and classrooms are naturally occurring clusters that can be easily randomised.

However, C-RCTs are at risk of being less robust compared to individual-level RCTs as there may be similarities between participants in the different clusters (Torgerson, 2001).

These potential similarities between participants imply that standard sample size calculations cannot be applied to C-RCTs as individuals are likely to correlate with one another (this is the intra-class correlation, ICC). This 'design effect' means a C-RCT requires a larger sample size than a traditional RCT to have the same statistical power in detecting intervention effects (Puffer et al., 2003). With this in mind, the literature advises that C-RCTs should have at least 5-8 clusters per condition in order to account for ICC and increase the robustness of C-RCTs (Medical Research Council, 2002; Torgerson, Torgerson & Styles, 2016). Another drawback of randomising at school level is that the increased chance of attrition bias, which is when there are systematic differences between participants withdrawing from a trial and those who continue (Giraudeau & Ravaud, 2009). Attrition is problematic when entire clusters withdraw from studies as a large number of participants will not be included in the analysis, particularly if attrition is higher in the treatment condition as the intervention will seem more effective than it is (Hahn et al. 2005). To minimise attrition bias, researchers can use intention to treat analysis (ITT). IIT is where participants are kept in the study and partake in the data collection process despite no longer receiving the intervention.

Despite the robustness of individual-level randomisation, the impracticalities placed upon participating schools would outweigh the statistical benefits of traditional RCTs. The same point can be extended to randomising at the classroom level, as the risk of contamination between pupils where schools could have control and intervention classes would compromise the validity of the study (Torgerson, Torgerson, & Styles, 2013). Individual or classroom level randomisation is also inappropriate when conducting a sub-group analysis, which is part of the protocol for the present study. As there is a focus on children that are eligible for FSM, particularly comparing their disruptive behaviours within the context of school poverty, randomising children or classes to the intervention condition would render sub-group analysis useless. Therefore, randomisation at the school level was deemed to be the most appropriate form of C-RCT for the present study. By using measures such as using intention to treat analysis and having an appropriate number of clusters in each condition were used to improve the robustness of the design.

5.5 Participants and their characteristics

In this section, details of how schools were recruited to the present study are presented here. Following this, the sample characteristics at the school, teacher and pupil level are then detailed.

5.5.1 Recruitment

In the summer term before the start of the trial, participants involved in the present study were sampled through the GBG evaluation (see sections 1.4 and 5.2). Researchers used

convenience sampling to enlist state-maintained schools through recruitment events held in partnership with the research team and the implementing team, Mentor UK. All primary schools in the areas of Greater Manchester, the Midlands, and parts of Yorkshire were invited to these events by their local learning authority. The following topics were covered in these recruiting events: The nature of the GBG, the layout of the RCT trial, the commitment to data collection and school visits from both the research team and Mentor UK, and the benefits of participating as either a control school or GBG school. The researchers then followed up with the headteachers that attended the events with phone calls and emails to establish whether the school was interested in participating in the RCT.

Before schools could be confirmed as participating in the study, senior members of the school had to sign a memorandum of agreement to demonstrate their commitment to the terms and conditions of participating in the wider evaluation and attached studies, such as the present study. The memorandum was important as schools were randomly allocated to either the control or intervention condition after consenting to participation, as it ensured that schools would continue to participate if they were randomised to a less desirable condition. As explained in section 5.4.3, minimisation was used in the randomisation process to ensure an equal representation as possible of schools in each condition, and so it was necessary to know which schools would be participating before randomising in order to use minimisation. The head teachers received a phone call and a letter from the research team confirming the condition in which the school had been randomised.

5.5.2 school-level characteristics

In total, 77 schools participated in the present study, with 38 randomised to the intervention condition and 39 in the usual practice condition. School-level demographic data was collected using the national pupil database (NPD) request that was made by wider evaluator study (see section 5.2.1). As depicted in table 5.2, schools participating in the present study were more likely to serve a larger population of children compared to the national average. However, the SD suggests that school sizes had a high degree of variability. The variability is likely to be explained by the number of forms per year group a school can hold. For instance, a one-form entry primary school has six classes that make up the pupil population, whereas a 3-form entry primary school comprises of 18 classes. If the average class size contains 27.4 children, then the average difference in pupil numbers between a 1-form and 3-form school is 328.8 (DfE, 2016a).

Table 5.2
Mean school-level characteristics of the sample presented by condition and total against the national average.

School characteristics	GBG (*n*= 38)		UP (*n*=39)		Total (*n*=77)		National average[2,3,4]
	M	SD	M	SD	M	SD	
School size	355.36	168.44	366.53	202.53	359.72	186.39	275
% FSM uptake	27.32	12.50	25.40	12.81	26.32	12.68	14.5

% White British	59.41	33.98	61.63	32.83	60.59	33.38	67.9
% EAL	30.30	32.44	27.80	29.79	29.02	31.12	20.1
% Absence	4.28	.86	4.32	1.01	4.30	.93	4.0
% SEND	22.06	9.91	17.78	5.17	20.20	8.27	13.4
% Achievement[1]	74.27	11.50	74.29	10.80	75.08	11.19	80.0

[1]% of children achieving level 4 or above in English and Maths.
[2]National averages for school size, FSM uptake, language group, and ethnicity were taken from DfE (2016a).
[3]SEND national averages taken from DfE (2016b).
[4]Pupil absence and achievement national averages taken from DfE (2016c).

However, participating schools also had higher numbers of pupils eligible for FSM, children speaking EAL, and children with SEND compared to the national average. With higher than average numbers of children considered to be at risk, it is unsurprising that schools in both conditions had lower than the average number of children reaching achievement standards set by the government. While the sample is not representative of primary schools in the UK, the schools that participated in the present study are the target population that the intervention is designed to aid. Schools that have more favourable positions about the national averages depicted in table 5.2 are less likely to perceive a "need" for the GBG and could have been less likely to seek to participate in such a study.

5.5.3 Teacher characteristics

As the present study was conducted over a two-year period, the characteristics of all participating teachers (*N*= 118) are presented below. Teacher demographics were self-reported. Table 5.3 presents a disparity between males and female teachers; 22.8% of teachers that participated in the present study were male. Government reports have replicated this disproportion consistently for many years, the most up to date figure is 15.2% (DfE, 2016d). Although the DfE does not offer the average number of years a teacher will stay in the profession, descriptive statistics about the distribution of teachers according to their age could give some insight into the average number of years teaching experience the sample from the present study had.

Table 5.3
Teacher demographic descriptive statistics presented by trial condition and total for the sample.

Conditions	Gender		Years of teaching		
(*N*)	Males	Females	M	SD	Range
GBG (58)	15	43	8.32	7.36	35
UP (60)	12	48	6.88	8.01	39
Total (118)	27	91	7.63	7.68	39

The majority of primary school teachers in the UK (54.7%) are 30 to fifty years old. Based on the qualification requirements an individual must meet, the minimum age a teacher can begin practice is 21-22 years of age. This age restriction is because university level qualifications are requisite for the profession. With this in mind, the majority of primary

teachers in the UK may have been practising for 9-29 years. Therefore, small SD and low mean would suggest the majority of teachers would not be considered to be new to the profession but are also likely to be younger professionals.

5.5.4 Pupil-level characteristics

In total, 3085 pupils across the two conditions were involved in the present study. The demographic data provided below was obtained through the NPD request mentioned previously. Table 5.4 shows disparities between the two conditions with the national average. In the intervention group, the percentages of children classified as an ethnic minority or as having SEND provision was approximately 10% lower than the national average. Meanwhile, the control condition's figures were more representative of the wider population.

Table 5.4
Pupil-level demographic statistics for each trial condition with the national average for comparison.

Pupil-level characteristics	GBG (N=1560)	UP (N=1525)	National average[1,2]
Gender %			
Male	50.4	54.9	51.0
Female	49.6	45.1	49.0
FSM %			
Eligible	27.4	22.8	14.5
Not eligible	72.6	72.2	85.5
Ethnicity %			
White British	66.7	64.9	67.9
Other	23.2	35.1	32.1
EAL %			
EAL	26.2	29.5	20.1
EFL[3]	73.8	70.5	79.9
SEND %			
Receiving provision	22.9	17.9	13.4
Not eligible	76.9	82.0	86.6

[1]National averages for Gender, FSM uptake, language group, and ethnicity were taken from DfE (2016a).
[2]SEND national averages taken from DfE (2016b).
[3]English as a first language.

5.6 Methods & Instruments

As previously explained, the present study relied on the wider GBG evaluation for the data (see sections 1.4 and 5.2). This meant the researcher was restricted by the methods and instruments that were available as the majority of the methodological decisions surrounding the GBG evaluation had taken place before the researcher had joined the research team. When deciding on the methods and instruments for the present study, in some cases, the researcher had no choice but to use the only available option provided by the wider trial (e.g. disruptive behaviour). In such instances, it will be made known to the reader. In other cases, there were multiple methods collecting the same variable, and so the researcher had to decide which one was the most appropriate for the present study. Nevertheless, within

this section of the methodology the researcher details the strengths and weaknesses of the methods and instruments, as well as providing assurance that those chosen were credible and accurate.

5.6.1 Proxies chosen for poverty

As detailed in chapter two (see section 2.3.2), relative poverty was selected as the defition for poverty within the present study. More specifically, sixty percent of the median, the official poverty line used by the UK government (DWP, 2017). The primary reason for using this definition is because complements the proxies that were available to the researcher from the wider trial. This is because many of the legislations driving provision for those experiencing poverty, such as FSM, use the cut-off point to determine which households require financial support (DfE, 2015). As the present study is interested in poverty at both the school-level and the individual-level, the chosen proxies are presented in this section.

Pupil-level poverty

The two available measures of pupil-level poverty were FSM eligibility and the Income Deprivation Affecting Children Index (IDACI). FSM eligibility is categorical data that represents whether children attending school live in households that are either claiming benefits relating to a lack of/low income (excluding working tax credits) or claiming child tax credits and have a gross annual income of no more than £16,190 (Department for Education, 2016). This criteria is based on the relative definition of poverty the UK government uses to classify those who are in need of assistance because of their low income (see section 2.3.2). Meanwhile, IDACI scores reflect the proportion of children that live in poverty within a specific geographic area and are then assigned to the individual pupil based on the postcode of their home-address (Office for National Statistics, 2015).

The IDACI does have an advantage over FSM eligibility, it provides continuous data which reflects the nature of poverty, i.e. a continuum. Meanwhile, the categorical nature of FSM eligibility means that the researcher would not know the extent of poverty the children in receipt of this benefit is experiencing. However, this is also an issue for the IDACI too, as the measure is only capable of providing the *likelihood* a pupil was living in poverty, not an accurate reflection of the household's income situation. As FSM eligibility was based on the income of the participant's household (as parents/carers would have to provide proof to receive the benefit) rather than probability, the researcher decided that FSM eligibility would be the most appropriate proxy for pupil-level poverty.

While FSM eligibility is a more accurate measure than IDACI, the researcher can not garuntee that all pupils experiencing poverty will be perfectly represented within the sample. This is because parents and carers have to provide evidence of their need for the benefit to the school before pupils can receive the benefit after Year One (DfE, 2015). It is therefore

possible that there were particiapnts that have been categorised as not receiving the benefit but are experiencing poverty. However the weaknesses of the measure have to be viewed alongside their strengths. The major advantage of using FSM eligibility was that the eligibility status could determined for every participating pupil through the NPD request for the wider evaluation. This is a non-invasive and time-saving way that allowed the researcher to access sensitive data without increasing the data burdern of the participating schools. Therefore, with both the weaknesses and strengths of FSM eligibility considered along side alternative proxies available, FSM eligibility was chosen to represent pupil-level poverty.

School-level poverty

Through the NPD request, the wider trial only provided one source of data that could be used to determine the level of poverty at the school-level, this was the percentage of the school population that claims FSM. However, for the ease of analysing the data using MLM, the researcher grouped the schools into two categories based on the amount of poverty at the school-level. The schools were then plotted onto a histogram and a cut-off point of one standard deviation (SD= 27.2%) above the national average FSM uptake (mean= 14.5%; Department for Education, 2016) was generated (see appendix D). Schools with an FSM uptake below the cut-off point were placed in the "moderate poverty category", while schools with an FSM uptake above the threshold would be placed in the "high poverty category. Table 5.5 shows the number of schools in both the usual practice and intervention conditions that have been assigned to each poverty category. There are a balanced number of schools from each condition in the two categories.

Table 5.5
Descriptive statistics concerning the % of FSM uptake across for both the poverty statuses and conditions utilised in the present study.

Poverty status	GBG condition			UP condition			Total		
	(N)	Min. %[1]	Max. %[1]	(N)	Min. %[1]	Max. %[1]	(N)	Min. %[1]	Max. %[1]
Moderate poverty	(19)	1.0	27.1	(21)	1.9	26.8	(40)	1.0	27.1
High poverty	(19)	28.5	56.7	(18)	27.4	50.5	(37)	27.4	56.7
Total	(38)	1.0	56.7	(39)	1.9	50.5	(77)	1.0	56.7

[1]School FSM uptake.

The term "moderate poverty" was chosen over "low poverty" to describe schools under the cut-off point as many of the schools in this category were still above the national average FSM uptake. As demonstrated in table 5.5, there were schools in the moderate category that had an FSM uptake approximately 12% higher than the national average. Therefore, it

was deemed that using the term "low poverty" to describe these schools would be deemed to be inappropriate as some of these schools were technically serving impoverished areas.

5.6.2 Disruptive behaviour

The Teacher Observation of Children's Adaptation: Checklist (TOCA-C: Koth, Bradshaw, & Leaf, 2009) was the only measure that was available to measure the pupil's disruptive behaviour for the present study. This was because the wider evaluation's research team had to make instrument choices for the trial before the author of the present study joined the research team. Therefore this section is dedicated to justifying the suitability of the TOCA-C (Koth et al., 2009) for the current study inspite of it being a predetermined decision by the wider evulation.

The TOCA-C is a non-clinical teacher report tool used to assess pupils' classroom behaviour (Koth, Bradshaw, & Leaf, 2009). The measure consists of three subscales: concentration problems, disruptive behaviour, and pro-social behaviour. Only the disruptive subscale was utilised as the researcher deemed the other subscales not relevant to the present study. The nine items reflect aggressive, disobedient, and disruptive behaviours that are deemed to be inappropriate in a classroom setting (see table 5.6 for a list of the items). The rater responds for each child in the class by indicating their level of agreement on a six-point scale (Never/Rarely/Sometimes/Often/Very Often/Almost Always). A higher score indicates more disruptive behaviours in the classroom. The highest score an individual could receive is 54, while the lowest score would be 6. Teachers filled in the TOCA-C for each pupil in their class as part of an battery of questionnaires that could be accessed on a secure website. Teachers completed the questionnaire at three time points of the study: at baseline, at the end of the first year, and at the end of the trial.

Table 5.6
Items that comprise of the TOCA-C disruptive behaviour sub-scale.

TOCA-C: Disruptive behaviour sub-scale
Breaks rules
Does not get along
Harms others
Gets angry when provoked
Yells
Fights
Lies
Harms property
Teases classmates

The disruptive behaviour subscale of the TOCA-C was developed with the GBG's programme theory in mind, social field theory (SFT: Kellam et al., 1975). The interviewer version of the TOCA-C, the TOCA-R (Werthamer-Larsson, Kellam, & Wheeler, 1991) has been used in large-scale school-based preventive intervention trials in the past, such as Fast Track (Conduct Problems Prevention Research Group, 2002) and the GBG (Kellam et al., 1998). Koth, Bradshaw, and Leaf (2009) found very few quantitative differences between

the TOCA-R and the TOCA-C and suggested that the TOCA-C would also be suited for large-scale research. The disruptive behaviour subscale has been reported to have strong internal reliability ranging from .92 to .95 as well as having a consistent factor structure over time (Bradshaw, Waasdorp, & Leaf, 2012; Koth, Bradshaw, & Leaf, 2009). A Chronbach's alpha test was run using the basline TOCA-C data for the present study and found that internal reliability was not as high as previous reportings, but it was still high enough to be considered reliable (α= 0.80).

However, a disadvantage of the TOCA-C is that rater is not able to score behaviours in terms of intensity, as well as frequency. Some of the items in the disruptive behaviour subscale may not occur frequently enough for the rater award a high score, but when the behaviour is displayed; it may be particularly problematic in the classroom. A good example would be the item "harms others". However, one could extend this criticism to most classroom behaviour measures. Therefore, the TOCA-C can be deemed to be an adequate measure to be used in the present study as it is short, quick to fill in, and cost-effective (Koth, Bradshaw, & Leaf, 2009).

5.6.3 Implementation

As the wider GBG trial was conducting an implementation process evaluation alongisde the RCT, there was a plethora of data sources the researcher of the present study could have utilised for the present study (for further details about this aspect of the wider evaluation, see: 1.4). Futhermore, implementation data was available for both years of the trial. This was advantageous as including implementation data at both time points could provide an insight into any fluctuations in behaviour changes observed in answering RQ1. As a reminder for the reader, fidelity, quality and dosage were the aspects of implementation focused on for the present study. This data was collected via a structured observation, and an electronic scorebard.

Structured observation for the implementation of the Good Behaviour Game

A structured observation schedule alongside an annotated rubric was formulated by the research team of the wider GBG trial (see section 1.4.2 and Appendix A). This was achieved by drawing from key components of the GBG identified as necessary by the coaches' fidelity checklist and the implementation manual, both of which were developed by AIR (Ford et al., 2014). As the GBG was heavily prescriptive, the majority of the schedule focused on the extent to which teachers delievered the intervention with fidelity. Taking inspiration from the coaches' checklist, the majority of the items had binary responses. As there were many items for the observer to be aware of, and many intervention sessions would be short (approximately 10 minutes), it was necessary for the schedule to be quick and easy to fill in while capturing a large amount of information in a short period of time. Another feature that the schedule had to aid the observers was subsections for the different stages of the game: the pre-game, during the game, and post game. Each subsection had the releavant items

in chronological order so the observer could mark off whether the teacher completed the action at the correct part of the game.

Once the structured observation was developed, video footage of teachers playing the GBG that had never been seen by the researchers, taken from the pilot study in Oxford (Chan et al., 2012) was used to practice using the observation. This was to see if it was user friendly, to train the researchers in how to use it, and establish inter-rater reliability between the observers. This involved discussing differences in scoring or altering the schedule to reflect common understanding. The final measures of inter-rater reliability coefficients were "almost perfect", with the intra-class coefficient (ICC: >.74) used for ordinal items, and Cohen's kappa coefficient (Kappa: >.8) used for nominal items (Hallgren, 2012).

Three trained researchers collected the fidelity and quality data through observing teachers implementing the GBG as part of the evaluation study.

Fidelity

The structured observation had 32 categorical items that measured fidelity (See Appendix A). As already mentioned, the researchers based the items on the instructions stipulated in the manual and the coaches' checklist (e.g. "Pupils are in clear teams"). The items required trained observers to mark each response with either a "yes" or "no", thus the minimum achievable score a teacher could receive was zero. The maximum score that could be achieved was 32.

Quality

Quality was measured using five ordinal items that were based on the definition of quality established in the literature review (see section 4.2.2). A teacher could receive a score that ranged 0-2 per quality item (e.g. "Interest and enthusiasm"). The higher the score, the higher the level of quality the teacher displayed, this meant the minimum score that could be received was 0, while the highest score was 10.

Justification of utilising the observation schedule

While the evaluation study offered other available methods of assessing implementation, e.g. teacher self-report surveys or data from the coaches, the researcher chose the observation schedule as the primary measure of implementation. One of the advantages of using independent observations is that they have higher validity compared to self-report measures, which are usually subject to biases such as social desirability (Dane & Schneider, 1998). Lillejoj, Griffin, and Spoth (2004) found in their study the mean proportions of scores teachers had awarded themselves were consistently higher than the observer ratings of implementation. It was also established that the independent observations of intervention implementation were better indicators of pupils' outcomes than the self-report data, and reasoned that discrepancies in scores were a major factor (see

section 4.2.4). This is because research suggests that self-report implementation data from teachers and coaches are likely to be biased, in that teachers and coaches could over-estimate their scores. Both sources of data may also create the feeling that judgment being passed on teaching or coaching abilities.

However, one disadvantage of using the independent structured observation data is that the observers are basing their scores on one implementation session of GBG. Using one sessions' worth of data could mean the score awarded to the implementer may not be reflective of the teacher's day-to-day implementation. Many factors could contribute to a change in implementation delivery when the observers are present. These could be observer effects, such as nervousness or social desirability, or unexpected circumstances that may cause the observation to be of an "a-typical" GBG lesson. One solution to this would be to conduct multiple observations over the academic year.

While multiple observations by an independent observer over the academic year may ease observer effects and is generally thought to yield stronger associations with outcomes compared to a single observation (Resnicow et al., 1998), there are drawbacks to this approach. For every round of observations conducted in classrooms implementing the GBG, it is an increase in data burden for the participating schools, which is likely to be a factor in schools discontinuing participation in the trial, let alone recruiting schools to participate in the trial in the first place. This issue is pertinent for the present study as implementation data from the second year of the GBG evaluation trial is utilised; this was also the year the wider trial collected data on academic outcomes, which also required a visit to the school. Schools were unlikely to be willing to agree to further visits. Therefore, the researcher decided that a single observation by an independent observer is the most appropriate and feasible method to measure implementation.

Electronic Scoreboard Data

The research team at the University of Manchester developed an online scoreboard as part of the wider trial. The researcher of the present study then used the frequency and length of the GBG implementation. Teachers were able to log onto the site and could set a built-in timer for the desired length of time and use the scoreboard during gameplay, instead of the paper versions provided by the developers. Alternatively, teachers were able to enter the dosage and score data into the scoreboard post-implementation if that was more convenient. The scoreboard was chosen over other available measures of dosage, such as the structured observation because the data provided by the teachers was the most accurate measure of game length and frequency. One limitation of the yearly structured observation was that it was unable to measure the frequency teachers implemented the GBG. A solution to this would have been to rely on teacher self-report data; however, as discussed earlier, the validity of teachers' responses is questionable.

One drawback to using data collected from the electronic scoreboard is that there was no guarantee that all the teachers would use the resource. It was not compulsory for teachers to use the electronic scoreboard, as they were provided with resources by the developers in order to monitor dosage for the coaches. Teachers with a preference for using the online scoreboard may be younger as their teaching practices are already likely to be integrated with the use of information technology in the classroom. This could mean that the teachers used in the analysis for implementation with regards to dosage may not be representative of teacher's implementing the GBG.

Dosage

By utilising the electronic scoreboard, the dosage was measured by establishing the average number of minutes a teacher delivered the GBG across a school week (five days). The average was calculated by dividing the total number of minutes played with the number of academic weeks available in a trial year. This approach was chosen over using the total number of minutes the GBG was played for the academic year as it would give some indication of how intensely the class were playing the game (see section: 4.2.5).

5.7 Procedure

1. From March 2015 to May 2015 the researchers contacted schools that attended recruitment events in order to secure a sample. Interested schools signed an MOA ahead of the start of the trial.

2. At the start of June 2015, year two teachers in all participating schools completed the TOCA-C for each child in their class.

3. Each time a school submitted their baseline TOCA-C data, a researcher contacted the Clinical Trials Unit at the University of Manchester to find out which condition the school had been randomised to. This was communicated both over the phone and in the form of a letter to the head teacher.

4. In September 2015, year three teachers from schools that had been randomised into the intervention condition attended a two day training course. The researchers attended the training events to match pupils to their teachers in schools that were more than one form entry.

5. After training, teachers began implementing the GBG in their classrooms. Throughout the academic year, teachers would submit their dosage data through the electronic scoreboard. Coaches continued visiting the teachers as part of the technical support ptovided by the GBG model.

6. Then, in December 2015, a data request was issued to the NPD to collect school level and pupil level data with regards to FSM eligibility and other socio-economic data.

7. From January to April 2016 a series of structured observations were completed. Each teacher participating in the intervention condition partook in a one-off

observation, which was conducted by a trained researcher in order to collect implementation data.

8. In June 2016, the year three teachers in both conditions filled in the TOCA-C for each pupil in their classroom.

9. Then, in September 2016, the year four teachers within the intervention condition attended the GBG training course. Again, the researchers attended the training events for pupil-matching.

10. Much like the previous year, after training the teachers began implementing the GBG in their classrooms. Again, the year four teachers submitted their dosage data through the electronic scoreboard.

11. After training, teachers began implementing the GBG in their classrooms. Throughout the academic year, teachers would submit their dosage data through the electronic scoreboard. Coaches periodically visited the teachers as part of the technical support ptovided by the GBG model.

12. From January to April 2017, a series of structured observations were completed. Each teacher participating in the intervention condition partook in a one-off observation, which was conducted by a trained researcher in order to collect implementation data.

13. Finally, in June 2017 the Year 4 teachers from both the control and intervention schools completed the follow-up TOCA for all children in their classes.

5.8 Analytical strategy

The purpose of this section is to justify the appropriateness of the analyses selected for the present study. Below, the researcher presents details on the following: the design's ability to detect intervention effects for both RQs; the steps that will be taken to screen and validate the data; the analytical strategies for both RQs. Key terminology and information, which will help with interpreting the results in later chapters.

5.8.1 Justification for statistical tests used in the present study

This section provides the reader with the rationale behind the core statistical methods utilised for the present study. Two tests were employed, multi-level modelling and hierarchal cluster analysis.

Multi-level modelling

A multi-level model (MLM) is a statistical technique used to analyse relationships between different variables that occur at different levels of a hierarchy within a data structure (Twisk, 2006). The concept came about within the field of education as data tended to follow a nested pattern: i.e. pupils are nested within classrooms, which in turn are nested in schools (Goldstein, 2011). MLM is an extension to linear regression but has the added strength of taking into account the complexities of the hierarchical nature of data. Firstly, MLM

considers the clustering found within multi-level data (Twisk, 2006). The consideration is particularly useful when modelling data collected from schools. For example, MLM can account for the fact that two children in a classroom are more likely to have similar outcomes compared to two children from different classrooms.

Secondly, MLM also allows one to analyse the effects of the wider context on individual outcomes. For instance, cluster-level variables can be included in analyses to determine their relationship with individual outcomes. Cross-level interactions are also a useful analysis that takes contextual factors into account; for example, where a cluster-level variable modifies the effects of an individual level variable on an individual's outcome (Goldstein, 2011). This feature is particularly pertinent to RQ1d and RQ3c, where the hypotheses predict that school-level poverty modifies the effects of individual-level poverty on children's behavioural outcomes.

Finally, MLM can analyse the heterogeneity of data-points between clusters, as well as the variation in covariate effects across clusters (Twisk, 2006). Regarding the present study, this would see whether behavioural outcomes vary across classes, which then could be further explored by analysing behavioural outcome variations based on FSM status across classes. Using regression as the primary analytical method within the present study would, therefore, be inappropriate and would likely lead to incorrect statistical inferences such as a type I error (Goldstein, 2011). By using MLM as the primary analytical component, one can place greater confidence in the conclusions drawn to answer the present study's hypotheses.

Hierarchical cluster analysis

The purpose of cluster analysis is to discover stable classifications within data using objective methods (Everitt, Landau, Leese, & Stahl, 2011). The method achieves this by partitioning the data so that individuals can be placed into a cluster, with the total number of clusters produced accounting for all the data. There are two defining characteristics of a cluster: one is based on internal cohesion (i.e. the data points are homogenous), and the other is external isolation (i.e. the data points are separated from the rest of set) (Cormack, 1971; Gordon, 1987). Although there are multiple forms of cluster analysis, hierarchical cluster analysis is utilised here because it is suited to smaller data sets such as the present study (Tan, Steinbach & Kumar, 2014). This form of analysis is unique in that data is partitioned multiple times, running from a single cluster containing all individuals to *n* clusters that comprise a single data point (Everitt, Landau, Leese, & Stahl, 2011). It is therefore up to the researcher to decide which number of clusters is the most appropriate. However, this can lead to the researcher imposing structure on the data rather than producing categories in the data that are inherently present.

The researcher decided on utilising cluster analysis because the present study subscribes to a micro-definition of implementation, as well as a person-centred approach to measuring the construct (see section 4.2.1). Cluster analysis was deemed appropriate in this regard as separate variables could be operationalised to represent the subordinate constructs (i.e. fidelity, quality of delivery, and dosage) that will be tested in the present study. The approach also acknowledges that teachers were likely to vary their implementation across different dimensions. The researcher chose Cluster analysis over creating three separate implementation categories based on external cut off points. The main reason for this was that cluster analysis did not require the dichotomising of variables, which reduces the statistical power of a study, unlike the alternative method (Altman & Royston, 2006).

Another reason cluster analysis was chosen over other available grouping strategies was that dosage comprised of two variables: frequency of games played in a week, and the length of games played. These variables would have been difficult to create a cut-off point for, as the implementation manual recommends increasing both the frequency and the length of games throughout the academic year (Ford et al., 2014). While the manual stipulates that there is an aspiration for implementers to be delivering the GBG 3 times a week, there is no guidance regarding at what point in the year teachers are expected to reach that stage. This suggests that teachers are expected to increase dosage as they see fit, based on whether their pupils are succeeding regularly at the GBG parameters already in place. Therefore it would be difficult to suggest an appropriate cut off point as teachers may only be implementing the GBG once or twice a week or playing short games based on the needs of their class, as suggested by developers.

5.8.2 Power analysis for the present study

Using the software *Optimal Design*, power calculations for the RQs concerning disruptive behaviour as the primary outcome were made in order to work out the minimal detectable effect size (MDES). This analysis was run to establish the sensitivity of the MLM that would answer the research questions of the thesis. It was therefore imperative to ascertain the extent to which the inferential statistics would be able to detect and effect if there was one to identify.

Research question one

Using the standard assumptions of Power=0.8 and Alpha=0.05, power calculations for RQ1a revealed that the MDES for ITT would be 0.15 (total sample N=3093). This calculation was based on a cluster size of 41 (the average number of children per school), and 77 schools. The calculation also required an intra-class coefficient (ICC= 0.04), which measures the similarities displayed within the different clusters. In this case, as expected, 4% of the variance explained in children's behaviour is due to differences between schools. The ICC estimate is based on PATHS SDQ data (Humphrey et al., 2016). A pre-test-post-test correlation (rho= 0.3) was also included in the analysis, as previous disruptive behaviour

scores account for 9% of the variance in post-test scores; this figure too was based on PATHS SDQ data (Humphrey et al., 2016). In conclusion, for RQ1a the proposed study is more than adequately powered to detect treatment effects of the GBG on disruptive behaviour, as mean effect sizes of between 0.3 and 0.75 have been reported in the literature (Ialongo et al., 1999). A subgroup analysis was conducted for FSM uptake (26%, sample size= 804) for RQ1b-1d. Using the same assumptions reported for RQ1a, the *a priori* MDES is 0.2.

Research question two

RQs 2a and 2c did not require power calculations, as these research questions were not concerned with detecting effect sizes. Meanwhile, the remaining sub-questions required power calculations. As before, the researcher utilised the standard assumptions of power and alpha for the calculations, but adjustments had to be made to account for the smaller number of schools (N= 38) and the introduction of a class-level within the analyses. Firstly, this is because data concerning the implementation of the GBG could only apply to the experimental group, so the sample size was drastically decreased. Secondly, as the GBG was delivered at the class-level, teachers within the same school could have implemented the intervention at different levels of fidelity, dosage, or quality of delivery. By introducing an additional level within the data, the researcher was able to acknowledge these differences. However, these adjustments did reduce the MDES to 0.5 for RQ2. The low MDES means that RQ2 is not adequately powered. While this indicates that inferences made for this portion of the study would not be suitable for explanatory purposes, the findings can be used for exploratory purposes (Varadhan et al., 2013).

5.8.3 Analytical strategy for research question one

In this section, the full proposed analytical strategy for RQ1 is presented. This is to give the reader an understanding of how the data was intended to be screened and analysed once the wider trial released the data to the author.

Data validation and screening

First, the percentage of missing data for key variables at the school-level and pupil-level had to be established. This was achieved using the descriptive statistics function on *SPSS*. The percentages were calculated for FSM status and the disruptive behaviour scores taken from the TOCA-C subscale. As missing data was inevitable in a large longitudinal study, it was necessary to determine the randomness of the missing data. To achieve this, a logistic regression was performed, it was an essential step for MI.

Next, the data had to meet a set of criteria in order to be suitable for MLM. Using *SPSS* and *MLwiN* several statistical tests were used. See table 5.7 for the checks and tests that were utilised.

Table 5.7
Requirements and tests used to determine the data for RQ1 is suitable for MLM.

Requirement	Test	Reason
Linerarity	Residuals plot	Analysis is meant to the best rectilinear line that explains the data based on chosen parameters (Goldstein, 2011).
Homogeniety of variance	Levene's test?	Variance of the residuals must be equal across groups (Goldstein, 2011).
Residuals are normally distributed	Q-Q plot	Residuals must not strongly deviate from the provided line (Goldstein, 2011).

Finally, baseline comparisons between the two conditions at the school-level and pupil-level had to be made. It was necessary to identify any major differences between the two conditions in case they could be responsible for any effects detected within the MLM. It is important to identify such differences in order to present an informed representation of the findings.

Once, the data was screened, the outcome data i.e. TOCA-C responses, were converted to z-scores. This form of standardisation allows for the direct comparison of effect sizes across and within multi-level models.

Case-wise multi-level model

1. Using *MLwiN,* an empty model was constructed to ascertain the unexplained variance accounted for by the school-level and pupil-level.

2. Then, a background model was constructed before adding the explanatory variables. This approach allowed for comparison once the explanatory variables were included in the full model. The background model contained the demographic variables that require controlling for in the full model. Gender was included at the pupil level as a control variable, as the literature suggested males have poorer behavioural outcomes compared to females (Deater-Deckard & Dodge, 2009; Kellam et al., 1998). Pupil's baseline TOCA-C score was also included as a control variable, as past scores often account for a high proportion of the variance.

3. Next, the background model had the IV added to it, in this case, the delivery of the GBG at the school level.

4. The full model included the interaction variables required to answer RQ1. See table 5.8 for a list of variables included in the full model, and a justification for their inclusion within the analysis.

Table 5.8

Interactions	Justification/explanation
IF GBG*IF post-test	RQ1a: This interaction compares both the usual practice and intervention condition's TOCA-C scores to determine whether there is a main effect of the GBG on disruptive behaviour.

Multi-level model	IF GBG *IF FSM eligible*IF post-test	RQ1b: This interaction compares the TOCA-C scores of FSM eligible pupils in the intervention condition, with their non-eligible counterparts to determine if there is a difference between these two groups.
	IF GBG *IF high school poverty*IF post-test	RQ1c: This interaction compares the TOCA-C scores of pupils in the high % of FSM uptake cluster of schools with the moderate % cluster between the two conditions.
	IF GBG* IF high FSM* IF FSM eligible *IF post-test	RQ1d: This interaction compares the TOCA-C scores of FSM eligible children in the high % of FSM uptake with FSM eligible children in the low % cluster between the two conditions.

interactions required to answer research question one.

Multiple imputation

REALCOM-Impute, multiple imputations (MI) of missing data was utilised and then compared to the case-wise MLM models in order to establish which set of models would be the most credible to report (see section 6.3 for further information on how MI was conducted).

Multi-level model using data from multiple imputation

1. Again, an empty model was constructed to ascertain the unexplained variance accounted for by the school-level and pupil-level.
2. Then, the background model was constructed before adding the explanatory variables. The same demographic variables that were controlled for in the case-wise model were included.
3. Next, the background model had the IV added to it, in this case, the delivery of the GBG at the school level.
4. The full model included the interaction variables required to answer RQ1. See table 5.8 for a list of variables included in the full model, and a justification for their inclusion within the analysis.

5.8.4 Analytical strategy for research question two

In this section, the full proposed analytical strategy for RQ2 is presented. This is to give the reader an understanding of how the data was intended to be screened and analysed once the wider trial released the data to the author.

Data validation and screening

In addition to the missing data at the school-level and pupil-level data described for RQ1 (see section 5.8.3), the missing implementation data taken from the electronic scoreboard and the structured observation would also be calculated at each time point. Again, logistic

regression would be performed to determine the randomness of the data. In the case of RQ2, teacher variables taken from the wider trial would be used to establish any patterns of missingness in the implementation data.

In addition to establish the suitability of the data for MLM (see section 5.8.3), it had to be suitable for χ_2 and HCA. The following requirements were needed in order determine an association with implementation profiles and school-level poverty through use of χ_2 (Field, 2009):

- The two variables must be categorical data.
- The two variables must have two or more independent categorical groups.
- An expected value of ≥ 5 must be calculated for each cell.

Meanwhile, the requirements for HCA were as follows (Gordon, 1987):

- Variables cannot be a mix of categorical data and interval data.
- Variables must not correlate highly. A correlational test was used to test this.
- Variables must not have many outliers. A scatterplot was generated to test this.
- Variables must be standardised.

Hierarchical cluster analysis
Using *SPSS,* HCA was conducted to determine if there are distinct implementation profiles amongst the teachers delivering the GBG. The profiles will be based on the input variables of fidelity, quality, and dosage.

Chi-square
For RQ2c, SPSS was used to conduct χ_2 in order determine an association with implementation profiles and school-level poverty.

Case-wise multi-level model
1. Using *MLwiN,* an empty model was constructed to ascertain the unexplained variance accounted for by the school-level and pupil-level.
2. Then, the background model was constructed before adding the explanatory variables. The same demographic variables that were controlled for in RQ1 were included.
3. The full model included the implementation profile clusters and interaction variables required to answer RQ 2. See table 5.9 for a list of variables included in the full model, and a justification for their inclusion within the analysis.

Table 5.9

Interactions	Justification/explanation

Multi-level model	IF strong implementation cluster*IF post-test[1]	RQ2c: This interaction compared the TOCA-C scores of children in th[ ] implementation cluster where teachers implemented the GBG mos[ ] accurately, with their counterparts who were in the implementation cluste[ ] where GBG delivered at a lower implementation status.
	IF strong implementation cluster*IF FSM eligible* IF post-test[1]	RQ2d: This interaction compared the TOCA-C scores of childre[ ] experiencing poverty placed in a strong implementation cluster, with the[ ] counterparts in the weak implementation cluster.
	IF strong implementation cluster* IF high school poverty* If post-test[1]	RQ2e: This interaction compared the TOCA-C scores of children placed in [ ] strong implementation cluster that attend a school with a high FSM uptak[ ] with their counterparts in the weak implementation cluster.
	IF strong implementation cluster* IF high school poverty* IF FSM eligible*IF post-test[1]	RQ2f: This interaction compared the TOCA-C scores of childre[ ] experiencing poverty that attend a school with a high FSM uptake, with the[ ] counterparts in the weak implementation cluster.

interactions required to answer research question two.

[1]In the event that more than two clusters were generated by HCA, then parallel interactions would be utilised to test for differences between all the available implementation clusters.

5.9 Ethical considerations

The present study utilised data from a wider evaluation of the GBG, which received ethical approval from the University of Manchester Ethics Committee in May 2015 (Ref: 010415). Nevertheless, the ethical considerations that are relevant to the current study are discussed alongside the protocols followed in order to address these issues. It is important to account for the ethical treatment of participants for moral reasons as well as enhance the study's credibility.

5.9.1 Informed consent

In order to ensure fully informed consent from all parties involved in the study, the teachers and parents of pupils of participating schools received information letters. The letters clearly outlined what data were collected and how the data would be used for the study (see appendices B & C). By providing all the relevant information about the nature of the study, as well as providing a point of contact for parents and teacher to answer any outstanding questions the present study ensured that they were fully informed. The teacher's information letter further explained the research team's expectations regarding data collection because they would be completing surveys on their implementation of the GBG, as well as DB outcome data on behalf of their pupils. Consent was established through an opt-out form, this meant that teachers or parents fully consented to the study if they did not return the opt-out form.

5.9.2 The use of randomisation

The majority of schools had volunteered to participate in order to access the GBG in the hope to improve behaviour in their classrooms. It was, therefore, essential to make schools aware of the fact that by consenting to participate, there was an equal chance of being

allocated to either condition. It would be misleading to allow schools to participate in the research being unaware or having high expectations of the likelihood of being randomised into the intervention condition. To ensure that schools understood the randomisation process, the EEF made it mandatory that schools had to sign a memorandum of the agreement before random allocation could take place (Humphrey et al., 2018).

5.9.3 The use of observation

While the present study investigates the GBG's effectiveness in behavioural outcomes, particularly for children experiencing poverty, RQ2 focuses on implementation as a moderator for behavioural outcomes. To achieve this, researchers observed the teachers when delivering the GBG in the classroom in order to collect relevant implementation data. As the data used in the current study was attached to the GBG evaluation trial, the observation process could have led to teachers feeling like researchers were evaluating their teaching practices on a more personal level. To prevent any potential harm to teachers' self-esteem, before the observation took place, the observers reiterated the purpose and nature of the stud.

5.9.4 The right to withdraw

Schools and pupils were able to withdraw from the study at any point. This was made known to the teachers and parents of pupils within the information letter and opt-out form. Teachers and parents of children were required to sign and return the form to either the school or the University of Manchester if they wished to withdraw from the study. The information letter clearly stated that there was no penalty for withdrawing from the study and that a specific reason did not have to be given either.

5.9.5 Confidentiality and anonymity

As part of the wider evaluation trial, potentially identifying information about child participants was collected from the national pupil database (NPD). Such information was only used to provide background information such as FSM eligibility status. Once baseline data was established, pupils were given unique reference numbers to match data across the multiple time points of the study. Teachers were asked to provide information such as their age and number of years teaching as part of the data collected for the implementation survey; however, the teachers also received unique reference number to protect their identity. Schools also received unique reference numbers to aid data matching across the different time points of the study. Any identifying information was destroyed before the analytical strategy was put into practice.

In order to preserve anonymity, teachers were informed through the distribution of the information letters that answers on both the implementation survey and the pupil outcomes survey would be kept confidential. Both teachers and parents were made aware that all responses would not be seen outside the University of Manchester. Both parties were also

informed that it would not be possible to identify individual responses from the present study, as the raw data for the study was stored in a password-protected file in order to maintain confidentiality.

5.9.6 Incentives for participation

Although teachers were not directly given incentives for participating in the study, schools in the UP condition received a monetary incentive from the EEF to remain in the wider evaluator study. Furthermore, parents did not receive a reward for agreeing to participate in the study.

5.9.7 Best practice

In addition to the above considerations, the researcher abided by ethical guidelines from additional, reputable sources such as the *Code of Ethics and Conduct* (British Psychological Society, 2009) and the *Ethical Guidelines for Education Research* (British Educational Research Association, 2011) in order to ensure participants were protected when participating in the study. Appendices B and C contain the relevant consent forms used by the study.

5.10 Chapter summary

Firstly, the researcher presented the context of the current study and highlighted the critical divergences from the intervention evaluation. Secondly, the chapter presented justifications for the adoption of a post-positive perspective for the present study. Then the researcher explored the objective stance and the impact it had on the methodological decisions made for the present study. This fed into a discussion regarding the use of the experimental design in education research, and as the design of choice for the present study. The chapter then presented arguments for the use of RCTs in the field of education, as well as exploring the design's limitations. Afterwards, the chapter explained how the participants were recruited, as well as their demographic details. The randomisation process was also presented. The researcher then presented the power and sample size calculations for each research question. The penultimate section discussed the selection and development of instruments used for data collection, as well as the choice of proxies used for child and school level poverty. The protocol for data collection, validation and analysis were also presented. Finally, the ethical considerations required for the study were outlined.

Chapter 6
Results

6.1 Introduction

The steps taken to answer each RQ were presented in chronological order, from data screening to the construction of the MLMs. It was decided by the researcher that presenting the analytical steps taken for each RQ in this way would provide the reader with a detailed understanding of the data analysis process that was undertaken.

6.2 Research question one: does poverty moderate the effects of the GBG on children's disruptive behaviour?

This section details the steps taken to analyse the data collected in line with the analytical strategy presented for RQ1 in section 5.8.3, each subsection provides detail of the purpose of the statistical tests performed and how they serve to prepare the data for MLM or presents the modelled data. The final subsection provides a summary to clearly answer the RQs for the benefit of the reader.

6.2.1 Missing Data

As described in Chapter 4 (see section 4.5 for details on the participants characteristics) 77 schools (N= 3084 pupils) were originally recruited and randomly allocated to either the intervention condition (N=38 schools) or the usual practice condition (N=39). There was no school-level attrition. However, there was pupil-level attrition. Complete data was available for 89.6% of pupils at T2 (*N*=2722). Table 6.1 details attrition for disruptive behaviour (TOCA-C) at the three time points of the study as well as the relevant missing demographic data. Possible explanations for the loss of data throughout the study are explored below.

Table 6.1
Missing data for research question one.

Trial condition	FSM status (missing)	Disruptive behaviour (TOCA-C)		
		T1 (missing)	T2 (missing)	T3 (missing)
GBG	1544 (16)	1497 (63)	1410 (150)	1202 (358)
Usual practice	1493 (31)	1469 (55)	1359 (165)	1310 (214)
Total	3037 (47)	2966 (118)	2769 (315)	2512 (572)

Missing outcome data

It is possible that missing disruptive behaviour data for pupils is a result of teachers not completing the TOCA-C as part of the online survey. This could be due to teachers finding the items not appropriate for children within different subgroups (e.g. children eligible for FSM or children listed as having SEND), or fatigue from completing a survey for each child

in their class. A more likely reason for missing outcome data is that teachers did not fill in surveys for students that had left the school.

Missing demographic data

As demographic data such as FSM eligibility was requested from NPD during the baseline phase of the RCT, children who could not be matched with the data provided were only included within the MI analysis. However, as demonstrated in table 5.1, the number of pupils with missing FSM status was deemed negligible as this represented 1.5% of the sample (Humphrey et al., 2018).

6.2.2 Missing data analysis

Although common practice amongst education researchers is to screen data using list-wise and pair-wise deletion (Peugh & Enders, 2004); questions regarding "why" data is missing were rarely asked up until relatively recently. However, with missing data analysis becoming a staple to the quantitative research model, three "explanations" offer insight into this area (Pampaka, Hutcheson & Williams 2016; Allison, 2000):

1. *Missing completely at random:* All cases have the same probability of being missing.
2. *Missing at random:* The probability of missing data for a variable may be dependent on other measured variables.
3. *Missing not at random:* The missingness of data is dependent on both measured and unobserved data.

Ignoring any underlying patterns that may explain missing data increases the risks of drawing incorrect conclusions regarding research questions, particularly if there are different characteristics between the participants with missing data and the remaining sample (Pampaka et al., 2016). This raises questions about the validity of conclusions drawn from the study and extrapolated to the wider population (Little, 1988). Such inferences are usually drawn through an unrepresentative dataset as well as an uneven number of participants between conditions (Pampaka, et al., 2016). With regards to the present study, although the overall sample started off relatively balanced, through attrition there appears to be a disparity between the intervention and control conditions in terms of the outcome variable (disruptive behaviour), particularly at T2 (see table 6.1). Therefore, a preliminary analysis to ascertain any differences between individuals with complete and missing cases was conducted on the data. This would increase the rigour of inferential analyses later presented in this chapter.

Missing data analysis for research question one

A logistic regression was performed using SPSS (version 22) to ascertain whether demographic variables and a participants' baseline disruptive behaviour score affected the likelihood an individual's data was incomplete. The model correctly classified 83% of cases and was statistically significant (χ_2 = 17.83, *df*=8, *p*= .02). The regression explained 2.4%

(Nagelkerke R_2) of the variance in missing data. As demonstrated in table 6.3, higher disruptive behaviour scores at baseline were associated with a decreased likelihood of being categorised as having missing data ($p < .0001$). Likewise, males were .77 times more likely to exhibit missing data compared to females ($p = .01$). The remaining explanatory variables did not have predictive power. This suggests that missing data within the present study can be explained as "missing at random" (Pampaka et al., 2016). This has implications on how the final analyses would be conducted, as there is some dependence on measured variables within the dataset, which accounts for the missingness of the data. Based on the guidance provided by Pampaka Hutcheson & Williams (2016), multiple imputations of missing data were performed on the data utilised in the MLMs and compared to the case-wise MLMs to determine which model is most appropriate to draw inferences from.

Table 6.2
Results for missing data analysis using logistic regression analysis for research question one.

Covariates	B	SE	p	Exp (B)	95% CI for Exp (B)	
					Lower	Upper
Disruptive behaviour[1]	-.279	.059	.000	.756	.673	.849
Male	-.267	.106	.011	.766	.623	.942
Eligible for FSM	.009	.117	.942	1.009	.802	1.268
Ethnicity	-.046	.142	.744	.955	.722	1.262
EAL	-.141	.154	.359	.868	.642	1.174
SEND	.346	.123	.005	1.413	1.111	1.798

[1]Baseline TOCA-C subscale.

6.2.3 Suitability of data for multi-level modelling

Following the analytical strategy devised in section 5.8.3, the data for both T2 and T3 had to be assessed for its suitability for MLM. There were three assumptions that had to be met: linearity, homogeneity of variance, and normal distribution of variance. Ordinarily, the assumptions refer to the actual outcome data, but for MLM these assumptions refer to the residuals (Twisk, 2006). The residuals are the difference between the observed value and the model estimate value that are calculated for the outcome variable (Goldstein, 2011). The residuals were calculated for both T2 and T3 using SPSS 22.0 using the linear regression function.

Linearity

A residuals plot for both T2 and T3 was produced to test this assumption (See Appendix D.1 and D2). As no pattern emerged, the assumption was not violated.

Homogeniety of variance

A Levene's test confirmed that for both T2 ($F(1, 2510) = 2.74$, $p < .097$) and T3 data ($F(1, 2510) = 2.53$, $p < .084$), the variance was homogenous at all levels.

Normal distribution of residuals

Q-Q plots for the residuals of both T2 and T3 were generated (see Appendix D.3 and D.4), these followed the expected pattern of a diagonal line with some deviation at the tails. Further to this, histograms are also provided to further support the normality of distribution for the residuals of T2 and T3 (see Appendix D.5 and D.6).

As there were no violations of the above assumptions, the researcher concluded that the data was suitable for MLM.

6.2.4 Baseline comparisons

Although the descriptive statistics for schools and pupils randomised to the usual practice and intervention conditions were reported in tables 5.2 and 5.4, significant differences between the conditions at baseline must be investigated before the final analyses. This is because any significant differences are likely to alter interpretations of analyses, as there will be less confidence that differences could be considered to have been subsequent to data collection at baseline. Therefore, the amount of variance between schools implementing GBG and schools continuing their usual practice was examined in relation to demographic variables listed in table 5.3.

School-level variables

A 'multiple analysis of variance' (MANOVA) was conducted on school-level demographic variables, which showed that there were no significant differences between schools in both conditions ($F_{(8,67)}= .852$, $p=.56$). The results in table 6.3 show the statistical significance and magnitude for each dependent school-level variable used in the baseline analysis. The magnitude values (Partial Eta$_2$) displayed in table 6.3 are minute and are consistent with the non-significant effect found within the overall MANOVA model. Therefore, it can be confidently concluded that there were no meaningful differences between schools in the

School-level variable	*df* (error)	*F*	*p*	Partial eta$_2$
School size	1(74)	.21	.64	.00
% Pupils eligible for FSM	1(74)	.06	.80	.00
Urbanicity	1(74)	1.05	.31	.01
% Pupils ethnicity is White British	1(74)	.02	.89	.00
% Pupils with EAL	1(74)	.04	.85	.00
% Absence	1(74)	.03	.86	.00
% Pupils with SEND	1(74)	3.26	.08	.04
% Pupils achieving level 4 or above	1(74)	.26	.61	.00

intervention and usual practice conditions at the start of the study.

Table 6.3
Analysis of variance comparing school-level characteristics between trial conditions.

Pupil-level variables

As all pupil covariates are categorical, a chi-square analysis was utilised in order to assess the baseline comparisons between pupils attending schools that were implementing GBG and usual practice schools. All pupil level predictors were statistically significant, suggesting that there is an overall difference between the two conditions. However, the percentage differences between the two conditions, as depicted in table 6.4 suggest that the difference is small. This is most likely due to the large sample size rather than representing any meaningful difference between the intervention and control condition.

Table 6.4

Pupil-level variable	df	χ^2	p	% Difference between the two conditions
Gender	1	6.36	.01	+9.84% Males in the usual practice condition.
FSM eligibility	1	8.37	.004	-6.65% FSM eligible in the usual practice condition.
Ethnicity	6	16.21	.01	+3.68% White British in the intervention condition.
EAL	2	6.14	.05	-6.54% children with EAL in the intervention condition.
SEND	3	12.62	.006	-10.43% children with SEND status in the GBG schools.
Gender	1	6.36	.01	+9.84% Males in the usual practice condition.
FSM eligibility	1	8.37	.004	-6.65% FSM eligible in the usual practice condition.
Ethnicity	6	16.21	.01	+3.68% White British in the intervention condition.
EAL	2	6.14	.05	-6.54% children with EAL in the intervention condition.
SEND	3	12.62	.006	-10.43% children with SEND status in the GBG schools.

χ^2 analyses of pupil covariates and % difference in pupil-level variables.

6.2.5 Preliminary analyses for research question one

Before conducting MLM models, as dictated by the analytical strategy (see section 5.8.3), preliminary analyses in the form of ANOVAs were conducted. These analyses included key variables, in this case, gender and FSM eligibility, that would be controlled for in both the case-wise and MI MLMs.

Gender

In order to see if there were any gender by trial condition effects before MLM was utilised, two-way ANOVAs were run using DB scores at T2 and T3 as the dependent variable (see table 6.5). Both gender and trial condition had a statistically significant difference at both time points, but the interaction between the two IVs were not significant. This is further demonstrated by the ANOVA plots in appendices E.1 and E.2.

Table 6.5
Two-way ANOVAs comparing gender by trial condition effects on DB at both study time points.

Disruptive Behaviour	T2					T3				
	Type III Sum of Squares	*df*	Mean Square	*F*	*p*	Type III Sum of Squares	*df*	Mean Square	*F*	*p*
Corrected model	98.61a	3	32.87	58.05	.001	95.72b	3	31.91	46.90	.001
Intercept	7912.30	1	7912.30	13974.04	.001	7073.55	1	7073.55	10398.53	.001
Gender	89.21	1	89.21	157.55	.001	90.24	1	90.24	132.66	.001
Trial condition	12.88	1	12.88	22.78	.001	7.37	1	7.37	10.84	.001
Gender*Trial condition	.24	1	.24	.42	.51	.14	1	.14	.21	.65
Error	1566.15	2766	.57			1706.06	2508	.68		
Total	9715.28	2770				8991.55	2512			
	1664.76					1801.77	2511			

a R Squared = .059 (Adjusted R Squared = .058)
b R Squared = .053 (Adjusted R Squared = .052)

Free school meal eligibility

A further set of two-way ANOVAs were run using the same dependent variables in order to test for FSM eligibility by trial condition effects (see table 6.6). For both time points, the individual IVs were statistically significant, but the interaction between trial condition and FSM eligibility was not significant. The ANOVA plot provided in appendices E.3 and E.4 further demonstrates the lack of interaction between the two variables.

Table 6.6

Two-way ANOVAs comparing trial condition by FSM eligibility effects on DB at both study time points.

Disruptive Behaviour	T2					T3			
	Type III Sum of Squares	*df*	Mean Square	*F*	*p*	Type III Sum of Squares	*df*	Mean Square	*F*
Corrected model	45.30[a]	3	15.10	25.75	.001	29.27[b]	3	9.76	13.81
Intercept	6318.40	1	6318.40	10775.58	.001	5640.48	1	5640.48	7981.7(
Trial condition	5.12	1	5.12	8.72	.001	3.40	1	3.40	4.82
FSM eligibility	35.55	1	35.55	60.62	.001	23.40	1	23.40	33.11
Trial condition* FSM eligibility	.09	1	.09	.01	.95	.01	1	.01	.013
Error	1616.02	2756	.586			1767.40	2501	.71	
Total	9683.68	2760				8964.74	2505		
	1661.31	2759				1796.66	2504		

[a] R Squared = .027 (Adjusted R Squared = .026)
[b] R Squared = .016 (Adjusted R Squared = .015)

6.2.6 Case-wise multi-level models

When constructing MLMs, it is common practice to build the models in stages by adding variables to an empty model (Twisk, 2006). In this section, each model for T2 and T3 will be presented side by side and explained so the reader can understand the purpose of each model, as well as compare the results of each model for the two different time points.

Empty models

The first model, the empty model, accounts for the variance at each level. In this case, the pupil-level (referred to as with-in school variance in table 6.7) and the school-level. There are no explanatory variables added to this model, only the intercept. This is the expected value of Y if X had the value of zero (Twisk, 2006).

As table 6.7 suggests, for both time T2 and T3, the pupil-level variance is larger than at the school-level. This suggests that pupils' DB scores have a wider variation within the school

than the scores between schools. It is worth noting that the amount of pupil-level variance, how spread out the DB scores are, increases between T2 and T3.

Table 6.7

Unexplained variance for disruptive behaviour in empty models for T2 and T3.

Disruptive behaviour	Est.	S.E.	Z	p
T2				
B_0 Intercept	.03	.05	.67	.25
σ_{2r} Between-School Variance	.13	.10		
σ_{2r} Within-School Variance	.47	.10		
Model fit	7647.30			
T3				
B_0 Intercept	.02	.039	.54	.30
σ_{2r} Between-School Variance	.08	.019		
σ_{2r} Within-School Variance	.92	.026		
Model fit	7016.18			

Background models

The purpose of the background model is to show the extent of the variance identified by the previous can be explained by demographic variables. FSM eligibility, gender, T1 DB scores, and school poverty status were selected for the background model based on the demographic variables covered in the literature review (see section 3.2.3).These variables were included in order to control for these effects in the future models which answer the RQs;

Table 6.8 shows that for both time points, FSM eligibility, gender, and the baseline DB score were significant predictors of a slight increase in DB scores. School poverty was not statistically significant for both T2 and T3. As more variables have been added to the background models, the model fit number has decreased from the figures provided in table 6.7. This suggests, alongside the significant χ_2 distribution, the background model is better at explaining the variance than the previous model.

144

Table 6.8

Demographic variables explaining variance in background models for T2 and T3.

Disruptive behaviour	T2				T3			
	Est.	*S.E.*	*Z*	*p*	*Est.*	*S.E.*	*Z*	*p*
B_0 Intercept	-1.46	.07	-21.68	.00005	-1.443	.062	-23.24	.00005
B_1 FSM Eligible	.18	.03	5.64	.001	.10	.036	2.825	.003
B_2 Male	.15	.03	4.66	.001	.16	.03	5.118	.00005
B_3 DB T1	.78	.02	42.35	.001	.81	.02	39.763	.00005
B_4 School Poverty	.13	.09	1.43	.08	.03	.078	.384	.351
σ_{2r} Between-School variance	.10	.06			.09	.018		
σ_{2r} Within-School variance	.26	.06			.51	.015		
Model fit	5871.29				5466.32			
χ_2 distribution	*p< .005*				*p< .005*			

IV models

The next set of models included the trial condition along with the variables controlled for from the previous model. The IV was entered in this way to determine the extent to which the variable contributed to the explained variance before the interactions were entered (See table 6.9). Note, that although the model fits for the IV models are marginally smaller, they are not statistically significant. This suggests that the previous model (see table 6.8) was a better explanation of the variance so far.

Table 6.9

T2 and T3 models with the school-level RCT condition entered as an explanatory variable.

Disruptive behaviour	T2				T3			
	Est.	*S.E.*	*Z*	*p*	*Est.*	*S.E.*	*Z*	*p*
B_0 Intercept	-1.27	0.084	-15.06	.001	-1.25	0.08	-16.11	.001
B_1 FSM Eligible	0.15	0.03	4.68	.001	0.10	0.04	2.81	0.01
B_2 Male	-0.16	0.03	-5.67	.001	-0.16	0.03	-5.14	.001
B_3 DB T1	0.78	0.02	42.34	.001	0.81	0.02	39.74	.001
B_4 School Poverty	0.12	0.09	1.39	0.16	0.03	0.08	0.35	0.72
B_5 GBG	-0.06	0.09	-0.69	0.49	-0.07	0.08	-0.86	0.39
σ_{2r} Between-School variance	0.13	0.02			0.09	0.02		
σ_{2r} Within-School variance	0.48	0.01			0.09	0.02		
Model fit	5870.82				5465.59			
χ_2 distribution	*p>.05*				*p> .05*			

Full models

The final case-wise models for RQ1 included the interactions required to answer the remaining RQs for RQ1. When comparing the model fit to the background model (the most appropriate model so far, see table 6.8) the reduction in value is statistically significant (see table 6.10).

146

Table 6.10

Full models for disruptive behaviour at T2 and T3 using case-wise data.

Disruptive behaviour	T2				T3			
	Est.	*S.E.*	*Z*	*p*	*Est.*	*S.E.*	*Z*	*p*
B_0 Intercept	-1.45	.09	-16.67	.001	-1.43	.078	18.264	.001
B_1 FSM Eligible	.16	.03	5.71	.001	-.03	.08	-.316	.38
B_2 Male	.24	.07	3.27	.001	.16	.03	5.26	.001
B_3 DB T1	.78	.02	42.34	.001	.81	.02	39.811	.001
B_4 School Poverty	.06	.13	.45	.33	-.06	.11	-.502	.31
B_5 GBG	-.02	.12	-.20	.42	.01	.105	.045	.48
B_6 School Poverty* FSM	-.21	.10	-2.18	.001	.13	.103	1.281	.10
B_6 GBG*FSM	-.13	.10	-1.29	.10	.05	.109	.452	.33
B_6 GBG*School Poverty	.15	.18	.82	.20	.09	.158	.543	.29
B_6 GBG*School Poverty* FSM	.33	.13	2.52	.01	.08	.143	.524	.30
σ_{2r} Between-School variance	.08	.06			.09	.018		
σ_{2r} Within-School variance	.27	.06			.51	.015		
Model fit	5861.70				5457.21			
χ_2 distribution	*p< .005*				*p< .005*			

6.2.7 Multiple imputation for research question one

MI replaces missing values with suitable estimates based on auxiliary variables available within the original data set and allows for statistical techniques reserved for complete datasets to be utilised in later analyses as the original sample size is maintained (Pampaka et al., 2016). Education researchers continue to remain sceptical of MI, as the addition of simulated data is often misinterpreted as data manipulation (ibid). However, simulated data minimises biases and yields a higher efficiency compared to case deletion (Durrant, 2009).

REALCOM-Impute was utilised to conduct MI, using the missing at random assumption (Carpenter, Goldstein & Kenward, 2011). Demographic variables (i.e. Gender, FSM eligibility, ethnicity, EAL, and SEND status) and explanatory variables (i.e. T1 scores of the

prosocial and concentration subscales of the TOCA-C), as well as the constant were entered as auxiliary variables. These were used to input the missing values using the default setting of 1000 iterations with a burn-in of 100, and a refresh of 10. This was based on the guidance provided by the developers, Carpenter and colleagues (2011), for multi-level imputation with mixed response types.

6.2.8 Imputed multi-level models

The full models for T2 and T3 were generated once the REALCOM-impute file was integrated into MLWiN. As table 6.11 demonstrates, the interaction between the school condition, school poverty status and FSM eligibility remained significant for T2. However, the model fit for both time points were not statistically significant. In fact, for time T3, the model fit increased, suggesting that it was a poorer fit. Therefore, the case-wise full models were used to answer RQ1 (see table 6.10).

Table 6.11
Imputed full model for RQ1 at T2 and T3.

Disruptive behaviour	T2				T3			
	Est.	S.E.	Z	p	Est.	S.E.	Z	p
B_0 Intercept	-1.43	.087	-16.42	.001	-1.41	.08	-17.91	.001
B_1 FSM Eligible	.22	.074	2.08	.001	-.02	.08	-.18	.43
B_2 Male	.17	.028	5.99	.001	.16	.03	5.36	.001
B_3 DB T1	.77	.019	40.10	.001	.80	.02	38.05	.001
B_4 School Poverty	.06	.127	.47	.32	-.04	.11	-.38	.35
B_5 GBG	-.03	.0118	-.22	.42	-.01	.11	.03	.49
B_6 School Poverty* FSM	-.18	.098	-1.83	.03	.12	.10	-1.19	.12
B_6 GBG*FSM	-.10	.101	-.95	.17	-.11	.10	-1.08	.28
B_6 GBG*School Poverty	.14	.178	.79	.21	.13	.18	.78	.44
B_6 GBG*School Poverty* FSM	.23	.133	2.08	.001	.29	.13	2.19	.61
σ_{2r} Between-School variance	.13	.023			.09	.02		
σ_{2r} Within-School variance	.50	.014			.52	.02		
Model fit	5864.70				5564.71			
χ_2 distribution	p> .005				*p*> .005			

148

6.2.9 Answering research question one

This section provides the direct statistical information required to answer the sub-questions within RQ1 (see table 6.10).

RQ1a: does the GBG reduce disruptive behaviour?

There was no relationship between GBG and behavioural outcomes at both time points (T2: β_{0j}= -.02, *p*= .42; T3: β_{0j}= .01*, p*= .48). Therefore, pupils' disruptive behaviour was not significantly improved by attending a school implementing GBG.

RQ1b: are there differential effects of the GBG among children who are eligible for FSM?

At both time points, there were no significant behavioural differences between children experiencing high levels of poverty and their less impoverished counterparts attending schools implementing GBG, relative to those in usual practice schools (T2: β_{0j}= -.13, *p*= .10; T3: β_{0j}= .05*, p*= .33).

RQ1c: does school-level poverty moderate the effects of the GBG on disruptive behaviour?

At both time points, there were no significant behavioural differences between children who attended schools implementing GBG and serving poor populations with schools serving more affluent populations (T2: β_{0j}= .15, *p*= .20; T3: β_{0j}= .09*, p*= .29).

RQ1d: do school-level poverty and child-level FSM eligibility interact to moderate the effects of the GBG on disruptive behaviour?

There was a significant interaction between school-level intervention status, school-level poverty status, and pupil-level FSM eligibility (β_{0ij}= .33, *p*= .01) in the first year of implementation. This predicted an *increase* in disruptive behaviour scores instead of a decrease. This meant that children who were identified as experiencing poverty and attending a school serving a high poverty population, which was also implementing GBG, had higher scores compared to peers who were also eligible for FSM but were attending moderately poor schools implementing GBG. However, the interaction was not replicated in the second year (β_{0ij}= .08, *p*= .30).

6.3 Research question two: does poverty moderate the implementation of GBG?

This section details the steps taken to analyse the data collected in line with the analytical strategy presented for RQ2 in section 5.8.4. Many of the steps that were taken for data analysis were similar to RQ1, so to avoid repetition, the in-depth explanations have been

omitted if they have already been covered in the previous section. The final subsection provides a summary to clearly answer the RQs for the benefit of the reader.

6.3.1 Missing data

Data for RQ2 was treated as two separate time points for two main reasons: new teachers were assigned to classes in the second year of the trial, and children often were re-assigned to different classes. Schools that had ceased implementation were not treated as missing data, so the data set was refined to remove these cases as part of the data screening process. Table 6.12 depicts the attrition for disruptive behaviour (TOCA-C), implementation data, and FSM eligibility across the three time points of the study. Possible explanations for the missing implementation data are explored below.

Table 6.12
Missing data for research question two.

Missing data for time point	Level of data		
	Pupil	Teacher	School
T1 (missing)			
TOCA-C$_1$	1497 (63)	-	-
FSM status	1544 (16)	-	-
T2 (missing)			
TOCA-C$_1$	1410 (150)	-	-
Observation	1425 (41)	54 (2)	35 (0)
Scoreboard	1306 (160)	28 (7)	31 (4)
T3 (missing)			
TOCA-C$_1$	1202 (358)	-	-
Observation	1179 (24)	45(1)	28 (0)
Scoreboard	1203 (0)	46 (0)	28 (0)

$_1$Disruptive behaviour subscale only.

Missing fidelity data

Teachers with missing data for fidelity were unavailable for observation visits; these were often cancelled at short notice, usually due to Ofsted visits or illness. These teachers had to be excluded from the analysis as HCA only uses complete cases (Tan et al., 2014).

Missing dosage data

A small number of teachers were not recording their GBG sessions using the electronic scoreboard. This may have been due to teachers preferring the regular scoreboard or were unable to connect to the Internet in order to enter dosage data. Again, these teachers had to be excluded from the analysis as HCA can only use complete cases to generate cluster solutions (ibid).

6.3.2 Missing data analysis

As detailed in section 6.2.2, it is important to establish the extent the missing data was random. Therefore, a logistic regression was performed using teacher's demographic variables alongside their baseline responses to wellbeing questions which were administered to participants as part of the wider trial (Humphrey et al., 2016). The model

correctly classified 89.5% of cases but the regression was not statistically significant (χ_2 = 2.72, df =5, p= .74). The model explained 6.4% (Nagelkerke R_2) of the variance in missing data. Table 6.13 demonstrates that none of the observed variables predict missingness.

Table 6.13
Missing data analysis in the form of logistic regression for research question two.

Covariates	B	SE	*p*	Exp (B)	95% CI for EXP (B)	
					Lower	Upper
Male	.226	.919	.806	1.253	.207	7.597
Years teaching	-.044	.039	.262	.957	.886	1.033
Efficacy[1]	-.199	.421	.637	.820	.359	1.872
Stress	-.218	.518	.674	.805	.292	2.218
Retention	-.194	.321	.546	.824	.439	1.546

[1]Teacher Efficacy scale

MI was not used to replace the missing values for RQ2 as REALCOM-Impute is unable to impute MLMs containing 3 levels of data.

6.3.3 Data suitability for hierarchical cluster analysis

Although cluster analysis is not considered to be an inferential statistical method, data should meet a set of assumptions, otherwise meaning cannot be drawn from the cluster solutions produced. Therefore, factors such as outliers should be screened for in order to determine whether the data set is suitable for HCA. A summary of assumptions and the extent to which they were met are presented in table 6.14. As the majority of the requirements were met, it was deemed that the data was suitable for HCA analysis.

Table 6.14
Data requirements for hierarchical cluster analysis.

Requirement	Evidence	Comment
Not a mix of categorical and interval variables.	See section 4.6	Only interval variables were used.
Variables must not correlate highly.	*rho*= .186, *p*= .075	Dosage and fid-qual were not significantly correlated.
Variables must not have many outliers.	See Appendix F.1	There was one outlier for dosage.
Variables must be standardised.	See section 5.8	Variables were standardised prior to HCA.

6.3.4 Hierarchical cluster analysis

Using fidelity-quality scores and dosage data, cases from both years of implementation were analysed using HCA on SPSS version 22. HCA works using a "bottom-up" process, in this case each value starts off a cluster in their own right (Gordon, 1987). Then, the two most similar clusters/or values are then merged to form a larger cluster, thus reducing the overall number of clusters. This process continues until singular observations become group members of larger clusters. Finally, all the clusters merge to form one cluster.

Although there is no statistical test to suggest which cluster solution is the most appropriate, the agglomeration schedule (see appendix G) was used to determine suitable cluster solutions. This was achieved by measuring the change in coefficients between stage jumps. The larger the coefficient change, the better the cluster solution. This method produced two cluster solutions that could be considered meaningful (see tables 6.15 & 6.16).

Table 6.15
Descriptive statistics for a two-cluster solution of implementation.

Cluster	N	Mean		SD	
		Fid-qual	Dosage[1]	Fid-qual	Dosage[1]
1	67	68.58	4.58	11.41	5.32
2	26	76.64	33.67	7.80	10.93

[1] Average minutes per week.

Table 6.16
Descriptive statistics for a three-cluster solution of implementation.

Cluster	N	Mean		SD	
		Fid-qual	Dosage[1]	Fid-qual	Dosage[1]
1	67	68.58	4.58	11.41	5.32
2	19	76.18	28.11	7.91	6.56
3	7	77.90	48.74	7.95	2.48

[1] Average minutes per week.

The three-cluster solution had the largest increase in coefficient, making this solution the most "valid" (the two-cluster solution had a difference of 0.06 compared to the three-cluster solution, which had a coefficient difference of 0.74). However, the two-cluster solution was more suitable for MLM, as the 3_{rd} cluster is very small in size (k=7). Despite this, the mean dosage (average minutes per week the GBG was played) was markedly different between clusters 2 and 3 in the three-cluster solution, even though the mean for fid-qual remained consistent. The difference in dosage is likely to be an important factor that should be acknowledged and so, therefore, the three-cluster solution was accepted.

Implementation profiles
As the interpretation of clusters is determined by the researcher, as opposed to a statistical test, a description and justification of each implementation profiles is provided below.

Medium fidelity, low dosage
Medium fidelity, low dosage (MF-LD) had the largest number of teachers within the profile (*n*=67) and was, therefore, the reference category for MLM. Although there was not much variation in terms of fidelity between this profile and high fidelity, medium dosage (HF-MD), teachers were delivering the GBG considerably less often than the developer's recommendation.

High fidelity, medium dosage

HF-MD had the second largest number of teachers within the profile (*n*=19). Teachers delivered the intervention at approximately the level the developers recommend.

High fidelity, high dosage

High fidelity, high dosage (HF-HD) consisted of the smallest number of teachers (*n*=7). Teachers in this cluster delivered the intervention above the minimum expectations of the developers.

6.3.5 Suitability of the data for multi-level modelling

An important consideration that had to be accounted for was the distribution of teachers with the cluster profiles across the two years of implementation. This would have implications in proceeding with MLM, particularly if the smallest implementation profile, HF-HD (n=7) was split across both years of implementation. See table 6.17 for the distribution of teachers across implementation years.

Table 6.17
Distribution of teachers across implementation years according to implementation profile.

Cluster	Implementation year		Total
	Year 1	Year 2	
MF-LD	23	44	19
HF-MD	18	1	7
HF-HD	7	0	7
Total	48	45	93

It was decided that as there was only the representation of HF-HD in the first year of implementation and very little representation of HF-MD in the second year of implementation, only data from the first year of the study would be used to answer the remain sub-questions of RQ2. Then, subgroup figures were calculated in order to determine the number of participants in each sub-group pertinent to the interactions required for MLM calculations. As depicted in table 6.18, for both schools serving populations of moderate and high poverty, the HD-HF cluster had the smallest number of children, with schools classified in the high poverty bracket as having almost half the number of children. Within the HD-HF implementation cluster, the number of children experiencing poverty was even smaller, and below the minimum number of participants per sub-group (Mass & Hox, 2005). This means the inferential analyses below should be interpreted with caution as the HD-HF*FSM interaction term is underpowered. See table 6.16 for MLM analyses for RQ2.

Table 6.18

The number of pupils by sub-groups of interest.

School poverty status	FSM eligibility	
	No	Yes
Moderate		
MF-LD	177	52
MF-MD	232	51

HD-HF	89	23	
High			
MF-LD	201	101	
MF-MD	123	83	
HD-HF	48	16	

6.3.6 Chi-square analysis

As both implementation variation and school-level were in the form of categorical data, the most appropriate method of analysis was the chi-square. However, the use of three implementation profiles would have violated one of the assumptions of the chi-square. Table 6.19 shows that less than five teachers in the high implementation profile were in the high poverty category.

Table 6.19

Distribution of teachers across implementation profiles and school-level poverty classifications.

Implementation profile	School poverty level		Total
	Moderate	High	
Cluster 1	10	12	22
Cluster 2	11	8	19
Cluster 3	5	2	7
Total	26	22	48

Therefore, in order to answer the research question, the clusters had to be remodified to answer RQ2b. Cluster two and three were recombined and renamed to form adequate fidelity and dosage, while cluster one remained the same. The combining of the higher implementing profiles took all the cells to above five counts per observed cell and therefore the violation was cleared (see table 6.20).

Table 6.20

Observed and expected distributions of teachers across amended implementation profiles and school-level classifications

Implementation profile	School level poverty	Total

	Moderate	High	
Low implementation			
Observed	10	12	22
Expected	13	9	22
Acceptable implementation			
Observed	16	10	26
Expected	14	12	26
Total			
Observed	26	22	48
Expected	26	22	48

Although table 6.20 demonstrates some variability between the observed and expected values within the chi-square analysis, particularly for low implementing teachers in high poverty schools and teachers in moderate schools implementing the GBG at an acceptable level. The chi-square analysis was therefore not significant (χ^2= 1.50, *df*=1, *p*=.22). The study concluded that implementation did not vary as a function of school-level poverty.

6.3.7 Case-wise multi-level modelling

Following a similar format to section 6.2.6, the MLMs for RQ2 were built in stages starting with an empty model, followed by a background model and finally the full model. It is pertinent to remind the reader at this stage that the MLMs constructed for RQ2 are underpowered and the results are reported for exploratory purposes.

Empty model

In this empty model, there are three-levels. In this case: the pupil-level, the class-level, and the school level (referred to as: within-class, between class, and between school in table 6.21). As expected, the higher the level, the smaller the amount of variance is demonstrated (see table 6.21). Therefore, pupils' DB scores have a wider variation within classes, than class DB scores and school DB scores.

Table 6.21
Unexplained variance for disruptive behaviour at the three levels of the empty model.

Disruptive Behaviour	*Est.*	*S.E.*	*Z*	*p*
B_0 Intercept	.02	.07	.33	.73
σ_{2r} Between-School Variance	.09	.05		
σ_{2r} Between-Class Variance	.05	.03		

σ$_{2r}$ Within-Class Variance	.86	.04	
Model fit	3057.48		

Background model

The same demographic variables as RQ1 were entered into the background model for RQ2: FSM eligibility, gender, T1 DB scores, and school poverty status. These variables were included to control for their effects in the final model. Table 6.22 shows that FSM eligibility, gender, and DB T1 were statistically significant in predicting an increase in the DB T2 score. Also, as more variables have been added to the model, the model fit has decreased considerably. As the χ_2 distribution was statistically significant, the background model is better at explaining the variance than the previous model.

Table 6.22

Demographic variables explaining variance in background model for RQ2.

Disruptive Behaviour	Est.	S.E.	Z	p
B$_0$ Intercept	-.16	.09	-1.76	.08
B$_1$ FSM Eligible	.13	.05	2.80	.005
B$_2$ Male	.12	.04	2.95	.003
B$_3$ Disruptive T1	.68	.02	29.64	.001
B$_4$ School Poverty	.18	.13	1.41	.16
σ$_{2r}$ Between-School Variance	.03	.04		
σ$_{2r}$ Between-Class Variance	.12	.05		
σ$_{2r}$ Within-Class Variance	.44	.02		
Model fit	2297.6			
χ_2 distribution	p< .005			

Full model

Although the introduction of the implementation categories into the MLM and the interactions improved the model fit, none of these additions were statistically significant (see table 6.23). Further to this, the only control variable to remain statistically significant was DB T1.

Table 6.23

Full three-level model for disruptive behaviour using case-wise data for RQ2.

Disruptive Behaviour	Est.	S.E.	Z	p
B$_0$ Intercept	-.07	.14	-.46	.64
B$_1$ FSM Eligible	-.02	.11	-.16	.87
B$_2$ Male	.12	.04	2.91	.09
B$_3$ Disruptive T1	.68	0.2	29.65	.001

B_4 School Poverty	.03	.03	.18	.86
B_5 HF-MD[1]	-.13	.18	-.69	.49
B_5 HF-HD[1]	-.16	.23	-.71	.48
B_6 HF-MD[1]*FSM	.23	.16	1.46	.15
B_6 HF-HD[1]*FSM	.22	.20	1.13	.26
B_7 HF-MD[1]* school poverty	.34	.25	1.37	.17
B_7 HF-HD[1]*school poverty	-.11	.39	-.27	.79
B_8 FSM*School poverty	.18	.14	1.29	.20
B_9 HF-MD[1]*school poverty*FSM	-.30	.20	-.47	.14
B_9 HF-HD[1]*school poverty*FSM	-.23	.29	-.80	.42
σ_{2r} Between-School Variance	.02	.04		
σ_{2r} Within-Class Variance	.12	.05		
σ_{2r} Within-School Variance	.43	.02		
Model fit	2291.85			
χ_2 distribution	p< .005			

[1]Reference category is MF-LD.

6.3.8 Answering research question two

This section provides the direct statistical information required to answer the sub-questions within RQ2.

RQ2a: are there distinct implementation profiles amongst teachers implementing GBG?

Three distinct implementation profiles were identified:
- Medium fidelity, low dosage (MF-LD; *n*= 67): Teachers were delivering the intervention less frequently than recommended by the developers but were implementing at an adequate level of fidelity.
- High fidelity, medium dosage (HF-MD; *n*= 19): Teachers delivered the intervention at the recommended level of implementation.
- High fidelity, high dosage (HF-HD; *n*= 7): Teachers delivered the intervention above the minimum expectations of the developers.

RQ2b: does implementation vary as a function of school-level poverty?

There was a non-significant association (χ_2= 3.71, *p*=.054) in implementation profiles across moderate and high poverty schools. The trend suggested that higher poverty was related to lower implementation.

RQ2c: does implementation variability moderate the effects of the GBG on disruptive behaviour?

There was no significant relationship between implementation profiles and disruptive behaviour (HF-MD: β_{0ij}= -.13, Z= -.69, p=.49; HD-HF: β_{0ij}= -.16, Z= -.71, p=.48).

RQ2d: do implementation variability and pupil-level poverty interact to moderate the effects of the GBG on disruptive behaviour?

There was no significant interaction between a teacher's implementation profile and pupil's poverty status in predicting disruptive behaviour outcomes (HF-MD: β_{0ij}= .23, *Z*= .16, *p*=.15; HD-HF: β_{0ij}= .22, *Z*= .20, *p*=.26). This meant children experiencing poverty in classes where the implementation of GBG was considered adequate or exceptional did not experience differential gains compared to peers in the lower implementing classes.

RQ2e: does school-level poverty and implementation variability interact to moderate the effects of the GBG on disruptive behaviour?

Again, there was no significant relationship between school poverty status and teacher implementation profile (HF-MD: β_{0jk}= .34, *Z*= .25, *p*=.17; HD-HF: β_{0jk}= -.11, *Z*= -.27, *p*=.79).

RQ2f: do School-level poverty, implementation variability and pupil-level poverty interact to moderate the effects of the GBG on disruptive behaviour?

Likewise, these interactions were non-significant (HF-MD: β_{0ijk}= -.34, *Z*= -.47, *p*=.14; HD-HF: β_{0ijk}= -.23, *Z*= -.80, *p*=.42). However, children who were experiencing poverty and in classrooms that not only within impoverished schools but also where teachers implemented GBG at an adequate or exceptional level had lower disruptive behaviour schools compared to their counterparts in lower implanting classrooms who were also experiencing poverty.

6.4 Chapter Summary

The chapter started with the reporting of the different steps taken to answer RQ1, from the data screening to the construction of the MLMs. Then the chapter focused on RQ2, which followed a similar manner. The majority of the RQs resulted in null findings, except for RQ1d where a three-way interaction between trial condition, school poverty status and pupil FSM status were statistically significant for T2 only.

Chapter 7

Discussion

7.1 Introduction to the chapter

The purpose of the present study is restated as a reminder for the reader when critically examining the findings in relation to the literature. The chapter then presents the summary of the results for RQ1 before discussing them in detail. This is followed by the summary and discussion of the RQ2 results. Next, the methodological and conceptual limitations of the present study are acknowledged. Afterwards, the researcher presents the implications of the current study for both the research community and practitioners. In this section, the discussion provided some recommendations for future studies. Subsequently, the contributions to knowledge are restated and expanded upon in light of the findings from the present study. Finally, the thesis concludes with a summary of the current study.

7.2 Purpose of the study

The primary purpose of the study was to assess the extent to which poverty at both the school level and the individual level moderated the implementation and behavioural outcomes of the GBG. The researcher utilised MLM in order to account for the hierarchical nature of the data. The data structure has been previously overlooked (e.g. Kellam et al., 1994; Kellam et al., 1998). It also meant that the researcher could control for factors such as gender in the analyses. Therefore, it is important to note that the current discussion will be concerning the role of poverty on GBG outcomes after gender has been controlled for. The secondary purpose of the study was to see if cluster analysis could group teachers into implementation profiles based on multiple aspects of implementation. The creation of implementation profiles was achieved through hierarchal cluster analysis. The section below summarises the findings of the study.

7.3 Summarised results for research question one

The over-arching question RQ1 attempted to answer was: does poverty moderate the effects of the GBG on children's DB? Results were reported for both years of implementation and generally indicated that poverty did not moderate the intervention effect of the GBG. However, a three-way interaction between the GBG, school-level poverty and pupil-level poverty was significant during the first year of the study. This finding indicated that economically disadvantaged children attending high poverty schools that implemented the GBG experienced increases in disruptive behaviour. Additionally, imputed models were generated and concurred with the reported case-wise models (see tables 6.10 & 6.11). This suggests the results are robust.

7.3.1 Does the Good Behaviour Game reduce children's disruptive behaviour?

For both years of the study, the GBG did not significantly reduce children's disruptive behaviour.

7.3.2 Are there differential effects of the Good Behaviour Game among children who are eligible for FSM?

For both years of the study, there were no significant differential effects among children who were FSM eligible.

7.3.3 Does school-level poverty moderate the effects of the Good Behaviour Game on disruptive behaviour?

In both years of implementation, schools serving highly economically disadvantaged neighbourhoods did not have significantly different behavioural outcomes when implementing GBG compared to control schools located in moderately economically disadvantaged areas.

7.3.4 Do school-level poverty and child-level FSM eligibility interact to moderate the effects of the GBG on disruptive behaviour?

In the first year of implementation, there was a significant interaction between school-level intervention status, school-level poverty, and pupil-level poverty. The results suggested that children experiencing poverty who attended a GBG school serving a moderately poor population were less disruptive compared to their equivalent peers in high poverty control schools. These results were not replicated in the second year of the study.

7.4 Discussion of findings for research question one

It was concluded that poverty did not moderate the effects of the GBG on disruptive behaviour. Many of the findings described for this RQ challenged the rationale proposed for the study. Therefore, potential explanations could be found by revisiting the present study's theoretical framework. As the sub-questions for RQ1 are additive, in that each sub-research question builds upon the previous one, there is a risk that the discussion points could become repetitive. To avoid repetition, fundamental theories and literature are presented in depth at critical points of the discussion and are briefly referred to afterwards when necessary.

7.4.1 Does the GBG reduce children's disruptive behaviour?

The findings suggested that the GBG was not responsible for the reductions in DB detected in both years of the trial. This is contradictory to a large proportion of the previous research testing the preventive intervention (e.g. Dion et al., 2011; Mitchell et al., 2015). For instance,

160

Kleinman and Saigh (2011) reported significant reductions in disruptive behaviour within a classroom of high school students. However, the majority of studies reporting GBG's success have been utilised small samples, many of which did not utilise a control condition (e.g. Kleinman & Saigh, 2011; Lannie & McCurdy, 2007). Meanwhile, large RCT trials, such as the one conducted in Baltimore (Kellam et al., 2014), have also reported no main effect for the GBG. Instead, these larger trials have reported significant subgroup effects, particularly for males categorised as persistently disruptive (Kellam et al., 2014; Kellam et al., 1998). For instance, Poduska and colleagues (2007) reported the reduced use of mental health services for males identified as disruptive and exposed to the GBG during their elementary years at the school. As these findings reflect the lack of the main effect universal prevention interventions are theorised to produce, it suggests that the present understanding of the mechanisms in which universal preventive interventions (UPIs) operate may be flawed. Firstly, the findings can be explained by the normative behaviour development patterns in children during primary school. Secondly, the GBG failed to encourage the desired behaviours to generalise outside of intervention delivery.

The reductions observed in the present study follow the normative trend reported in many longitudinal studies (for a review see Tremblay, 2010). As described in section 3.2.2, individuals are most disruptive during infancy, showing a sharp increase in aggressive behaviour and inattention between the ages of 2 and 6 (Tremblay, 2010; Brame et al., 2001). As infants mature, they develop their cognitive, motor, social, and language skills that allow children to better articulate and regulate their emotions in a more socially appropriate manner (Eisenberg et al., 2010). However, longitudinal studies also identify a small proportion of the sample, approximately 5% that consistently do not follow the normative developmental trajectory (Brame et al., 2001). Within this group, the children not only become more disruptive, but their actions escalate in extremity with age (Loeber & Stouthamer-Loeber, 1998; Côté et al., 2007). By applying the above knowledge to the context of the present study, the null findings for RQ1a can be explained. The majority of the cohort that participated were aged seven (Year 2) when baseline disruption levels were measured. This meant the majority of children would have received behaviour ratings which indicated they were not aggressive and had adequate attention spans. As there would have been a limited scope of improvement, except for those who were at risk, the GBG was unlikely to cause any reductions in behaviour that would not have occurred naturally as a result of the process of maturation. This suggests that selective interventions may be a more effective way of addressing disruptive behaviour as the children who fall under a risk factor are the individuals who have scope for behavioural improvement.

Another reason for the lack of significant interaction between the GBG and reduced DB scores could have been due to the behaviours encouraged in the GBG failing to generalise outside of intervention delivery. Qualitative evidence from the wider trial's implementation suggested that pupils were responsive to the principles of the intervention and improved

their behaviour when playing the GBG (Humphrey et al., 2018). Similar findings were also reported in the Oxford trial (Chan et al., 2012). The qualitative evidence of pupils responding to the intervention but not generalising these learnt behaviours could be due to two reasons: intervention failure or implementation failure. Intervention failure would indicate that the logic model for the GBG did not adequately facilitate the generalisation of desired behaviours to outside gameplay. Meanwhile, implementation failure would indicate that teachers were delivering the GBG incorrectly, and it is the incorrect delivery that is obstructing adaptive behaviours from generalising to outside the intervention context. In this section of the discussion, intervention failure as a potential cause for the lack of generalised adaptive behaviour will be explored, as RQ2 will focus on implementation failure (see section 7.5).

When assessing the GBG at a surface level, the intervention components would appear to facilitate the generalisation of adaptive behaviours to the wider classroom environment. For instance, the intervention was designed to encourage pupils to actively engage in the adaptive behaviours set by the rules during game play (see sections 3.4.1 & 3.4.2). The four rules were designed to reflect the everyday expectations of the classroom while completing a learning exercise that would typically be completed within the lesson the GBG was being administered in (Ford, et al., 2010). These features separates the GBG from other classroom-based interventions which have a focus on encouraging positive behaviour development where the programme content is delivered at a specific slot in the classroom timetable and is treated like a separate lesson (e.g. PATHS: Greenberg, 2010; The Incredible Years: Webster-Stratton & Reid, 2003). However, the integration of the intervention into the classroom environment, in the case of the present study, appears to have been unsuccessful at improving DB. Therefore, another aspect of the GBG may have inhibited the generalisation of adaptive behaviours to the wider classroom context.

The other unique feature of the GBG is the amount of positive reinforcement pupils receive during intervention sessions. To remind the reader the end of a typical GBG session goes as follows (Ford et al., 2010): First, the team captains places a star on the weekly winner board as the entire class celebrates each team's success. Then, each winning pupil stamps their GBG booklet that records their progress. Every winning pupil also receives a small extrinsic reward at the end of the game, and a larger weekly prize if their team wins at least one game in the week. The amount of positive reinforcement the pupils receive at the end of the game is likely to be considerably more than they receive outside of the game where children are expected to display the same behaviours. It is possible that the amount of positive reinforcement had the opposite effect to developers intended, instead of encouraging pupils to internalise the adaptive behaviours, the lack of incentives outside gameplay discouraged children from generalising desirable behaviours. In fact, teachers involved in the qualitative strand of the wider trial were concerned that the number of extrinsic rewards pupils would receive would create an expectation that they should always receive incentives for good behaviour (Humphrey et al., 2018). Moreover, some teachers

had switched to daily intrinsic rewards instead in an attempt to rectify this concern. However, a surface level change of reward type may not have been enough to prevent the expectation of incentives for everyday expectations as the celebratory nature of a GBG session would still be a contrast from outside gameplay. Developers may have to address the logic model so all the elements of the GBG encourage behaviour generalisation, rather than hinder it.

7.4.2 Are there differential effects of the GBG among children who are eligible for FSM?

The results reflected the null results reported by Kellam and colleagues (1998) whereby individual poverty did not moderate the GBG's effect, despite individual poverty being predictive of later disruption levels in middle school males. The researchers argued that individual poverty was likely to be mediated by previous ratings of disruptive behaviour, which is indicative of maladaptive behaviour management practices used within economically disadvantaged households. This raises the issue of peer influence, and whether an intervention such as the GBG can compete with the multiple sources, such as parent-child relationships, DB can be learnt from. This may mean children experiencing poverty are particularly resistant to the influence of the GBG because of their risk to increased exposure to maladaptive behaviour within the household and broader neighbourhood.

While it is well established that the relationships children have with their parents/caregivers are crucial for their social development, the association between household poverty and parenting practices have been widely documented within the literature (See Shaw & Shelby, 2014 for an overview). Researchers argue that a lack of income and resources induce psychological distress upon the caregiver(s), which can lead to inconsistent parenting practices or the use of, particularly punitive methods to exert control over the household (Neppl, Senia & Donnellan, 2016; Warren & Font, 2015). These factors have been found to moderate aggressive and attentional behaviours within children who are at risk of escalating their behaviours to delinquent levels (Scott et al., 2012; Holmes & Kiernan, 2013). Furthermore, the social structures economically disadvantaged households are part of are likely to be maintaining the transmission of maladaptive behaviours associated with disruptive behaviour. This is because individuals with low levels of social capital often live in residential areas with poor social structures that facilitate social order (see section 4.4.1). Economically disadvantaged communities can be viewed as fractured social structures that consist of small, isolated clusters which are high in strong social ties between household members that allow for the propagation and persistence of behaviours with little influence from the wider community. Therefore, if maladaptive behaviours are established within a family unit experiencing poverty, preventive interventions that address the strong social connections between parents and children would be a more appropriate method to reduce disruptive behaviour compared to a classroom-based intervention such as the GBG.

For instance, Incredible Years (Webster-Stratton & Reid, 2003) is beneficial in economically disadvantaged communities (Garnder, Burton & Klines, 2006; McGilloway et al., 2012). Another example is ParentCorps, a pre-kindergarten intervention that lasts 13 weeks (Brotman et al., 2008). Dawson-McClure and colleagues (2014) found that the intervention reduced behaviours indicative of conduct problems at age six within a high poverty community-based in the US. The intervention effect was particularly strong for children that had parents who used either harsh or inconsistent behaviour management strategies and reported high levels of daily hassles. However, schools and other educational establishments are often the delivery sites for such preventive interventions. Like classroom-based interventions, these are reliant on the weak social ties between the parent/caregiver and a school staff member or implementer, but there is the further complication whereby parents are difficult to recruit to attend and commit to multiple interventions (Roggman, Boyce & Innocenti, 2008). This suggests that parenting UPIs are likely to be no more successful than classroom-based interventions such as the GBG because of the reliance on weak social ties to disseminate behaviour changes within a socially fragmented community.

Additionally, outcomes used to test the success of these forms of prevention interventions are often based on parent's knowledge of the practices delivered during programme sessions rather than observations of parents demonstrating taught skills (Roggman et al., 2008). Although utilising self-report of knowledge is often advantageous in terms of economic savings, scores are likely to be conflated due to social desirability biases and may not reflect the behaviours parents and caregivers engage in at home (Linares, Montalto, Li & Oza, 2006). However, parenting preventive interventions have the potential to foster healthy social connections between parents, as community ties are an effective way of encouraging behaviour change (Sanders, 2000). Individuals are more likely to adopt new behaviours if they observe multiple people engaging in the behaviours, particularly friends (Centola & Macy, 2005). If parenting interventions are successful in strengthening community ties amongst parents, the promoted practices will be more successful at reducing disruptive behaviour as the interactions between parents and children will be more positive. This could then lead to improved behaviour levels in the classroom among economically disadvantaged children.

7.4.3 Does school-level poverty moderate the effects of the GBG on disruptive behaviour?

The findings indicated that the GBG was not responsible for the observed increases in disruptive behaviour for children in high poverty schools. Again, the present study replicated the null findings reported by Kellam and colleagues (1998), whereby school-level poverty did not moderate the behavioural outcomes of the GBG. This is despite their findings

164

suggesting that both classroom and school level poverty were significant moderators of aggressive behaviour in classrooms. Although the present study adapted the analytical strategy adopted by Kellam and colleagues to account for the hierarchical nature of the data, the methodological decisions surrounding school poverty statuses may have contributed to the null findings for this particular research question.

As a reminder for the reader, the researcher of the present study created discrete categories whereby schools were grouped based on the percentage of the school population's eligible for FSM. A cut off point of one standard deviation above the national average FSM uptake for a school was utilised in order to group participating schools into two categories (moderate poverty and high poverty). As discussed in section 5.6.1, the cut-off point was necessary in order to test RQ1 without violating the principles of MLM. This meant that the study was comparing high poverty with moderate poverty instead of schools with low levels of poverty. As many schools in the moderate poverty category were likely to display similar characteristics to high poverty schools, it is possible that the majority of schools within the sample experienced a level of social disorganisation which affected disruptive behaviour levels in the classroom, as well as the implementation of the GBG. Therefore, the observed null findings may be due to a lack of difference between the two poverty statuses.

7.4.4 Do school-level poverty and child-level FSM eligibility interact to moderate the effects of the GBG on disruptive behaviour?

The findings indicate that in the first year of the study school-level poverty and child-level poverty interacted to moderate the intervention effect. The interaction indicates that children experiencing poverty within schools serving a high poverty population were more likely to become disruptive if exposed to the GBG. The three-way interaction is a particularly striking finding, as the two levels of poverty did not moderate GBG outcomes when modelled separately (see sections 7.4.2 and 7.4.3). Furthermore, the three-way interaction between poverty and the GBG was not statistically significant during the second year of the study. This section of the discussion will explore two explanations for this finding, the role of peer influence and implementation failure.

Peer influence

As mentioned in section 7.6.3, children learn behaviours through multiple sources, not just the explicit feedback given by a teacher. While the parental-child relationship was identified as a key influence that may mean children experiencing poverty could require more support in order to learn adaptive behaviours, the role of peer influence may explain the significant

three-way interaction[16] in the first year of the study. This is because the peer-to-peer relations between pupils in schools with high numbers of children experiencing poverty could have further impeded the generalisation of adaptive behaviours learnt from the GBG. In this case, the developers of the GBG may have underestimated the influence classmates would have on influencing intervention outcomes.

While the developers incorporated peer influence within the GBG logic model through the use of SLT and SFT (see section 3.4.3), the main focus was the explicit ratings of behaviour given by the teacher. As a brief reminder to the reader of a cornerstone principle of the GBG logic model, individuals within a shared environment rate one another's' behaviours based on their ability to meet the behavioural expectations of the social context (Kellam et al., 1975). The intervention was designed so the teacher was clearly identifiable as the primary rater by explicitly awarded behaviour ratings by awarding infractions as well as awards (Ford et al., 2010). Meanwhile, peers within teams would model adaptive behaviour and encourage team mates to avoid infractions in order to gain positive ratings from the teacher. Yet, outside the gameplay, pupils do not have the explicit incentives to hold one another accountable for their behaviour, and so the classroom adaptive behaviours were unlikely to generalise (see section 7.4.1) for the prior argument about the GBG failing to generalise behaviour).

In fact, outside of intervention, pupils may have valued one another as more socially desirable raters than the teacher (Iago et al., 1998). This could be the case within classes where there were high numbers of children experiencing poverty within the school population, in that there were likely to be greater numbers of disruptive peers within the class influencing one another to continue or escalate their misbehaviour when the GBG was not being implemented. Without the incentives of rewards given by the teacher through the intervention, peers were more likely to reinforce their embedded maladaptive behaviours learnt from not just one another, but from home too (see section 7.6.3). This could be due to peers from similar backgrounds identifying with each other more compared to their teacher, as suggested by SLT (Bandura, 1977).Further to this, Centola and Macy (2005) argue that the number of individuals within the same environment engaging in a behaviour can influence the likelihood another person will enact the same behaviour. This could mean that in schools with higher levels of children experiencing poverty, the influence of peers with a similar socio-economic background within a classroom continuing to be disruptive

when the GBG is not implemented could cause an increase in disruptive behaviour through reinforcement.

However, the above possible explanation is only applicable to the findings from the first year of the study, as the three-way interaction between school poverty, trial condition and FSM eligibility were not significant in the final year of the study. This meant that children experiencing poverty within a high poverty school that implemented the GBG were no more disruptive than their counterparts in the high poverty schools with the UP condition. The most likely explanation for the findings in the second year is the natural development of children's cognitive abilities surrounding better control over impulsive and aggressive behaviours. As explained in section 3.2.2, children develop better executive functions with age which enable individuals to place effort into inhibiting behaviours that will not be tolerated in the classroom or channel them into more socially accepted forms such as relational aggression (Tremblay, 2010; Björkqvist, 2017). Children experiencing poverty are more likely to have higher levels of DB as a result of developmental deficits in the brain regions associated with executive functions (Eisenberg et al., 2010). These developmental differences are possibly linked to poor nutrition and stimulation due to the lack of resources in households with low income (McLoyd, 1998). Nevertheless, the majority of children experiencing poverty will still reduce their DB levels with age, albeit from a more elevated stage compared to peers that are not from impoverished households (Cooper & Stewart, 2017). Therefore, children experiencing poverty would continue to reduce their DB scores between time points irrespective of the intervention.

Implementation of the GBG

The second explanation for the significant three-way interaction in the first year could be due to how teachers in high poverty schools implemented the GBG. A core argument for the present study was that schools in most need of preventive interventions often lack the organisational capacity to implement UPIs in order to produce programme outcomes (see section 3.4.2). Of course, teachers in high poverty schools from either condition were at a higher risk of experiencing factors such as access to fewer resources, experience higher levels of stress, engaging in more "daily fire-fighting", and increased time-pressure, compared to their counterparts in the moderate poverty category. The addition of the GBG may have thus been an increase in workload that further detracted from already low resources. This argument is based on the principles of social disorganisation theory (SDT: Shaw & McKay, 1942), whereby the social structures of a school hinder the collective from exerting social order, problem-solving and appropriately supporting one another (Domitrovich et al., 2008). If the GBG was burdensome to teachers working within a high poverty school during the first year of the study, children experiencing poverty might have responded to the increased disorganisation that may have occurred as a result of implementing the GBG by increasing their disruptive behaviour.

A further indication that implementation may explain the differing findings for this RQ across the two time points is that there was a change in the teachers implementing the GBG half-way through the study. This was due to the natural way in which the UK primary school system operates, with most children receiving a different class teacher as they progress through the academic years. Thus, it is possible that within high poverty schools, teachers in the second year of the study implemented the GBG at a lower level than teachers in the first year (see section 7.4.3). If this is the case, the introduction of new implementers during the establishment of an intervention within a school setting, particularly those serving economically disadvantaged areas, may be an area the field of implementation focuses more attention on in the future (see section 7.8).

In the second year of the study, teachers implementing the GBG in the high poverty schools may have reduced their level of delivery in response to the high levels of social disorganisation experienced within the classroom and the wider school. If the implementation of the GBG is responsible for the change reported between the first and second year of the study, there may be two possible areas of influence that could account for the differences in the delivery of the intervention. The first suggestion is that this may have been an executive decision made by the teacher. The second suggestion is external pressure from more senior members of staff. The first explanation surrounds the extent to which the teachers "bought into" the intervention. As the teachers in the second year were more detached from the study and GBG before participation, they may not have shared the same enthusiasm for delivering the intervention. Some aspects concerning the intervention may have led to teachers in high poverty schools less inclined to utilise the programme in their classes, but they can be broadly categorised under social validity (Wolf, 1978). This is a particular point that is further explored in section 7.6.2.

Alternatively, teachers may have been either explicitly or implicitly directed to reduce the delivery of the GBG by their senior leadership team (SLT) after the first year of implementation. Although the findings of the first year of the wider evaluation were not made immediately available to schools, staff members that had participated in the decision-making process to become involved in the trial may have been underwhelmed by the impact of the GBG or had become disinterested in participating in the study. This lack of impact may have led to SLT either actively discouraging teachers from implementing the intervention or by providing negligible resources and support to the teachers in the second year.

As schools serving highly economically disadvantaged neighbourhoods are more likely to experience a lower organisational capacity compared to their more affluent counterparts (Payne & Eckhert, 2009), the support required for teachers to continue implementing the

GBG was likely to already under strain due to other competing priorities from the broader school population. Further to this, the perceived failure of the intervention may have further eroded the collective efficacy of the school community, mainly the implementers and key stakeholders. This could have led to the further erosion of the collective efficacy of staff members, the common belief that mutual action brings about positive change (Bandura, 1977). While low collective efficacy is associated with schools with high levels of social disorganisation (Akiba et al., 2010), the construct has yet to be established through empirical evidence as a factor that could influence implementation (see section 7.8). Even so, the above postulations may have resulted in the teachers in high poverty schools reducing the extent they implemented the GBG within their classrooms during the second year as a response to the depleted resources or demotivation attributed to high levels of social disorganisation. However, these explanations for differing results between the two time points of the study are speculative. Discussion of the results for RQ2 may provide further insight into how implementation variability may have contributed to the findings discussed above.

7.5 Summarised results for research question two

The overarching question RQ2 attempted to answer was: does poverty moderate the implementation of the GBG. The generation of implementation profiles led to a restriction in the data available for subsequent MLM analysis. Therefore, the summarised results are regarding the first year of implementation, as all cluster options are represented within that particular time point. As the MLMs utilised for analysis were constructed of three levels, with a smaller sample than RQ1, the reported results should be treated as exploratory findings due to the underpowered nature of this aspect of the study (see section 6.7.2).

7.5.1 Are there distinct implementation profiles amongst teachers implementing GBG?

A three-cluster solution was accepted over a two-cluster solution because it was considered a more valid representation of implementation variability. The three implementation profiles were: medium fidelity-low dosage (MF-LD, N= 67), high fidelity-medium dosage (HF-MD, N= 19), and high fidelity-high dosage (HF-HD, N=7). The implementation profiles were distributed unevenly across the teacher sample, with the majority of teachers fitting in the MF-LD profile, and a select few in the HF-HD profile.

7.5.2 Does implementation vary as a function of school-level poverty?

The study concluded that implementation did not vary as a function of school-level poverty.

7.5.3 Does implementation variability moderate the effects of the GBG on disruptive behaviour?

Implementation variability did not moderate the effects of the GBG on disruptive behaviour.

7.5.4 Do implementation variability and pupil-level poverty interact to moderate the effects of the GBG on disruptive behaviour?

Implementation variability did not moderate the disruptive behaviour levels of children experiencing poverty.

7.5.5 Does school-level poverty and implementation variability interact to moderate the effects of the GBG on disruptive behaviour?

The interaction between school-level poverty and implementation variability was reported as a null finding. The interaction did not moderate the effects of GBG on disruptive behaviour.

7.5.6 Do school-level poverty, implementation variability and individual-level poverty interact to moderate the effects of the GBG on disruptive behaviour?

The interaction between school-level poverty, implementation variability and individual-level poverty were reported as a null finding. The interaction did not moderate the effects of GBG on disruptive behaviour.

7.6 Discussion of findings for research question two

Many of the studies utilising an RCT design report null results for preventive interventions assume the results are due to programme failure (that is, the intervention simply 'does not work'; Lendrum & Wigelsworth, 2013). Such an assumption is dangerous to make without considering the extent to which the intervention was implemented. This oversight is called a type III error, and with the lack of utilising implementation within their analyses, it is likely to be committed at an alarming level within the prevention science literature. With this in mind, RQ2 focused on implementation variability as a moderator of intervention outcomes. While the majority of research has taken a variable-focused approach to study implementation (e.g. Ialongo et al., 1999; Pettigrew et al., 2014; Rosenblatt & Elias, 2008), the present study took a person-centred approach, as individuals implement interventions across all the subordinate dimensions at variable levels. Therefore, this section of the discussion focuses first on the implementation profiles found within the implementation data, before exploring whether membership of said profiles moderates behavioural outcomes. As highlighted in the results chapter, the reader must treat the findings for this section with caution as the statistical model was underpowered for these sets of research questions, and therefore they can only serve as an exploratory function. Issues surrounding statistical power are discussed in the appropriate section of the discussion (see section 7.7.2).

7.6.1 Are there distinct implementation profiles amongst teachers implementing GBG?

Cluster analysis established three implementation profiles. These were generated based on fid-qual scores and dosage data collected over the two years of the study (see section 5.8.3), which were then labelled based on defining characteristics within each profile. These were: Medium fidelity-low dosage (MF-LD; N=67); high fidelity-medium dosage (HF-MD; N= 19), and high fidelity-high dosage (HD-HF; N=7). Teachers clustered within the MF-LD category implemented the GBG for an average of 5 minutes a week at a fidelity rate of 68.58%. Meanwhile, teachers clustered within the HF-MD category implemented the GBG for an average of 28 minutes per week at a fidelity rate of 76.18%. Finally, the HF-HD teachers were implementing the GBG for an average of 49 minutes per week at a fidelity rate of 77.9% as the majority of the teachers within the study were clustered to the MF-LD category. In relation to the wider literature, teachers within the present study were delivering the GBG at a similar fidelity level to teachers in previous studies (e.g. 82% in Domitrovich et al., 2015; 77% in Leflot et al., 2013; 60% in Ialongo et al., 2001). For dosage, there are very few studies that have utilised such data within their analyses. However, findings from previous research vary considerably and do not indicate a particular trend in which the dosage levels of the present study support. Hagermoser-Sanetti and Fallon (2011) reported that the majority of teachers (94%) delivered the GBG for the recommended length of time, despite only 56% delivered the intervention at the frequency recommended by developers. Meanwhile, Domitrovich and colleagues (2015) reported that the average frequency of GBG delivery met the developer's implementation criteria, but in terms of average duration, teachers were not meeting the developer's expectations. The contradictory findings suggest that dosage levels are likely to be influenced by factors that affect implementation that is unique to the context of each study or more specifically, each delivery setting.

One possible explanation for the differentiation of dosage between the three implementation profiles is the teacher's attitude towards the GBG. While it is striking that the majority of teachers clustered within the MF-LD profile, there was a small number clustered within the HF-HD profile could be considered "early adopters" of the GBG. Early adopters are a group of individuals who have a low threshold for investing in innovations compared to the majority of the population (Rogers, 1995). Implementer-level factors of implementation are likely to explain the difference between early adopting teachers from the majority of the teacher sample (Domitrovich et al., 2008). As described in section 4.3.2 certain factors such as professional and psychological characteristics of the teachers may have influenced the extent to which the high implementing teachers adopted the GBG. For instance, teachers with lower levels of stress and higher levels of self-efficacy implement interventions at a higher standard compared to their counterparts (Kessler, 1999). It may be that teachers with a HF-HD profile did not feel overwhelmed by implementing the GBG and felt confident in their delivery of the intervention.

Meanwhile, the majority of teachers may have been sceptical of the GBG, as other implementer-level factors can serve as a barrier towards implementation (Domitrovich et al., 2008). It may be that the teachers did not "buy in" to the intervention as they did not believe the GBG addressed the needs of their pupils (Durlak & DuPre, 2008). Negative perceptions of the intervention can lead to teachers feeling demotivated, and therefore the lower implementing teachers delivered the GBG less frequently and at a lower quality. Further to this, classroom and school level factors of implementation may have influenced teachers in the lower implementing profiles in spite of their own attitudes towards the intervention. A prime example of this is the level of social disorganisation teachers experienced within the classroom and the wider school (Domitrovich, 2008). Teachers with an MF-LD implementation profile may have lowered the intensity of intervention delivery out of necessity in order to preserve the little time and resources available for general day to day tasks. As there is an association between social disorganisation and school poverty levels (Domitrovich et al., 2008), this particular discussion point is explored in detail later (see section 7.4.2).

Implementation profiles over time

Another striking finding was that the distributions of implementation profiles changed over time, with no HF-HD teachers in the second year of the study (see section 6.6.3). Again, to remind the reader, different teachers implemented the GBG in second year of the study. The change in in the proportions of teachers that represented the three implementation clusters changed during the second year of the study may again be explained through school-level implementation factors, such as a lack of involvement in decision-making (Aldeman & Taylor, 2002). This is because teachers in the first year of the study would have been more involved in the decision-making process in order to participate within the RCT. Participation in the decision-making process may be vital to implementers frequently delivering preventive interventions. For instance, Rogers (1995) argues that there are several stages individuals pass through in order to decide on adoption: these include knowledge seeking and weighing up the advantages and limitations of adopting the innovation. However, decision-making often takes the form of an "authority innovation decision", whereby a member of superordinate power forces the adoption of a behaviour change or technology onto a subordinate (Rogers & Shoemaker, 1971). Although commonplace within organisations, such as educational institutions, this form of decision making is often considered harmful as pressured individuals feel like their opinions are ignored and agency undervalued valued (Evans, Murphy & Scourfield, 2015). Therefore, during the second year of implementation, teachers delivering the GBG may have felt reluctant to implement the intervention frequently due to a disparity between the programme fit (further discussed in section 7.4.3) and the demands of management. This pressure may be responsible for the lowering levels of dosage in the second year of the study.

Alternatively, teachers in the second year of implementation may have felt less supported by their superordinate colleagues compared to teachers in the first year of the study. Evidence suggests that the support of senior staff members is a major factor that influences the extent to which teachers implement a preventive intervention (Gregory, Henry, Schoeny & The Metropolitan Area Child Study Research Group, 2007; Kam et al., 2003; Thapa, Cohen, Guffey & Higgins-D'Alesssandro, 2013). For instance, Payne and Eckhert (2009) reported that high levels of headteacher support was a predictor of implementation quality for the range of interventions schools from the US utilised within 'The National Study of Delinquency Prevention in Schools'. Strong support from senior staff members may send a message to implementing members of staff that the intervention is a school priority by providing teachers with the necessary support and resources (Adelman & Taylor, 2003; Schien, 2010). If superordinate colleagues reduced the level of support for the implementation of the GBG during the second year, teachers might have responded by reducing dosage levels because the school no longer perceived the intervention as a priority (Humphrey et al., 2018).

SDT can be applied to explain the possible lack of senior staff interest in the GBG during the second year of the study. Key stakeholders within the school may have purposely reduced their resources on purpose as the GBG may have been perceived as unsuccessful during the first year of implementation[17]. With resources stretched across a pupil population with high needs, the executive decision to reduce support for the intervention may have been to protect initiatives or practices which were perceived as a higher priority or were particularly successful (Domitrovich et al., 2008). On the other hand, the level of social disorganisation may have meant despite the high priority placed on the GBG, Senior staff were unable to meet the needs of their colleagues in order for them to achieve high levels of implementation. If SDT is a contributor to the intervention variability observed within the study, then the level of school-poverty may provide an insight into whether this speculation has any weight (see section 7.4.3).

7.6.2 Does implementation vary as a function of school-level poverty?

One explanation for the null result may be due to the limitations of the poverty profiles given to schools during the study (see section 6.7.2). Moderate poverty schools and high poverty schools may *both* be experiencing high levels of social disorganisationeaning, and all schools may have been unable to prioritise the GBG due to lack of time and resources. However, the findings also suggest that the levels of implementation for the GBG may have been due to implementer level factors of implementation such as teacher buy-in. Further supporting this, the levels of fidelity across the three implementation profiles indicate that the teachers knew *how* to implement the programme and that the variation in dosage could

---

[17] This is without knowledge from the wider evaluation trial in regard to any changes in disruptive behaviour.

show the extent to which the teacher had committed to using the intervention as part of their behaviour management repertoire. Although there are many reasons why the majority of teachers did not buy into the GBG, they can be organised within two inter-related categories: a disparity between intervention fit and the cultural context, and social validity of the programme.

Intervention fit and UK cultural context

Since the GBG was developed and widely tested in the US (e.g. Kellam et al., 1998; Kellam et al., 2011), the intervention was designed to meet the needs of elementary school teachers. However, there are major differences between the UK and US in terms of their education systems, as well as cultural differences in teacher-child interactions (Alexander, 2001). With such differences usually expected to pose issues for implementation, preventive interventions should go through a revision and piloting process in the new cultural context (Castro et al., 2004). This is so the intervention can address the local needs while maintaining key intervention components that are thought to encourage behaviour change. It appears that this process may not have been as extemsove for the exporting of the GBG to the UK as it had been for other countries (e.g. The Netherlands). While a pilot of the programme was deemed successful in Oxfordshire (Chan et al., 2012), it appears that the developers valued fidelity over local needs as the present study utilised the same format delivered to elementary pupils in the US. This lack of consideration for the UK cultural context may have resulted in teachers attributing the intervention as having a poor fit as the programme did not meet the local needs of UK classrooms.

However, the GBG has been tested and implemented successfully in other countries, most notably The Netherlands (e.g. van Lier, et al., 2004; van Lier et al., 2009) where a substantial evidence base has found similar findings to the Baltimore trial (Kellam et al., 2014). For instance, van Lier and colleagues (2004) found there was no main effect in reducing disruptive behaviour in the GBG condition, but there were significant reductions observed amongst males. The stark difference in findings between the present study and those reported by Van Lier and colleagues could be attributed to the adaptions made to the GBG when exported to The Netherlands, with deeper modifications made to the intervention (see section 4.5.1 for a comparison). This lack of consideration for the local needs of teachers implementing the GBG within UK classrooms may be due to an assumed shared educational culture because there was an absent language barrier. Such a culturally blind approach to importing a preventive intervention is likely to lead to decreased participation and outcome effects (Kumpfer, 2002). Therefore, developers of the GBG should consider revising aspects of their intervention in order to meet the needs of teachers delivering the programme in UK classrooms to ensure higher levels of sustained implementation in future studies (see section 4.6).

The social validity of the Good Behaviour Game

A related construct to cultural fit is social validity. Social validity is where stakeholders attribute value to the intervention based on the extent to which it meets the needs of the social group in which it operates (Wolf, 1978). Traditionally, three subordinate constructs are utilised in order to determine an intervention's social validity within implementation science, these are acceptability, feasibility, and utility (Humphrey et al., 2016). The acceptability of an intervention is determined by whether the outcomes promised by the programme are perceived as needed or wanted (Wolf, 1978). It is possible that teachers delivered the GBG at a low dosage because they felt that their behaviour management practices were sufficient for controlling the disruption levels of the class.

Meanwhile, researchers assess the feasibility of an intervention by evaluating whether the programme components are sufficient for the environment it will operate in (Dusenbury, 2003; Dane & Schneider, 1998). A teacher will likely consider their own needs, the needs of their class and the ease of use when examining feasibility (Durlak & DuPre, 2008). In the case of the GBG, teachers may have implemented the intervention at a low frequency because one of these factors may have compromised their ability to utilise the intervention in order to reduce disruption levels in the classroom. Finally, the utility is assessed by considering the degree to which the effort required to implement the intervention is "worth" the outcome (Humphrey et al., 2016). The majority of teachers implementing the GBG may have perceived the input, i.e. the amount of effort required to plan and deliver the intervention was outweighing the decreases in disruptive behaviour and therefore reduced the level of dosage, irrespective of the poverty status of the school.

However, social validity within implementation science has taken an individualistic stance on the construct by focusing only the teacher's opinion on the extent to which an intervention meets local needs successfully. Meanwhile, contagion theory also stresses that social validity is important for the adoption of an innovation (Centola & Macy, 2005). The emphasis is on whether the surrounding people an individual knows have adopted the innovation too (Centola & Macy, 2005). This is because a high number of individuals within one's social network demonstrating success with the innovation gives it credibility, particularly if the reports of success are from close social connections (Dodds & Watts, 2004). For instance, research into the spread of medical technology revealed that many were unwilling to uptake the innovations until a trusted colleague had been seen using the same technology (Coleman et al. 1983). Therefore, the design of the trial may have implicitly impacted teacher's assessment of the GBG during the study as the majority of the staff were not participating in the study. If teachers did not have access to many examples of intervention success, then the teachers may have felt that the GBG lacked credibility and lowered their level of implementation as a result of this. Although it can only be speculated the extent to which such factors may have led to the majority of teacher's reducing their level of

175

implementation for the GBG, this highlights the need for implementation process evaluations to occur alongside feasibility and effectiveness studies in order for social validity to inform future developments within implementation science (see section 7.8).

7.6.3 Does implementation variability moderate the effects of the GBG on disruptive behaviour?

When compared to MF-LD, the differences in pupil-level disruptive behaviour for HF-MD and HF-HD implementation profiles were not statistically significant. However, as highlighted before, the combination of a reduced sample size with an additional structural level within the analysis meant that the research question was underpowered. While this means that the null findings cannot definitively indicate that implementation variability does not moderate intervention outcomes, the discussion of these findings can provide an exploratory function in order to direct future research (see section 7.8). For instance, the reductions in disruptive behaviour for HF-HD and HF-MD appear to improve based on the increase in dosage. This observation tentatively supports a widely used argument within implementation science, whereby implementers who are more invested in the intervention are more likely to perceive a benefit from it (see section 7.4.3 for further discussion on this point), and a cyclical relationship is formed, sustaining the implementation of the intervention as well as it's desired outcomes (Langley et al., 2010). However, when situating this finding about the broader results of the present study, even if a teacher has bought into the GBG, implementing the intervention to the specification desired by the developers, or exceeding it, would not necessarily improve the disruptive behaviour within the classroom.

This may be due two issues discussed previously in sections 7.4.1 and 7.4.3: the failure of the GBG's logic model to allow adaptative behaviours to generalise to outside gameplay and the influence of peers on classroom behaviour. Therefore, the intensity in which a teacher delivers the GBG is irrelevant in reducing disruptive behaviour within the classroom as they do not incentivise adaptive behaviours to the same extent as outside of the game. It would therefore seem an intervention that encompasses the teacher's every day practice would be more appropriate for DB, such as a whole school intervention. There are many preventive interventions which have taken a whole-school approach to address disruptive and anti-social behaviours such as KidsMatter (Graetz et al., 2008) and The Olweus Bullying Prevention Programme (Olweus, 1993). Both of which have substantial evidence bases that suggest pupils improve their behaviour in accordance to programme outcomes (For KidsMatter see: Hahn et al, 2010; For Owleus Prevention Programme see: Ttofti & Farrington, 2010). For instance, Nielsen and colleagues (2015) reported that the whole school SEL intervention *Up* significantly increased adaptive emotional regulation behaviours in pupils across the school. However, whole school interventions are often complex as multiple strands require implementation and are no more effective than classroom-based interventions (Dusenbury et al., 2003; Greenlaugh et al., 2004). This

suggests that although schools high in social disorganisation could potentially benefit from them, may be unable to sustain a whole school preventive intervention.

7.6.4 Do child-level poverty and implementation variability interact to moderate the effects of the GBG on disruptive behaviour?

The study concluded that the interaction between child-level poverty and implementation variability was not responsible for the observed disruptive behaviour scores during the first year of the study. However, for FSM eligible children in classes with teachers implementing the GBG to a higher standard (HF-MD and HF-HD), their disruptive behaviour appeared to increase compared to their counterparts in the MF-LD group. As this sample was underpowered for this particular research question, it is possible that there was a significant interaction effect which the MLM was unable to detect (see section 7.7.2). But it is more likely that moderate poverty schools and high poverty schools were likely to have a large degree of similarity in terms of school level and class level factors that affected the implementation of the GBG. In that many schools were still above the national average FSM uptake and therefore many moderate schools could have been experiencing a level of social disorganisation (this issue is further discussed in section 7.4.3). Thus, at-risk groups such as children experiencing poverty may have been more likely to increase their disruptive behaviour levels as teachers in higher implementing profiles (HF-MD; HF-HD) focused their energies on the GBG, leaving fewer resources reserved for general behavioural management outside of the intervention. Teachers within the MF-LD category may have also reduced their dosage levels in order to be able to conserve resources which could be directed to priorities which were valued higher by the school and did not have increases in disruptive behaviour as a result (see section 7.4.1).

Further to this, intervention failure could also be responsible for the null findings. As discussed in RQ1 (see section: 7.4.1), despite teachers reporting in the wider trial that the children were responsive to the GBG when it was being played, the quantitative data suggested the adaptative behaviours were not generalising to beyond the game (Humphrey et al., 2018).

7.6.5 Do school-level poverty and implementation variability interact to moderate the effects of the GBG on disrupt

It was concluded that school-level poverty and implementation did not interact to moderate the intervention outcomes during the first year of the study. This finding provides further support that the observed delivery variability may be a reflection on implementer level factors of implementation, as discussed previously in section 4.3. One such factor is intervention buy-in, as teachers within high poverty schools were still able to gain HF-HD profile despite the higher levels of social disorganisation present within the school and classroom (see section 7.4.3). Another implementer level factor could be related to

competencies in skills such as general behaviour management in the classroom. Teachers within the HF-HD implementation cluster may have stronger behaviour management techniques which allowed them to manage disruption levels within the classroom outside of intervention sessions. Therefore, HF-HD teachers may have reported that the children in their class were better behaved compared to teachers implementing the GBG at MF-HD.

Another set of implementation level factors touched on previously in the discussion are psychological characteristics of teachers, such as self-efficacy, and how they can affect the delivery of interventions (see section: 7.6.1). While the literature indicates teacher self-efficacy is a crucial attribute that influences implementation levels (e.g. McCormick et al., 1995; Rohrbach, Graham & Hansen, 1993), there are examples where this argument can be challenged (e.g. Baker et al., 2010; Swift et al., 2017). This may indicate that high levels of implementation may not be a product of efficacies surrounding intervention delivery; they may instead be an indicator for wider teaching self-efficacy. Within high poverty schools, teachers in the HF-MD cluster may have had stronger positive beliefs in their teaching and behaviour management skills compared to their counterparts in the MD-HF category. High implementing teachers may have felt the addition of the GBG to their workload was manageable despite the high levels of social disorganisation within their school. Meanwhile, in the MD-HF cluster, teachers may have been relying on the GBG as a means to improve their behaviour management skills. But due to a lack of intervention effect (as established in section 7.3), the increase in disruption levels for high poverty schools within this implementation profile may reflect their poor behaviour management outside of intervention sessions. However, as these are exploratory findings, further research is required to determine if the psychological characteristics such as self-efficacy is a protective factor for teachers implementing interventions within schools of low organisational capacity.

7.6.6 Do school-level poverty, implementation variability, and individual-level poverty interact to moderate the effects of the GBG on disruptive behaviour?

The findings indicate that changes to disruptive behaviour in FSM eligible children within high poverty schools were not due to the implementation standard of the GBG. However, as previously mentioned, the sample was underpowered, and so it is possible that a significant interaction did occur but was not detected due to the lack of sensitivity of the MLM (see section 7.5). However, the findings also provide support for further examination of teacher self-efficacy levels within schools that have high levels of social disorganisation, as teachers that fit the HF-HD profiles had a larger reduction disruptive behaviour for children experiencing poverty compared to their HF-MD and MF-LD counterparts. As the overall findings of the study suggest that despite the GBG failing to change disruptive behaviour levels, independent factors are likely to be responsible for any observed changes (e.g. section 7.4). However, the identification of a small group of teachers able to implement

the GBG to a high standard despite the low organisational capacity compounding the distribution of resources, suggests that there may be a common factor that allows HF-HF teachers manage this addition to their workload. This common factor, may provide further insight into improving the overall behaviour management systems that schools utilise within economically disadvantaged areas, and could provide an alternative to the use of preventive interventions (see section 7.8.). Therefore, the changes of disruptive behaviour levels across the implementation profiles within high poverty schools may have been a reflection of teachers' behaviour management skills used outside of the intervention sessions.

7.7 Limitations of the present study

Researching human behaviour requires compromising on both methodological and conceptual decisions in order to produce a robust piece of research that uses feasible methods within the constraints of economic and time pressures. With such compromises, limitations are inevitable and should be acknowledged in order to interpret the current findings. Such acknowledgements will allow the reader to make a relative assessment of the credibility of the present study alongside side the contributions to knowledge the conclusions provide. Within this section of the discussion, both the methodological and conceptual issues of the present study will be examined.

7.7.1 Methodological issues

As there were methodological limitations at each stage of the research process, the issues will be discussed according to four broad categories: issues relating to sample, issues relating to research design, issues relating to data collection, and issues relating to data analysis.

The sample

The sample of schools and pupils may not be fully representative of the wider population. For example, most participating schools exceeded the national average of FSM uptake. Other methodological decisions regarding research design and data analysis were affected by the poverty levels observed in the sample (see below). As the present study relied on the sample derived from the wider evaluation of the GBG, the researcher was limited in their ability to address this issue (see sections 5.2.1 and 5.5.1). Schools experiencing high levels of disruptive behaviour may have been more likely to apply to the evaluation, as the chance to gain access to an intervention which could improve that particular outcome would have been seen as an incentive. As children experiencing poverty are at higher risk of displaying disruptive behaviours in the classroom, the number of schools with high proportions of FSM eligible pupils that participated in the trial indicate a particular need for behaviour management assistance for schools serving economically disadvantaged areas. A further compounding factor for the lack of broader SES representation in the study is that schools in affluent areas may not have perceived a need to participate in the RCT.

While methodological alternatives could guide future research that further examines the differential gains of children experiencing poverty in terms of UPI outcomes, or the implementation of preventive interventions, these solutions would not have been appropriate for the present study. For instance, the researchers in the wider trial could have used other sampling strategies such as stratified sampling in order to ensure that schools serving more socio-economically varied areas were participating in the study. However, such alternatives have major limitations that could have impacted other strengths of the study. As an example, the time and resources required to find the correct number of schools per strata would have led to an unfeasible research design, as either violation of core methodological assumptions concerning MLM would have jeopardised the robustness of the study, or the number of schools required to meet the assumptions would be unrealistic to achieve. Therefore, the present study accepted the lack of representativeness of the sample at the school level, in order to maintain the overall robustness of the research design.

Attrition is a common limitation to longitudinal studies such as the present one (Giraudeau & Ravaud, 2009). Although the present study did not experience school-level attrition, pupil level attrition was observed. In this case, the attrition was caused by children leaving their local school. For RQ1, 81% of the data pupils at T2 had complete cases; meanwhile, 75% of pupils had complete cases of data for RQ2 (see section 6.2.1). In order to ascertain the extent to which the missingness of data had a cause, missing data analysis was performed for both RQs. For RQ1, incomplete cases were missing at random[18] while for RQ2, incomplete cases were missing completely at random[19] (see section 6.3 for further details concerning the missing data analysis). While attrition is an apparent limitation, the missing data analysis and the use of multiple imputation indicated that the attrition did not impact the findings presented by the current study.

Research design

The primary limitation concerning the present study's research design was the distribution of FSM eligibility across the two conditions of the RCT. Although the use of minimisation during the randomisation process was meant to ensure that the distribution of FSM eligibility was similar, there was a disparity between the conditions, with more children being eligible for FSM in the intervention condition. This issue arose because the FSM uptake of schools expressing an interest in participating in the wider evaluation did not accurately reflect the schools that eventually consented and were then subsequently randomised in the RCT.

[18] The probability of missing data for a variable may be dependent on other measured variables (Pampaka, Hutcheson & Williams 2016).
[19] All cases have the same probability of being missing.

However, this disparity was considered statistically insignificant (F (1,74) = .06, p= .80) when the school-level variance was analysed as part of the analytical strategy (see section 5.5).

Data collection

The lack of variables utilised to fully account disruptive behaviour could be considered a primary limitation of the present study. While there is evidence that a variety of explanations account for displays of disruptive behaviour, it would be beyond the present study's scope to collect data on all contributors towards the target construct. For instance, there are biological influences, such as hormones (Book et al., 2001), or neighbourhood influences such as exposure to violence (Patchin et al., 2006) which contribute towards fully accounting for disruptive behaviour (see section 3.2 for a more detailed discussion). While the present study acknowledges the above contributors are important variables, it would have been unfeasible to collect data that presented each variable. This argument surrounding variable usage could also be extended to measures of implementation, whereby additional variables could have been utilised for creating implementation profiles with HCA for RQ2. However, the variables selected as proxies for disruptive behaviour and implementation are based on the conceptualisation of these constructs. Therefore, it may be that the conceptualisation of disruptive behaviour and implementation may have been inappropriate for the present study (see section 6.5.2 for a discussion of the conceptual issues of the study).

Collecting disruptive behaviour data from teachers instead of utilising self-report or parent informed surveys brings natural limitations. Firstly, it is unlikely that the same teacher provided ratings across all time points. It could be argued that any observed change in behaviour levels was due to changes in the rater, as teachers may interpret disruptive behaviour levels differently. However, this change was constant across both conditions of the RCT, and so this would not have influenced the results. Further to this, the psychometric properties of the TOCA-C have been considered to be robust, as the measure has high internal reliability and strong factor structure over time (Koth, Bradshaw, & Leaf, 2009), suggesting that any observed changes in behaviour are unlikely artefacts of rater differences.

Secondly, fatigue effects may have contributed to observed behaviour changes throughout the study. The wider trial collected data on a broader set of outcomes compared to the present study (see section 5.2.1), which meant teachers had to fill out several psychometric tests for all participating children within their class. Although the present study utilised the disruptive behaviour subscale of the TOCA-C, the data may be inaccurate due to the fatigue of filling in multiple questionnaires. Even so, teachers could stop when feeling tired, and return at a later point as the questionnaires were online. Additionally, teachers were able to share the 'survey burden' with other members of staff, such as TAs, as the only expertise required to answer the questionnaires was their knowledge of the children being rated.

Finally, the use of the electronic scoreboard to collect dosage data may have led to inaccurate measures of GBG exposure. The electronic scoreboard was designed so teachers could record infractions, the length of intervention sessions, and the frequency of play in real-time. Teachers could also add such data post-game if they did not utilise the scoreboard. However, this function relied on teachers to remember to input the data afterwards and so it may be that teachers with low levels of dosage were playing the GBG more frequently than observed in the data. Although there were other forms of dosage data available from the wider trial, such as teachers' self-reports on their frequency and duration of game-play, the electronic scoreboard was the most accurate source. This is because evidence suggests that teachers are positively biased when reflecting on the level at which they implement an intervention, which may be due to social desirability (Hansen et al., 2014). Therefore, low dosage levels reported in the present study are most likely attributed to a lack of teacher buy-in or a lack of time and resources (see section 7.4).

Data analysis

A major limitation concerning the data analysis was that the sample utilised in RQ2 did not have the power required to provide an explanatory function for the present study. Three factors contributed to this: sample size, the addition of a level within the MLMs, and the use of complete cases in cluster analysis. As RQ2 focused specifically on implementation, data from the control condition of the RCT was not required as they were not delivering the GBG. Further to this, as teachers implemented the intervention, it was possible for multi-form schools to have teachers implementing at different levels. This meant an additional level had to be introduced to the MLM to account for the nesting of classrooms. However, with every addition of a level requires an increase in sample size in order to detect small changes in effect sizes, and this was not achieved. Finally, further reductions to the dataset were required in order to generate implementation profiles through cluster analysis. HCA required complete cases, and some teachers had missing data in one or both variables used to measure implementation. Although these issues prevented RQ2 from providing an explanatory function within the present research, the findings still provide an essential contribution to knowledge. As implementation has not been investigated using these particular methods, future studies could focus on designing studies which would have enough statistical power to provide further insight into implementation's role as a moderator of preventive intervention outcomes (see section 7.6).

Also, the cluster solutions utilised for RQ2 could be considered a limitation of the present study. Firstly, using HCA over other available clustering methods led to less than optimal usage of the available sample within the intervention condition. As mentioned previously, this was because complete cases were required in order to generate implementation profiles. However, alternative clustering methods, such as LCA, were not suitable for the present study as the creation of implementation profiles required producing distinct

categories between groups of participants (see section 5.8.3). Therefore, the loss of data was a compromise made in order to be able to categorise teachers into discrete cluster solutions. The second issue relates to the cluster solution chosen to represent the implementation profiles of the present study's sample. While it was determined that there were two viable cluster solutions generated, the favoured three-cluster solution led to the MLM analysis being restricted to the first year of implementation only. This would not have been the case if the two-cluster solution had been utilised instead. However, the three-cluster solution indicated that dosage was a clear factor that separated teachers into these distinct profiles, thus this finding was an important contribution to knowledge that could not be ignored (see section 7.6).

Finally, the two school-level poverty categories were not comparing economically disadvantaged schools with more affluent schools. This particular issue was discussed in detail in section 7.4.3.

7.7.2 Conceptual issues

The three concepts integral to the present study had issues surrounding the definition and operationalisation of constructs and variables, often within debates that still rage. The concepts of poverty, disruptive behaviour and implementation will be examined in turn, to acknowledge their associated limitations.

Poverty

The conceptualisation of poverty is a contentious area, with many schools of thought devoted to debating the constituent constructs that form the concept. As investigated in chapter one, there are two main perspectives surrounding the different concepts of poverty: the needs perspective and the income perspective (see section 2.2.1 and 2.2.2). The needs perspective argues that an individual's or neighbourhood's poverty level can be determined by the lack of basic resources, or restricted access to particular services and opportunities (Ringen, 1988). Meanwhile, the income perspective argues that discrete categories of income level can be generated through the use of cut-off points based on the distribution of a given population's income. Arguments that rage between the two perspectives in the quest to establish themselves as the "correct" conceptualisation of poverty are futile, as the epistemological stances of the two perspectives lie at the heart of the debate (Shaffer, 1996). As the researcher of the present study subscribed to a post-positivist view, an objective stance was therefore adopted for poverty, which resulted in the use of income related proxies to provide commentary on trends displayed by a sample and generalise findings to the broader population. Therefore, the income perspective was the more appropriate conceptualisation of poverty for the study.

Based on this argument, FSM eligibility was the proxy utilised to represent child-level poverty, as it is widely utilised within education research. Although FSM eligibility is an

objective marker for child poverty, it is not a perfect operationalisation of the present study's definition of poverty, which is 60% of the median of the national income (Department for Work & Pensions, 2016). FSM eligibility utilises a separate income criterion which is more conservative than study's definition, which means it is likely that a proportion of participants who were not eligible for FSM but were experiencing poverty as according to the definition (see 'relative poverty in the present study' within section 2.2.2). Although it cannot be known how many participants fall into this category, the risk for such participants to be wrongly categorised as not experiencing poverty would have been higher in schools serving economically disadvantaged neighbourhoods. This issue further compounds the representativeness of the sample and pulls into question the accuracy of the findings of the present study, this is because comparisons drawn between children who are not eligible for FSM may not be the same as individuals who genuinely are not experiencing poverty (as according to the study's definition). Nevertheless, the alternative proxy, the IDACI, would have been less accurate at determining the poverty status of children than FSM eligibility, as the measure utilises the participants' household postcode. This meant that children experiencing poverty in wealthy areas, and children not experiencing poverty in poor areas would be assigned the incorrect poverty status, as the IDACI at best represents the probability of experiencing poverty rather than attempting to classify an individual's poverty status. Therefore, FSM was the best choice of proxy for child-level poverty despite the more restrictive criterion it employs to establish eligibility.

Meanwhile, a conceptual limitation of school-level poverty surrounds the extent to which the proportion of FSM eligible pupils within participating schools was an appropriate proxy in the context of SDT. The researcher applied the theory the present study in order to explain the role of poverty in moderation implementation levels of preventive interventions. However, SDT places focus on the organisational capacity of the school community, which is a separate concept from poverty. It could be argued that the present study was investigating the effects of child-level poverty and the organisational capacity of the school using a crude proxy for the school-level construct. Although several measures such as the Staff Development and School Climate Assessment Questionnaire (SDSCAQ: Zigarmi, Edeburn Blanchard, 1991) can provide more reliable and valid assessments of school organisational functioning, these were not utilised by the EEF evaluation, which meant the present study was restricted to the numerical representation of FSM eligibility within school populations. While SDT can be considered an extension of SCT, whereby the social networks are equally crucial to other forms of capital in contributing towards poverty levels within individuals and communities, future research should consider building upon the conceptual limitation of the current study by utilising organisational capacity as a construct within studies.

Disruptive behaviour

In the past, DB has been poorly defined, leading to its conflation with constructs such as "aggression" and "conduct problems" (see section 3.2.1). Despite calls for an objective definition to be formulated (e.g. Wheldall & Merrett, 1988), the term has been liberally applied to variables that already embody well-established constructs within the literature base concerning behavioural problems. While the researcher has defined DB as a set of behaviours that cause interruptions to one's own or other's learning in the classroom, this conceptualisation of the construct has not been verified with the consensus of researchers within the field. Therefore, DB's conceptualisation as a set of low-level externalising behaviours may not complement the current understanding of behaviour problems within children and adolescents. However, the suggested formulation of DB is building upon an already existing literature base for externalising behaviours. By providing refinement to a term that is used broadly, it can be considered an attempt to address current issues with the construct in order to provide future research with clarity.

Meanwhile, teachers could have evaluated the disruption levels in their classrooms using different conceptualisations of DB from the current study, and from each other when rating children's behaviour. Teachers have their unique standards and expectations of behavioural conduct within the classroom (Cooper, Smith & Upton, 1994). Changes in DB over time could, therefore, represent the tolerance levels of teachers rather than actual changes in behaviour (Cooper et al., 1994). Even so, the current study's conceptualisation of DB is informed by research that indicates that low-level behaviours within the classroom are an issue for teachers, due to the high frequencies of interruptions to learning such displays make (Wheldall & Merrett, 1988). While this suggests that there is a collective agreement among teachers on the nature and form of DB, utilising a psychometric tool such as the TOCA-C (Koth, Bradshaw, & Leaf, 2009) further reduces the likelihood of subjective interpretations occurring.

Finally, the act of assessing children's behaviour may lead to biases that affect specific sub-groups of participants, with such individuals receiving inaccurate behavioural ratings because of characteristics such as age, gender, ethnic background, and poverty status (Papatheodorou, 2005). This meant that teachers may have unconsciously shown a negative bias when rating the behaviour of children from the aforementioned risk groups. Changes in the disruptive behaviour of participants experiencing poverty may not have been accurate, pulling into question the extent to which the increases in DB amongst FSM eligible children were valid. Although other sources of behaviour rating such as parental or self-report were unlikely to be swayed by such demographic biases, the risk of inflated scores due to social desirability are higher (Taylor et al., 2011). Therefore, the assessment of DB from a teacher can be considered a more valid source of data for the present study compared to other sources.

Implementation

A pertinent limitation to acknowledge is the extent to which the conceptualisation of implementation was appropriate for the present study. This is because the proposed formulation which formed part of the contextual framework did not translate as anticipated to the data analysis stage of the research. Although it was originally argued that implementation was concerned with processes involving the practical application of an intervention (Lendrum & Humphrey, 2012), the micro-view of the term was favoured over the macro-view (see section 4.2.1). The micro-view argues that implementation is comprised of subordinate constructs (e.g., fidelity, quality of delivery, dosage, programme reach, programme reach, and adaptations). The present study intended to operationalise three of the dimensions of implementation (fidelity, quality of delivery, and dosage) as three discrete variables that could be utilised to create implementation profiles for the participating teachers (see section 5.8.3). However, during data screening, it transpired that fidelity and quality correlated too highly to be considered separate constructs. The two variables were transformed into a "fid-qual score", which was used alongside the dosage variable in order to perform cluster analysis. This high correlation between fidelity and quality suggests that the macro-view of implementation, whereby the term is a broad construct that is dependent upon an implementer's delivery meeting the developer's stipulations (Humphrey et al., 2016), would have been a more appropriate conceptualisation. However, as argued in section 4.2.4, although it is possible to collect quantitative data on all the dimensions of implementation, some subordinate constructs, such as participant responsiveness are better assessed by way of qualitative data. Quality is likely to be such a dimension, as it has been difficult to define, with vague definitions such as "how well" and intervention is delivered (Dane & Scheiner, 1998). Despite utilising a robust implementation checklist for the GBG to collect implementation data, the present study suggests that a post-positivist approach to subjective constructs such as quality is not the most appropriate course to take when conducting implementation research.

7.8 Recommendations from the present study

This section provides suggestions that will allow various facets of society and the research community to continue to develop their work concerning poverty, prevention science, as well as the implementation of preventive interventions. These will be addressed in two parts, first the implications of the present study, and recommendations for future channels of research.

7.8.1 Implications for the present study

For the research community, the suggestions from the present study can be broadly categorised into theoretical and methodological implications. The theoretical recommendations are concerned with building upon the current understanding of using preventive interventions to reduce behavioural problems within schools. Whereas the

methodological implementations will improve understanding in the areas of implementation science and research concerning poverty. Meanwhile, there is a growing interest amongst policymakers and head teachers to apply preventive interventions within educational settings in order to prevent maladaptive behaviours escalating into costly and time-consuming negative outcomes both within school environments and for wider society (Gough, 2013). However, the findings of the present study indicate that classroom-based interventions may not be the most effective form of preventive intervention for children experiencing poverty or for schools serving economically disadvantaged neighbourhoods.

Theoretical implications

When placing the findings of the present study alongside other RCT trials of the GBG, (e.g. Kellam et al., 2011; van Lier et al., 2004), it appears that the logic model of the intervention requires revising. The problem theory, Social Field Theory (SFT: Kellam et al., 1975), argues that individuals pass through a series of social fields, such as school and work. In these fields, there are expectations on how to behave, and others in the same field rate the individual's ability to succeed in these tasks (see section 3.4.4). Pupils that are consistently rated with a low social adaptation status (SAS) are more likely to continue through future social fields with a low ability to meet social expectations (Kellam at al., 1975). The GBG is thought to promote SAS by providing explicit feedback from the teacher on clearly defined expectations of behaviour. Both SFT and GBG ephasises hierarchical structures when behaviour rating occurs, with authority figures such as teachers, overseeing the assignment of adaptation statuses. Although there is some acknowledgement of peers acting as a source of rating, the influence of peers is not fully credited. However, the current findings suggest that the social influence of peers within the classroom is stronger than given credit for. This is particularly the case for schools serving economically disadvantaged communities where disruptive behaviour increased despite the implementation of the GBG (see section 7.3.3). The number of raters within a social field may have more influence in behaviour modification compared to the status of the rater. In the case of classrooms with high proportions of FSM eligible pupils, children are more responsive to each other's task demands, rather than those set by the teacher. Therefore, the GBG logic model should be revised, so the majority of raters are in agreement of the task demands required to be successful in the classroom.

This call for acknowledging the influence of the majority of participants can be extended to the development of preventive interventions more generally. The current conceptualisation of UPIs has largely ignored the social structures in which the programmes operate. This has led to an oversight of an important factor individuals account for when modifying their own behaviour: whether peers are also engaging in the behaviour change. Peer influence has been found to be a key catalyst for behaviour changes across many facets of life and society, including social movements (e.g. McAdam, 1988), improvements in health behaviour (e.g. Pulerwitz, Barker & Segundo, 2004), as well as buying habits (e.g. Crane,

1999). However, current preventive interventions tend to focus on equipping recipients with the knowledge required to change behaviours rather than building a social network where the new behaviours are the norm. While many would argue that is precisely the purpose of the preventive intervention, the fact that dosage levels decreased in the present study suggests otherwise. Even so, the field of HIV prevention has recognised the importance of preventive interventions complementing the social structure it is operating in, as opposed to competing with it (Latkin, 1995).

For instance, Kelly and colleagues (1997) utilised a community level design when implementing an HIV preventive intervention within cities known to have high numbers of homosexual males at risk of the disease. An integral part of the intervention was that well-liked members of the local gay community disseminated knowledge of HIV prevention practices through casual conversation in popular locations. It was reported that compared to placement of educational materials in similar locations, safe-sex practices significantly increased within the experimental condition. However, school-based interventions that aim to improve behavioural and SEL outcomes are presently not designed to complement the social structures in which they operate. By acknowledging that behaviour changes rely on social influences based on the number of individuals rather than the status of individuals, the designing and implementing of UPIs will require revisions in order to be more effective within the social fields they operate in. For instance, classroom-based preventive interventions should not be limited to one cohort but implemented as a framework across the entire school, so individuals are exposed to the desired behaviours from multiple sources, providing the social persuasion required for behaviour adoption. Therefore, the educational arm of prevention science should consider integrating the literature surrounding social structures into their theoretical frameworks when researching preventive interventions.

Methodological implications

The primary methodological implication of the present study is the use of implementation profiles to provide valuable insight into the variation of implementation across samples, as well as across implementation dimensions. This method of establishing implementation levels is vastly different from common practice within the field, where individual variables that measure implementation dimensions are analysed in isolation (Aber et al., 1998). However, current approaches to implementation measurements are reductionist, as teachers deliver an intervention using all the implementation dimensions (i.e. fidelity, quality of delivery, dosage, participant responsiveness and adaptations) with various degrees of success. Meanwhile, a person-centred approach to implementation acknowledges that implementers may have strengths in certain areas of implementation while lacking in others (Aber et al., 1998). Research taking a person-centred approach to implementation should utilise multiple dimensions to build a more detailed picture of how an intervention was implemented, as the insight into implementation variability could be particularly useful if

implementation failure was responsible for the lack of intervention outcomes. In addition to this, utilising implementation profiles within MLMs may provide a valuable tool for studying implementation variability and its effects on intervention outcomes[20], an area which requires more research (Ozer, 2006). Therefore, the research community should consider utilising a person-centred approach when investigating implementation levels, particularly in the form of implementation profiles.

Implications for policy

Policy-makers have been particularly interested in incorporating preventive initiatives within their policies, as the upfront costs of these interventions provide a strong return on investment in the future by reducing the number of individuals that pass through the criminal justice system, require mental health treatment, or require financial support (Gough, 2013). Children experiencing poverty have been identified as a risk group for such outcomes (see McLoyd, 1998 for an overview), with escalating disruptive behaviour levels acting as a pathway (Tremblay, 2010). As children experiencing poverty represent a considerable proportion of the child population, the at-risk group has been identified as a potential candidate that could greatly benefit from classroom-based preventive interventions (Webster-Stratton, Reid & Stoolmiller, 2008). However, the present study, which is the first to have a primary objective to assess the differential gains of a behavioural intervention for children experiencing poverty, concluded that the at-risk group was unlikely to benefit from the programme being tested (see section 7.3.1). This is because children living in persistently economically disadvantaged households are more likely to be exposed to maladaptive parenting practices (Shaw & Shelby, 2014). If these parenting styles are sustained throughout an individual's infancy, children experiencing poverty are more likely internalise disruptive behaviours, which can lead to higher levels of children from this at-risk group less prepared for the social expectations required to succeed in formal education (see section 7.3.1).

Instead, policymakers should consider promoting the use of parenting interventions within economically disadvantaged communities in order to reduce the risk of developing behavioural problems within children experiencing poverty. The primary reason parenting interventions are a promising alternative is that by providing parents with adaptive parenting strategies, punitive or inconsistent practices are less likely to be used during infancy (McGilloway et al., 2012), and so children are less likely to be aggressive or inattentive when they start school (Welsh, Bierman & Mathis., 2014). This means children experiencing poverty would be better able to meet the behavioural expectations of the classroom as intervening during the critical period of development is more likely to improve disruption levels compared to after infancy (Tremblay, 2010). Additionally, parents within the same neighbourhood could form support networks with each other by attending parenting

[20] Providing the sample is powered adequately.

sessions. Such social ties between participants involved in a programme could reinforce the use of adaptive parenting practices outside of the intervention, as well as providing support during periods where the stresses of parenthood and low income become overwhelming. These additional benefits would be particularly important in promoting community ties within economically disadvantaged neighbourhoods as the social disorganisation of such locations leave parents isolated (Putman, 2001). Therefore, policy should consider using a location central to the community for parenting interventions in order to encourage the reversal of social disorganisation within the neighbourhood.

Implications for schools

Schools have become particularly interested in implementing preventive interventions in order to address problems within the pupil population. This is fuelled by the increasing opportunities provided by evidence-based policy schemes such as the EEF. However, schools are often unaware of the time and dedication to implementing a preventive programme in order to access the benefits promised by the intervention (Durlak & DuPre, 2008). This has led to an almost paradoxical situation within preventive science, as the schools in most need are often unable to commit to the regular implementation of an intervention due to a lack of organisational capacity and already stretched resources. Therefore, leadership teams that are considering the introduction of a preventive intervention within their school should examine in detail the organisational capacity of the staff, in order to assess if the implementation levels of the intervention can be realistically met. Schools should then address organisational issues by focusing on the following areas: improving communication between staff members; promoting collective efficacy; encourage problem-solving amongst colleagues; and streamlining existing tasks, policies or processes in order to reduce the burden for teachers (King & Bouchard, 2011). However, schools that focus on improving their organisational capacity by strengthening the existing social structures between staff may find that the need for a prevention intervention is reduced after addressing these issues.

A suggestion for schools interested in providing further behavioural support to children experiencing poverty would be to assist in the promotion of parenting interventions within the community. However, such support given to these interventions should be within the limits of the school's organisational capacity. This is particularly the case for schools serving a highly economically disadvantaged neighbourhood as organising and implementing parental interventions would divert a large number of resources from day to day functioning of the school. As schools are often a focal point of the community even small amounts of involvement, such as advertising sessions or holding classes within the school site, could be beneficial to sustaining parental interventions within the neighbourhood (Sanders, 2000). Alongside this, schools should communicate their behavioural expectations to the parents of children that will join the school population in the future and provide information on how parents could support the development of these behaviours at home. Although it is often

190

difficult for schools serving deprived neighbourhoods to develop positive community ties as they are often already overwhelmed with internal social disorganisation, both the community and the school could benefit in the long term.

7.8.2 Recommendations for future research

As the research process is an additive process, whereby previous research informs the direction of future investigations, the present study provides recommendations for the advancement of the following research areas: poverty, behavioural trajectories, prevention science, and implementation science.

Differential behavioural outcomes for children experiencing poverty

Although there is a substantial evidence base documenting the many negative outcomes children experiencing poverty are at risk of (See McLoyd, 1998 for a review), there is little research determining if there are differential outcomes from universal preventive interventions concerned with improving behaviour within classrooms. While the present study found that there was little evidence for poverty exerting a moderating effect on the GBG, this may not be the case for other preventive interventions. It is therefore in the interest of society to continue such research in order to ascertain which preventive interventions can help towards reducing the ever-increasing disparities between children experiencing poverty and more affluent peers. The analyses required for determining such differential gains would not require collecting additional data as many researchers collect the FSM status as standard demographic data within educational research. With the rise of large RCTs testing the effectiveness of preventive interventions within school-settings, the testing of moderators such as poverty on intervention outcomes should become standard practice within the field of prevention science. While the assessment of differential gains for children experiencing poverty is a requirement for funders such as the EEF, this only extends to academic attainment (EEF, 2011). Though the need to improve the attainment gap is great, so is the need for improving the social and behavioural outcomes of children experiencing poverty. Therefore, the outcomes tested for this at-risk group should expand in order to promote a holistic improvement in the developmental milestones of children experiencing poverty.

Additionally, this call for future research to investigate the moderating effects of poverty on child behavioural outcomes should be extended to other forms of preventive interventions. In particular, parenting programmes. This is because the present study found that an individual's poverty status was an independent factor that increased disruptive behaviour, irrespective of the intervention status of the school. As proposed earlier in the discussion (see section 7.3.1), maladaptive parenting may explain the resistance children experiencing poverty may have had to the GBG, and that earlier intervention could be a more effective strategy for preparing children living in economically disadvantaged neighbourhoods for classroom appropriate behaviour during their educational career (Kellam et al., 1998).

Although there is some research that has indicated that children experiencing poverty could benefit from parenting programmes (e.g. Webster-Stratton, Reid & Stoolmiller, 2008), this area requires further attention in order to establish more confidence in the use of such programmes within high poverty communities.

The moderating effects of school organisational capacity on implementation

While it is slowly becoming common practice to report the effective implementation has on intervention outcomes, there is a need for more research on how the factors associated with implementation influence programme delivery (Domitrovich et al., 2008). Although the present study has contributed to furthering understanding within this particular area of implementation science, the exploratory nature of the analysis meant that any insight was surface level. It was found that school-level poverty was not a moderator for the implementation of the GBG, but organisational capacity may still be responsible for the large proportion of teachers implementing the GBG at a low level of dosage. As there are a range of measures that are available which specifically assess the organisational capacity of schools (e.g. Staff Development and School Climate Assessment Questionnaire: Zigarmi, Edeburn & Blanchard, 1991), future research could utilise these tools in conjunction with school-level poverty data in order to refine current understanding of school-level factors associated with the implementation of interventions.

Wider use of implementation profiles within intervention research

Although the use of multiple implementation dimensions within analyses is not widely reported within the research community at present, further insight can be gained into the delivery of an intervention by examining several implementation variables in conjunction with one another. This would particularly be the case if a person-centred approach to implementation was adopted instead of a variable approach. The influence of specific dimensions could then be identified in order to determine their role in the success or failure of implementing a preventive intervention (Aber et al., 1998). For instance, through the use of HCA, the present study identified three distinct implementation profiles (MF-LD, HF-MD, HF-HD) which appeared to vary in dosage levels. As a large proportion of teachers were clustered within MF-LD ($N=$ 67), the likely cause for the lack of intervention effect across the study was a lack of regular GBG sessions, particularly within the second year of the study. However, further demonstrations of implementation profiles utilising data from other interventions are required to see if meaningful clusters can be created from other data sets, as well as with a broader range of implementation dimensions. Although the application of cluster analysis within implementation science is in its infancy, future research could help the field gain a deeper appreciation of the insights that can be achieved by utilising multiple dimensions of implementation within analyses.

7.9 Contributions to knowledge

The present study has contributed to furthering understanding of both a theoretical and methodological level. The theoretical contributions to knowledge are concerned with the creation of an objective definition for disruptive behaviour, as well as the expansion of current understanding of the role poverty plays in the outcomes and implementation of preventive interventions. Meanwhile, the methodological contributions to knowledge are concerned with the successful generation of meaningful implementation profiles using HCA, as well as the use of MLM to account for the hierarchical nature of data when testing for interaction effects concerning poverty and the GBG.

7.9.1 Theoretical contributions

Firstly, the present study recognised that within the field of developmental psychology, the term disruptive behaviour has often been conflated with other established constructs during investigations concerning the behavioural outcomes of children (see sections 3.2.1 and 3.2.2). In order to distinguish the use of the term within the present study from commonly confused terms, a framework was developed after examining common definitions of maladaptive behavioural outcomes (such as conduct problems). As the literature consistently suggests that children that follow a non-normative behavioural trajectory escalate the intensity and frequency with age (Tremblay, 2010), the terms used to describe the different stages of the developmental trajectory should reflect this increase in externalising behaviours. It was therefore proposed that disruptive behaviours represented a collection of actions that occurred at the lower end of the continuum, while terms such as conduct problems were utilised at the end of the scale (see section 2.2.2). The constructs aggression and impulsivity were also included within the framework as these terms have also been used as synonyms in previous research (eg. Kellam et al, 2011). It was argued that the actions that comprised of this continuum of externalising behaviour were either aggressive or impulsive in nature and that the severity of these actions determined whether these behaviours were disruptive or conduct problems (see section: 3.2.2). The aforementioned distinctions allowed for the following definition for disruptive behaviour to be utilised for the present study: "behaviours that cause interruptions to one's own, or other's learning in the classroom" (from section 3.2.1). This definition for disruptive behaviour in the classroom contributes to the broader literature base by providing both the objectivity and clarity to a term that has often been misused within behavioural research.

The present study's second theoretical contribution is the expansion of current understanding with regards to how poverty influences the behavioural outcomes of individuals participating in a school-based preventive intervention. Although there is a large evidence base documenting the association of childhood poverty with adverse behavioural outcomes (e.g. McLoyd, 1998; Tremblay, 2010), very little research has assessed whether children experiencing poverty benefit from classroom-based behavioural preventive

interventions, particularly individuals who attend high poverty schools. Although the results from the present study indicated that poverty at both the school-level and child-level did not interact with the GBG, it was suggested that children experiencing poverty had developed a level of resistance to the universal "light touch" approach classroom-based interventions utilise. This is because the disruptive behaviours displayed by children experiencing poverty have been internalised through maladaptive parenting practices, often exerted by parents under high levels of stresses caused by a lack of resources (Shaw & Shelby, 2014). These findings have furthered understanding by concluding that earlier intervention, by means of parenting interventions, may be more effective for improving behavioural outcomes for children experiencing poverty.

7.9.2 Methodological contributions

The generation of meaningful implementation profiles is the present study's primary methodological contribution to knowledge. This was achieved through the use of HCA and provided insight into the lack of intervention effects reported for RQ1. This approach was in keeping with the micro-view of implementation, a perspective that is slowly becoming dominant within prevention science (Durlak & DuPre, 2008). However, studies measuring implementation often only utilise one dimension within their analyses, and this increases the risk of committing a type III error (Humphrey et al., 2016). Further insight can be gained by taking a person-centred approach to analysing implementation data, which reduces the risk of misattributing the cause of failure. For instance, the present study utilised two variables (fidelity-quality and dosage) to generate implementation profiles. From this, it was ascertained that although teachers were implementing the intervention to an appropriate standard of fidelity, dosage levels were likely to be responsible for the null intervention effect. If the present study had only focused on fidelity, as many studies often do, a type III error would have been made. To date, the use of HCA has not been tested within prevention science as a method for treating implementation data, and so the testing of this analysis within the present study provides a promising alternative for future directions within the field of intervention implementation.

The second methodological contribution provided by the present study improved upon the current literature base concerning the GBG. This was achieved by utilising MLM as the primary method of analysis in order to account for the hierarchical nature of the data set. This is because the only existing study investigating the moderating effects of poverty on the intervention's outcomes was unable to use such statistical techniques to account for the shared variance that occurs within nested datasets, as these capabilities were unavailable at the time (Kellam et al., 1994). Using MLM over single-level regression is considered best practice as a more accurate understanding of the influence of the wider surrounding (e.g. the class or school) when calculating effect sizes (Twisk, 2006). This reduces the likelihood of a type I error occurring because the variance associated with the nesting effect is accounted for within MLM (Goldstein, 2011). Although the findings of the current study

emulate those found by Kellam and colleagues (1998), the present study still provides a valuable contribution in terms of furthering understanding of the independent nature of poverty and GBG the in relation to behavioural outcomes.

7.10 Summary of the present study

The present study set out to investigate the moderating effects of school-level and child-level poverty on both the behavioural outcomes and implementation of a classroom-based preventive intervention called the Good Behaviour Game (GBG). Although the intervention has a large evidence base, very little research specifically focused on the potential differential behavioural outcomes children experiencing poverty may have had as a result of the programme. Further to this, the literature concerning the implementation of the GBG is predominantly collected and analysed using a variable-approach and has often been utilised for descriptive purposes only (e.g. Dion et al., 2011). The present study endeavoured to add to the existing knowledge-base in three ways:

- Accounting for the hierarchical nature of data when investigating the moderating effects of poverty on the GBG where previous research did not account for it (e.g. Kellam et al., 1994; Kellam et al., 1998).

- Taking a person-centred approach to implementation by attempting to create implementation profiles in order to incorporate multiple dimensions of implementation.

- Testing whether poverty moderated the implementation of the GBG in order to further understanding the proposed factors that affect implementation.

The above aims were achieved through the adoption of an experimental design, in the form of a RCT. The RCT comprised of two conditions, where schools were randomised to either the intervention arm (N= 38) or the control arm of the trial (N= 39). Overall, 3084 pupils and 118 teachers participated in the study over a two-year period. Behavioural outcomes were measured at three time points: The baseline measure was taken at the end of the academic year when the cohort was in Year 2, with follow up measures collected at the end of subsequent academic years. Behavioural data consisted of the disruptive behaviour subscale of the TOCA-C (Koth, Bradshaw & Leaf, 2009) and required the teacher to rate each pupil's classroom behaviour on a Likert scale. The FSM status of pupils served as a proxy of poverty for the individual-level, the percentage of FSM eligible children within the pupil population functioned as the proxy for school-level poverty. Schools were later categorised as moderate or high poverty based on the cut-off point for the purposes of comparison. Both school and child-level poverty data were collected through an NPD request. Meanwhile, the majority of implementation data was collected by means of an annual on-site visit to schools delivering the GBG. These visits consisted of a structured observation where a researcher utilised a bespoke checklist in order to rate the fidelity and

quality of delivery of a GBG session. The remaining implementation dimension, dosage, was collected through online scoreboards that teachers used when delivering the GBG. This data was collected throughout the trial.

The primary means of data analysis was MLM. This method was chosen as the dataset had a natural hierarchical structure, whereby pupils were nested in classes, classes were nested in schools, and schools were nested within either the control or experimental condition. MLM also accounted for the interactions between different variables across levels of data, which allowed for the testing of potential effects of school-level poverty on child-level behavioural outcomes. Meanwhile, HCA was utilised to create the implementation profiles required to group teachers in order to compare different levels of implementation and their effects of disruptive behaviour. Although this was achieved, it was originally intended that three dimensions (fidelity, quality of delivery, and dosage) would contribute to the cluster analysis. However, during the preliminary stages of analysis, it was found that fidelity and quality of delivery were too correlated to be considered separate variables (See section 5.6.3). This lead to the two variables to be transformed into one variable, which was named fid-qual. The new variable and dosage data were then used to generate the implementation profiles.

The present study found that child-level poverty and school-level poverty did not moderate the intervention effect. Instead, poverty's effect on disruptive levels appeared to be independent, with disruption levels increasing over the two-years for children experiencing poverty, or children attending high poverty schools compared to counterparts in the usual practice condition. However, the three-way interaction between the GBG, school poverty status and FSM eligibility was statistically significant (β_{0ij}= .33, p= .006). But it indicated that children experiencing poverty in high poverty schools that delivered the intervention became more disruptive within the first year of the trial. This finding was not replicated in the second year of the study. Overall, these results indicated that the GBG did not positively impact the behavioural outcomes of children experiencing poverty, or schools serving economically disadvantaged neighbourhoods.

Further insight was gained from utilising the implementation data in exploring why the GBG was not successful in improving disruptive behaviour within classrooms. Three implementation profiles were created as a result of HCA. These were labelled Medium Fidelity-Low Dosage (MF-LD: N= 67), High Fidelity-Medium Dosage (HF-MD: N= 19), and High Fidelity-High Dosage (HF-HD: N= 7). It appeared that implementation decreased within the second year, as the proportion of teachers within the MF-LD profile increased, and the HF-MD profile ceased to exist. While these were interesting observations which may explain the lack of intervention effect in the second year of the study, it meant that only data from the first year of the study could be utilised to conduct exploratory analyses on poverty's moderating effects on implementation. The MLMs utilised for this portion of the

present study were classified as exploratory because the dataset was underpowered as an additional data level was introduced within a smaller pool of participants[21]. The findings suggested that poverty did not interact with implementation to moderate the behavioural outcomes of children, particularly individuals experiencing poverty.

In light of the above results, the present study concluded that poverty remained an independent factor that contributed to the increase in disruptive behaviour. As the GBG did not positively impact children experiencing poverty, it is likely that classroom-based interventions are not an effective counter to the maladaptive behaviour trajectories often observed for this particular risk group. It was suggested that earlier intervention could be more effective at improving the behavioural outcomes for children experiencing poverty, as disruptive behaviour observed in the classroom may be due to maladaptive parenting practices occurring within economically disadvantaged households. Therefore, future research and policy should consider further exploring parenting interventions to improve the behavioural outcomes of children experiencing poverty. A further conclusion drawn from the present study concerns the role of school-level poverty as a factor affecting implementation. Although the study was unable to establish an association between school-level poverty and implementation, this area of the study had an exploratory purpose. The large proportion of teachers that implemented the GBG at low levels of dosage indicate that organisational capacity may be a contributing factor to the lack of intervention effects reported throughout the study. As teachers working within high poverty schools may be overwhelmed with "daily fire-fighting" and receive a little support from colleagues and senior staff to implement preventive interventions sustainably. Future research should utilise specific measures for organisational capacity to ascertain whether this construct would be more appropriate when considering a proxy.

[21] This was because schools in the control condition had to be excluded from the analysis as RQ2 was specifically interested in the implementation of the GBG.

Reference List

Aber, J. L., Bennett, N. G., Conley, D. C., & Li, J. (1997). The effects of poverty on child health and development. *Annual Review of Public Health*, *18*, 463–483. https://doi.org/10.1146/annurev.publhealth.18.1.463

Abry, T., Rimm-Kaufman, S. E., Larsen, R. A., & Brewer, A. J. (2013). The influence of fidelity of implementation on teacher-student interaction quality in the context of a randomized controlled trial of the Responsive Classroom Approach. *Journal of School Psychology*, *51*, 437–453. https://doi.org/10.1016/j.jsp.2013.03.001

Achenbach, T. M., & Edelbrock, C. S. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development*, *46*, 1–82. https://doi.org/10.2307/1165983

Adelman, H. S., & Taylor, L. (2003). On sustainability of project innovations as systematic change. *Journal of Educational and Psychological Consultation*, *14*, 1–25. https://doi.org/10.1207/S1532768XJEPC1401

Aeschleman, S. R., Rosen, C. C., & Williams, M. R. (2003). The effect of non-contingent negative and positive reinforcement operations on the acquisition of superstitious behaviors. *Behavioural Processes*, *61*, 37–45. https://doi.org/10.1016/S0376-6357(02)00158-4

Airely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions*. New York: Harper Collins.

Akiba, M. (2010). What predicts fear of school violence among U.S. adolescents? *Teachers College Record*, *112*, 68–102. Retrieved from https://psycnet.apa.org/record/2011-06379-003

Alexander, R. (2001). *Culture and pedagogy: International comparisons in primary education.* Malden: Blackwell.

Aliber, R. Z., & Kindleberger, C. P. (2015). *manias, panics, and crashes*. London: Palgrave Macmillan UK. https://doi.org/10.1007/978-1-137-52574-1

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological Methods & Research*, *28*, 301–309. https://doi.org/10.1177/0049124100028003003

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ*, *332*, 1080–1801. https://doi.org/10.1136/bmj.332.7549.1080

Anderson, G., & Horvath, J. (2004). The growing burden of chronic disease in america. *Public Health Reports*, *119*, 263–270. https://doi.org/10.1016/j.phr.2004.04.005

Anderson, J., Hollinger, D., & Conaty, J. (1992). Poverty and achievement: re-examining the relationship between school poverty and student achievement: an examination of eighth grade student achievement using the national education longitudinal study of 1988. In *American Educational Research Association*. Washington, DC: U.S. Department for Education.

Anderson, R. M., & May, R. M. (1985). Vaccination and herd immunity to infectious diseases. *Nature*, *318*, 323–329. https://doi.org/10.1038/318323a0

Apouey, B., & Geoffard, P. Y. (2013). Family income and child health in the UK. *Journal of Health Economics*, *32*, 715–727. https://doi.org/10.1016/j.jhealeco.2013.03.006

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology*, *8*, 291–322. https://doi.org/10.1037/1089-2680.8.4.291

Atkinson, E. (2000). In defence of ideas, or why "what works" is not enough. *Management in Education*, *14*, 6–9. https://doi.org/10.1177/089202060001400302

Attar, B. K., Guerra, N. G., Tolan, P. H., Children, U. E., & Attar, B. M. (2010). Neighborhood disadvantage, stressful life events and adjustments in urban elementary-school children. *Journal of Clinical Child Psychology*, 12, 37–41. https://doi.org/10.1207/s15374424jccp2304

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*, 7–39. https://doi.org/10.1080/19345747.2016.1232459

Baker, C. N., Kupersmidt, J. B., Voegler-Lee, M. E., Arnold, D. H., & Willoughby, M. T. (2010). Predicting teacher participation in a classroom-based, integrated preventive intervention for pre-schoolers. *Early Childhood Research Quarterly*, *25*, 270–283. https://doi.org/10.1016/j.ecresq.2009.09.005

Balas, E. A., & Boren, S. A. (2000). Managing clinical knowledge for health care improvement. *Yearbook of Medical Informatics*, *9*, 65–70. https://doi.org/10.1055/s-0038-1637943

Baldacchino, S. (2017). Mental health promotion in schools: Cross-cultural narratives and perspectives. In C. Cefai & P. Cooper (Eds.), *Mental Health Promotion in Schools: Cross-Cultural Narratives and Perspectives* (pp. 121–140). Rotterdam: Sense Publishers. https://doi.org/10.1007/978-94-6351-053-0

Ball, S. (2007). "You've been NERFed!" dumbing down the academy: National Educational Research Forum: "a national strategy? consultation paper": a brief and bilious response. *Journal of Education Policy*, *16*, 265–268. https://doi.org/10.1080/713807558

Bandura, A. (1978). Social learning theory of aggression. *Journal of Communication*, *28*, 12–29. https://doi.org/10.1111/j.1460-2466.1978.tb01621.x

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *The Journal of Abnormal and Social Psychology*, *63*, 575–582. https://doi.org/10.1037/h0045925

Barnes, J., Belsky, J., Broomfield, K. A., & Melhuish, E. (2006). Neighbourhood deprivation, school disorder and academic achievement in primary schools in deprived communities in England. *International Journal of Behavioral Development*, *30*, 127–136. https://doi.org/10.1177/0165025406063585

Baron, R. A., & Richardson, D. R. (2004). *Human aggression*. New York: Springer. Retrieved from https://books.google.co.uk/books?id=MOljg8xXI1oC

Baron, S., Field, J., & Schuller, T. (2000). *Social capital: critical perspectives.* Oxford: OUP Oxford. Retrieved from https://books.google.co.uk/books?id=nvivgiFfPr0C

Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, *2*, 119–124. https://doi.org/10.1901/jaba.1969.2-119

Bartle, J., Birch, S., & Skirmuntt, M. (2017). The local roots of the participation gap: Inequality and voter turnout. *Electoral Studies*, *48*, 30–44. https://doi.org/10.1016/j.electstud.2017.05.004

Beaman, R., Wheldall, K., & Kemp, C. (2007). Recent research on troublesome classroom behavior: a review. *Australasian Journal of Special Education*, *31*, 45–60. https://doi.org/10.1080/10300110701189014

Becker, K. D., Bradshaw, C. P., Domitrovich, C., & Ialongo, N. S. (2013). Coaching teachers to improve implementation of the good behavior game. *Administration and Policy in Mental Health and Mental Health Services Research*, *40*, 482–493. https://doi.org/10.1007/s10488-013-0482-8

Beets, M. W., Flay, B. R., Vuchinich, S., Acock, A. C., Li, K. K., & Allred, C. (2008). School climate and teachers' beliefs and attitudes associated with implementation of the positive action program: A diffusion of innovations model. *Prevention Science*, *9*, 264–275. https://doi.org/10.1007/s11121-008-0100-2

Beresford, P. (2002). Thinking about "mental health": Towards a social model. *Journal of Mental Health*, *11*, 581–584. https://doi.org/10.1080/09638230020023921

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: an integrated model of program implementation. *Prevention Science*, *12*, 23–33. https://doi.org/10.1007/s11121-010-0186-1

Bernal, G., & Sáez-Santiago, E. (2006). Culturally-centered psychosocial interventions. *Journal of Community Psychology*, *34*, 121–132. https://doi.org/10.1002/jcop.20096

Bernard, K., Zwerling, J., & Dozier, M. (2015). Effects of early adversity on young children's diurnal cortisol rhythms and externalizing behavior. *Developmental Psychobiology*, *57*, 935–947. https://doi.org/10.1002/dev.21324

Bernstein, E., & Harvey, E. C. (1911). *Evolutionary socialism: A criticism and affirmation.* New York: Huebsch. Retrieved from https://books.google.co.uk/books?id=sK2ZAAAAIAAJ

Berthoud, R., Bryan, M., & Bardarsi, E. (2004). *The dynamics of deprivation: The relationship between income and material deprivation over Time* (No. 219). London.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Bevans, K., Bradshaw, C., Miech, R., & Leaf, P. (2007). Staff and school-level predictors of school organizational health: A multilevel analysis. *Journal of School Health*, *77*, 294–302. https://doi.org/10.1111/j.1746-1561.2007.00210.x

Beveridge, W. (1942). *Social Insurance and Allied Services*. London. Retrieved from http://www.nationalarchives.gov.uk/pathways/citizenship/brave_new_world/welfare.htm

Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., McMahon, R. J., & Pinderhughes, E. (2010). The effects of a multiyear universal social–emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology*, *78*, 156–168. https://doi.org/10.1037/a0018607

Biesta, G. (2007). Why '"what works"' won't work : Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, *57*, 1–22. https://doi.org/10.1111/j.1741-5446.2006.00241.x

Bishop, D. C., Pankratz, M. M., Hansen, W. B., Albritton, J., Albritton, L., & Strack, J. (2014). measuring fidelity and adaptation: reliability of a instrument for school-based prevention programs. *Evaluation and the Health Professions*, *37*, 231–257. https://doi.org/10.1177/0163278713476882

Björkqvist, K. (2018). Gender differences in aggression. *Current Opinion in Psychology*, *19*, 39–42. https://doi.org/10.1016/j.copsyc.2017.03.030

Björkqvist, K., Lagerspetz, K. M. J., & Kaukiainen, A. (1992). Do girls manipulate and boys fight? Developmental trends in regard to direct and indirect aggression. *Aggressive Behavior*, *18*, 117–127. https://doi.org/10.1002/1098-2337(1992)18:2<117::AID-AB2480180205>3.0.CO;2-3

Blakely, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., & Emshoff, J. G. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology*, *15*(3), 253–268. https://doi.org/10.1007/BF00922697

Blau, D. M. (1999). The effect of income on child development. *Review of Economics and Statistics*, *81*, 261–276. https://doi.org/10.1162/003465399558067

Bøe, T., Sivertsen, B., Heiervang, E., Goodman, R., Lundervold, A. J., & Hysing, M. (2014). Socioeconomic status and child mental health: The role of parental emotional well-being and parenting practices. *Journal of Abnormal Child Psychology*, *42*, 705–715. https://doi.org/10.1007/s10802-013-9818-9

Bonell, C. P. P., Fletcher, A., Jamal, F., Wells, H., Harden, A., Murphy, S., & Thomas, J. (2013). Theories of how the school environment impacts on student health: Systematic review and synthesis. *Health & Place*, *24*, 242–249. https://doi.org/10.1016/j.healthplace.2013.09.014

Book, A., Starzyk, K., & Quinsey, V. (2001). The relationship between testosterone and aggression: A meta-analysis. *Aggression & Violent Behaviour*, *6*, 579–599. https://doi.org/10.1016/S1359-1789(00)00032-X

Borglin, G. (2015). The value of mixed methods for researching complex interventions. In D. Richards & I. Hallberg (Eds.), *Complex interventions in health: An overview of research methods* (pp. 29–45). New york, NY: Routledge.

Botvin, G. J. (2004). Advancing prevention science and practice: Challenges, critical issues, and future directions. *Prevention Science*, *5*, 69–72. https://doi.org/10.1023/B:PREV.0000013984.83251.8b

Bourdieu, P. (1972). Les stratégies matrimoniales dans le système de reproduction. *Annales. Histoire, Sciences Sociales*, *27*, 1105–1127. Retrieved from https://www.jstor.org/stable/27578159

Bowman-Perrott, L., Burke, M. D., Zaini, S., Zhang, N., & Vannest, K. (2016). Promoting positive behavior using the good behavior game: a meta-analysis of single-case research. *Journal of Positive Behavior Interventions*, *18*, 180–190. https://doi.org/10.1177/1098300715592355

Boyden, J., & Cooper, E. (2007). *Questioning the Power of Resilience : Are Children Up To the Task of Disrupting the Transmission of Poverty ?* (No. 73). Oxford: The University of Oxford.

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics*, *130*, 1136–1145. https://doi.org/10.1542/peds.2012-0243

Bradshaw, C. P., Sawyer, A. L., O'Brennan, L. M., & O'Brennan, L. M. (2009). A social disorganization perspective on bullying-related attitudes and behaviors: The influence of school context. *American Journal of Community Psychology*, *43*, 204–220. https://doi.org/10.1007/s10464-009-9240-1

Brame, B., Nagin, D. S., & Tremblay, R. E. (2001). Developmental trajectories of physical aggression from school entry to late adolescence. *Journal of Child Psychology and Psychiatry*, *42*, 503–512. https://doi.org/10.1017/S0021963001007120

Brehm, J., & Rahn, W. (1997). Individual-level evidence for the causes and consequences of social capital. *American Journal of Political Science*, *41*, 999–1023. https://doi.org/10.1021/jo401992d

Bridges, D. (1998). Research for sale: Moral market or moral maze? *British Educational Research Journal*, *24*, 593–607. https://doi.org/10.1080/0141192980240507

British Educational Research Association. (2011). Ethical guidelines for educational research. In British Educational Research Association (Ed.). London. Retrieved from https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf?noredirect=1%0Ahttps://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2011

Bronfenbrenner, U. (1994). Ecological models of human development. In M. Gauvain & M. Cole (Eds.), *Readings on the development of children* (2nd ed., pp. 37–43). New York: Freeman. https://doi.org/10.4135/9781446215159.n280

Brooks-gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children*, *7*, 55–71. https://doi.org/10.2307/1602387

Brotman, L. M., Kingston, S., Bat-Chava, Y., Caldwell, M. B., & Calzada, E. J. (2008). Training school personnel to facilitate a family intervention to prevent conduct problems. *Early Education and Development*, *19*, 622–642. https://doi.org/10.1080/15374410802231057

Burchardt, T., Le Grand, J., & Piachaud, D. (1999). Social Exclusion in Britain 1991-1995. *Social Policy & Administration*, *33*, 227–244. https://doi.org/10.1111/1467-9515.00148

Burrell, G., & Morgan, G. (1979). Two dimensions: Four paradigms. In D. Burrell & G. Morgan (Eds.), *Sociological paradigms and organizational analysis* (pp. 21–87). Ashgate: Heinemann.

Butts, C. T. (2009). Revisiting the foundations of network analysis. *Science*, *325*, 414–416. https://doi.org/10.1126/science.1171022

Cabinet Office. (2019). Guidance: What Works Network. Retrieved June 17, 2018, from https://www.gov.uk/guidance/what-works-network

Callan, T., Nolan, B., & Whelan, C. T. (1993). Resources, deprivation and the measurement of poverty. *Journal of Social Policy*, *22*, 141-172. https://doi.org/10.1017/s0047279400019280

Cancian, M., Yang, M.-Y., & Slack, K. S. (2013). The effect of additional child support income on the risk of child maltreatment. *Social Service Review*, *87*, 417–437. https://doi.org/10.1086/671929

Cantillion, B., Chzhen, Y., Handa, S., & Nolan, B. (2017). *Children of austerity: Impact of the great recession on child poverty in rich countries*. Oxford: Oxford University Press.

Card,Noel A., Stucky,Brian D., Sawalani,Gita M., & Little,Todd D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development*, *79*, 1185–1229. https://doi.org/10.1111/j.1467-8624.2008.01184.x

Carpenter, J. J. R., Goldstein, H., & Kenward, M. G. M. (2011). REALCOM-IMPUTE software for multilevel multiple imputation with mixed response types. *Journal of Statistical Software*, *45*, 1–12. https://doi.org/http://dx.doi.org/10.18637/jss.v045.i05

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*, 1–9. https://doi.org/10.1186/1748-5908-2-40

Casazza, A. (2015). How are all countries rich and poor to define poverty? Retrieved January 1, 2017, from http://www.africa.undp.org/content/rba/en/home/ourperspective/ourperspectivearticles/2015/10/18/how-are-all-countries-rich-and-poor-to-define-poverty-.html

Case, K. E., Fair, R. C., Oster, S. E., & Oster, S. M. (2017). *Principles of Economics* (12th ed.). London: Pearson. Retrieved from https://books.google.co.uk/books?id=ijkDswEACAAJ

Cash, R., & Nealis, L. (2004). Mental health in schools: It's a matter of public policy. In *Public Policy Institute of the National Association of School Psychologists*. Washington, DC.

Castro, F. G., Barrera Jr., M., & Martinez Jr., C. R. (2004). The cultural adaptation of prevention interventions: resolving tensions between fit and fidelity. *Prevention Science*, *5*, 41–45. https://doi.org/10.1037/e628592012-269

Castro, F. G., & Yasui, M. (2017). Advances in EBI development for diverse populations: Towards a science of intervention adaptation. *Prevention Science*, *18*, 623–629. https://doi.org/10.1007/s11121-017-0809-x

Catalano, R. F., Fagan, A. A., Gavin, L. E., Greenberg, M. T., Irwin, C. E., Ross, D. A., & Shek, D. T. L. (2012). Worldwide application of prevention science in adolescent health. *The Lancet*, *379*, 1653–1664. https://doi.org/10.1016/s0140-6736(12)60238-4

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing when the "have-nots" gain but the "haves" gain even more. *American Psychologist*, *60*, 149–160. https://doi.org/10.1037/0003-066X.60.2.149

Centola, D., & Macy, M. (2005). Complex contagions and the weakness of long ties. *American Journal of Sociology*, *113*, 702–734. https://doi.org/10.1086/521848

Chan, G. C., Foxcroft, D. R., Smurthwaite, B., Coomes, L., & Allen, D. (2012). *Improving child behaviour management: An evaluation of the good behaviour game in uk primary schools*. Oxford: Oxfordshire County Council.

Chang, H. J. (2014). *Economics: The user's guide*. London: Pelican. Retrieved from https://books.google.co.uk/books?id=lcUjngEACAAJ

Chew, B. (1988). The no-nonsense guide to measuring productivity. *Harvard Business Review*, *66*, 110–118. Retrieved from https://hbr.org/1988/01/no-nonsense-guide-to-measuring-productivity

Chinman, M., Hannah, G., Wandersman, A., Ebener, P., Hunter, S. B., Imm, P., & Sheldon, J. (2005). Developing a community science research agenda for building community capacity for effective preventive interventions. *American Journal of Community Psychology*, *35*, 143–157. https://doi.org/10.1007/s10464-005-3390-6

Cho, H., Hallfors, D. D., & Sánchez, V. (2005). Evaluation of a high school peer group intervention for at-risk youth. *Journal of Abnormal Child Psychology*, *33*, 363–374. https://doi.org/10.1007/s10802-005-3574-4

Chung, W., Ha, H., & Kim, B. (2016). Money transfer and birth weight: Evidence from the Alaska permanent fund dividend. *Economic Inquiry*, *54*, 576–590. https://doi.org/10.1111/ecin.12235

Clotfelter, C., Ladd, H. F., Vigdor, J., & Wheeler, J. (2007). High-poverty schools and the distribution of teachers and principals. *North Carolina Law Review*, *85*, 1345–1380. https://doi.org/10.1037/e722752011-001

Coburn, C. E. (2004). Beyond de-coupling: rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, *77*, 211–244. https://doi.org/10.1177/003804070407700302

Coe, R. (2002). It's the effect size, stupid. In *British Annual Research Association Conference*. Exeter.

Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J., … Long, B. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist*, *48*, 1013–1022. https://doi.org/10.1037/0003-066X.48.10.1013

Coleman, J. (2000). Social capital in the creation of human capital. In *Knowledge and Social Capital*, 22, 17–41. https://doi.org/10.1016/B978-0-7506-7222-1.50005-2

Conger, R. D., Conger, K. J., & Martin, M. J. (2010). Socioeconomic status, family processes,and individual development. *Journal of Marriage and Family*, *72*, 685–704. https://doi.org/10.1111/j.1741-3737.2010.00725.x

Connell, J. P., & Klem, A. M. (1998). You can get there from here: Using a theory of change approach to plan urban education reform. *Journal of Educational and Psychological Consultation*, *11*, 93–120. https://doi.org/10.1207/S1532768XJEPC1101_6

Coombes, L., Chan, G., Allen, D., & Foxcroft, D. (2016). Mixed-methods evaluation of the good behaviour game in english primary schools. *Journal of Community & Applied Social Psychology*, *25*, 432–446. https://doi.org/10.1002/casp

Cooper, K., & Stewart, K. (2017). *Does money affect children's outcomes? An Update*. London. Retrieved from http://eprints.lse.ac.uk/54435

Cooper, P. (2011). Teacher strategies for effective intervention with students presenting social, emotional and behavioural difficulties: an international review. *European Journal of Special Needs Education*, *26*, 71–86. https://doi.org/10.1080/08856257.2011.543547

Cooper, P., Smith, C., & Upton, G. (1994). *Emotional and behavioural difficulties*. London: Routledge. https://doi.org/10.4324/9780203416297

Cormack, R. M. (1971). A review of classification. *Journal of the Royal Statistical Society. Series A (General)*, *134*, 321–367. https://doi.org/10.2307/2344237

Côté, S. M., Vallancourt, T., Barker, E. D., Nagin, D., & Tremblay, R. E. (2007). The joint development of physical and indirect aggression: Predictors of continuity and change during childhood. *Development and Psychopathology*, *19*, 37–55. https://doi.org/10.1017/S0954579407070034

Crane, D. (1999). Diffusion models and fashion: A reassessment. *The Annals of the American Academy of Political and Social Science*, *566*, 13–24. https://doi.org/10.1177/000271629956600102

Creswell, J. W., & Poth, C. N. (2017). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (2nd ed.). New York: SAGE Publications. Retrieved from https://books.google.co.uk/books?id=DLbBDQAAQBAJ

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. New York: SAGE Publications, Inc.

Cunha, F., & Heckman, J. (2007). *The technology of skill formation* (Vol. 97). Cambridge, MA. https://doi.org/10.3386/w12840

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology* (4th ed.). Harlow: Pearson. Retrieved from https://books.google.co.uk/books?id=QjfQ0_DqyNQC

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

Dariotis, J. K., Bumbarger, B. K., Duncan, L. G., & Greenberg, M. T. (2008). How do implementation efforts relate to program adherence? Examining the role of organizational, implementer, and program factors. *Journal of Community Psychology*, *36*, 744–760. https://doi.org/10.1002/jcop.20255

Davis, E. P., & Sanchez-martinez, M. (2014). *A review of the economic theories of poverty. National Institute of Economic and Social Research*.

Dawson-Mcclure, S., Brotman, L. M., Theise, R., Palamar, J. J., Kamboukos, D., Barajas, R. G., & Calzada, E. J. (2014). Early childhood obesity prevention in low-income, urban communities. *Journal of Prevention and Intervention in the Community*, *42*, 152–166. https://doi.org/10.1080/10852352.2014.881194

Dearing, E., Kreider, H., Simpkins, S., & Weiss, H. B. (2006). Family involvement in school and low-income children's literacy: Longitudinal associations between and within families. *Journal of Educational Psychology*, *98*, 653–664. https://doi.org/10.1037/0022-0663.98.4.653

Deković, M., Slagt, M. I., Asscher, J. J., Boendermaker, L., Eichelsheim, V. I., & Prinzie, P. (2011). Effects of early prevention programs on adult criminal offending: A meta-analysis. *Clinical Psychology Review*, *31*, 532–544. https://doi.org/10.1016/j.cpr.2010.12.003

Dempsey, N., Bramley, G., Power, S., & Brown, C. (2011). The social dimension of sustainable development: Defining urban social sustainability. *Sustainable Development*, *19*, 289–300. https://doi.org/10.1002/sd.417

Department for Communities and Local Government. (2011). *The English indices of deprivation 2010*. London: Department for Communities and Local Government.

Department for Education. (2015). *Statistical working paper measuring disadvantaged pupils ' attainment gaps over time*. London: Department for Education.

Department for Education. (2016a). *Pupil absence in schools in England: 2014 to 2015*. London: Department for Education.

Department for Education. (2016b). *Schools workforce in England 2010 to 2015: Trends and geographical*. London: Department for Education.

Department for Education. (2016c). *Schools, pupils, and their characteristics: January 2016*. London: Department for Education.

Department for Education. (2017a). *National curriculum assessments at key stage 4 in England 2017*. London: Department for Education.

Department for Education. (2017b). *National curriculum assessments at key stage 2 in England 2017*. London.

Department for Education. (2018). *Pupil abscence in schools in england: 2016 to 2017*. London: Department for Education.

Department for Work and Pensions. (2017). *Households below average income: 1994/95 to 2015/16*. London. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/600091/households-below-average-income-1994-1995-2015-2016.pdf

Deutsch, C. (1973). Social class and child development. In B. Caldwell & H. Ricciuti (Eds.), *Review of Child Development Research Vol. 3* (pp. 233–282). Chicago: University of Chicago Press.

Dewey, J. (1976). *Essays on logical theory. 1902–1903.* Chicago: Southern Illiinois University Press.

Dion, E., Roux, C., Landry, D., Fuchs, D., Wehby, J., & Dupéré, V. (2011). Improving attention and preventing reading difficulties among low-income first-graders: A

randomized study. *Prevention Science*, *12*(1), 70–79. https://doi.org/10.1007/s11121-010-0182-5

DiPietro, J. A. (1981). Rough and tumble play: A function of gender. *Developmental Psychology*, *17*, 50–58. https://doi.org/10.1037/0012-1649.17.1.50

Dodds, P. S., & Watts, D. J. (2004). Universal behavior in a generalized model of contagion. *Physical Review Letters*, *92*, 1–4. https://doi.org/10.1103/PhysRevLett.92.218701

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

Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., & DeRousie, R. M. S. (2010). Implementation quality: Lessons learned in the context of the Head Start REDI trial. *Early Childhood Research Quarterly*, *25*, 284–298. https://doi.org/10.1016/j.ecresq.2010.04.001

Domitrovich, C. E., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation*, *11*(2), 193–221. https://doi.org/10.1207/S1532768XJEPC1102_04

Domitrovich, C. E., Pas, E. T., Bradshaw, C. P., Becker, K. D., Keperling, J. P., Embry, D. D., & Ialongo, N. (2015). Individual and school organizational factors that influence implementation of the PAX Good Behavior Game intervention. *Prevention Science*, *16*, 1064–1074. https://doi.org/10.1007/s11121-015-0557-8

Dowdy, E., Ritchey, K., & Kamphaus, R. W. (2010). School-based screening: A population-based approach to inform and monitor children's mental health needs. *School Mental Health*, *2*, 166–176. https://doi.org/10.1007/s12310-010-9036-3

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta analytic review. *American Journal of Community Psychology*, *30*, 157–198. https://doi.org/10.1023/A:1014628810714

Duncan, G. J., Brooks-gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development*, *65*, 296–318.

Duncan, G. J., Magnuson, K., Kalil, A., & Ziol-Guest, K. (2012). The importance of early childhood poverty. *Social Indicatots Research*, *108*, 87–96. https://doi.org/10.1007/sll205-011-9867-9

Duncan, G. J., Morris, P. a, & Rodrigues, C. (2011). Does money really matter? *Development Psychology*, *47*, 1263–1279. https://doi.org/10.1037/a0023875.Does

Duncan, G. J., Yeung, W. J., Brooks-Gunn, J., & Smith, J. R. (1998). How much does childhood poverty affect the life chances of children? *American Sociological Review*, *63*, 406. https://doi.org/10.2307/2657556

Durlak, J. A., & Wells, A. M. (1998). Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. *Am J Community Psychol*, *26*, 775–802. https://doi.org/10.1023/A:1022162015815

Durlak, J., & DuPre, E. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327–350. https://doi.org/10.1007/s10464-008-9165-0

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*, 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Durrant, G. B. (2009). Imputation methods for handling item-nonresponse in practice: methodological issues and recent debates. *International Journal of Social Research Methodology*, *12*, 293–304. https://doi.org/10.1080/13645570802394003

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

Education Endownment Foundation. (2017). *Annual Report 2017*. London. Retrieved from https://educationendowmentfoundation.org.uk/about/annual-reports/

Eisenberg, D., & Neighbors, K. (2007). *Economics of Preventing Mental Disorders and Substance Abuse Among Young People.* Washington, DC.

Eisenberg, N., Spinrad, T. L., & Eggum, N. D. (2010). Emotion-Related Self-Regulation and Its Relation to Children's Maladjustment. *Annual Review of Clinical Psychology*, *6*, 495–525. https://doi.org/10.1146/annurev.clinpsy.121208.131208

Eisner, M. (2009). No effects in independent prevention trials: Can we reject the cynical view? *Journal of Experimental Criminology*, *5*, 163–183. https://doi.org/10.1007/s11292-009-9071-y

Elder, G. (1974). *Children of the Great Depression*. Chicago: University of Chicago Press.

Elias, M. J., Zins, J. E., Graczyk, P. A., & Weissberg, R. P. (2003). Implementation, sustainability, and scaling up of social-emotional and academic innovations in public schools. *School Psychology Review*, *32*, 303–319. Retrieved from https://www.scopus.com/record/display.uri?eid=2-s2.0-0242511164&origin=inward&txGid=e2a0c15048e7b71fee079d35d7839c9e

Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs: Blending prevention research and practice in schools. *Prevention Science*, *5*, 47–53. https://doi.org/10.1023/B:PREV.0000013981.28071.52

Embry, D. D. (2002). The Good Behavior Game: A best practice candidate as a universal behavioral vaccine. *Clinical Child and Family Psychology Review*, *5*, 273–297. https://doi.org/10.1023/A:1020977107086

Evans, G. W., & Schamberg, M. A. (2009). Childhood poverty, chronic stress, and adult working memory. *Proceedings of the National Academy of Sciences*, *106*, 6545–6549. https://doi.org/10.1073/pnas.0811910106

Evans, G. W. (2006). Child development and the physical environment. *Annual Review of Psychology*, *57*, 423–451. https://doi.org/10.1146/annurev.psych.57.102904.190057

Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist*, *59*(2), 77–92. https://doi.org/10.1037/0003-066X.59.2.77

Evans, G. W., Boxhill, L., & Pinkava, M. (2008). Poverty and maternal responsiveness: The role of maternal stress and social resources. *International Journal of Behavioral Development*, *32*, 232–237. https://doi.org/10.1177/0165025408089272

Evans, G. W., Chen, E., Miller, G., & Seeman, T. (2012). *how poverty gets under the skin: A life course perspective*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199769100.013.0001

Evans, G. W., & Kim, P. (2012). Childhood poverty and young adults' allostatic load: The mediating role of childhood cumulative risk exposure. *Psychological Science*, *23*, 979–983. https://doi.org/10.1177/0956797612441218

Evans, M. E. (2012). Barriers and facilitators to implementation of the institute of medicine recommendations on preventing mental, emotional, and behavioral disorders among young people. *Journal of Child and Adolescent Psychiatric Nursing*, *25*, 60–65. https://doi.org/10.1111/j.1744-6171.2012.00323.x

Evans, R., Murphy, S., & Scourfield, J. (2015). Implementation of a school-based social and emotional learning intervention: understanding diffusion processes within complex systems. *Prevention Science*, *16*, 754–764. https://doi.org/10.1007/s11121-015-0552-0

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology*, *146*, 348–361. https://doi.org/10.1007/PL00005481

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). Hierarchical clustering. In *Cluster analysis* (6th ed., pp. 71–110). Chichester: John Wiley & Sons.

Fabiano, G. A., Chafouleas, S. M., Weist, M. D., Carl Sumi, W., & Humphrey, N. (2014). Methodology considerations in school mental health research. *School Mental Health*, *6*, 68–83. https://doi.org/10.1007/s12310-013-9117-1

Farah, M. J., Shera, D. M., Savage, J. H., Betancourt, L., Giannetta, J. M., Brodsky, N. L., … Hurt, H. (2006). Childhood poverty: Specific associations with neurocognitive development. *Brain Research*, *1110*, 166–174. https://doi.org/10.1016/j.brainres.2006.06.072

Farrington, D. P. (2003). Methodological quality standards for evaluation research. *Annals of the American Academy of Political and Social Science*, *587*, 49–68. https://doi.org/10.1177/0002716202250789

Fernald, L. C., Gertler, P. J., & Neufeld, L. M. (2008). Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico's Oportunidades. *The Lancet*, *371*, 828–837. https://doi.org/10.1016/S0140-6736(08)60382-7

Fields, B. A. (1986). The nature and incidence of classroom behaviour problems and their remediation through preventive management. *Behaviour Change*, *3*, 53–57. https://doi.org/10.1017/S0813483900009116

Finegood, E. D., Raver, C. C., DeJoseph, M. L., & Blair, C. (2017). Parenting in poverty: Attention bias and anxiety interact to predict parents' perceptions of daily parenting hassles. *Journal of Family Psychology*, *31*, 51–60. https://doi.org/10.1037/fam0000291

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine, 15*, 451–474. https://doi.org/10.1016/0091-7435(86)90024-1

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., … Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*, 151–175. https://doi.org/10.1007/s11121-005-5553-y

Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., & Vega, R. (2014). Effects of the Good Behavior Game on challenging behaviors in school settings. *Review of Educational Research, 20*, 546–571. https://doi.org/10.3102/0034654314536781

Ford, C., Keegan, N., Poduska, J., Kellam, S., & Littman, J. (2014). Implementation MANUAL. Washington, DC: American Institute for Research.

Forman, S. G., Olin, S. S., Hoagwood, K. E., Crowe, M., & Saka, N. (2009). Evidence-based interventions in schools: Developers' views of implementation barriers and facilitators. *School Mental Health, 1*, 26–36. https://doi.org/10.1007/s12310-008-9002-5

Forman, S. G. (2015). *Implementation of mental health programs in schools: A change agent's guide*. Washington, DC: American Psychological Association.

Forsetlund, L., Chalmers, I., & Bjørndal, A. (2007). When was random allocation first used to generate comparison groups in experiments to assess the effects of social interventions? *Economics of Innovation and New Technology, 16*, 371–384. https://doi.org/10.1080/10438590600982467

Foxcroft, D. R. (2013). Can prevention classification be improved by considering the function of prevention? *Prevention Science, 15*, 818–822. https://doi.org/10.1007/s11121-013-0435-1

Fraser, M. W., & Galinsky, M. J. (2010). Steps in intervention research: Designing and developing social programs. *Research on Social Work Practice, 20*, 459–466. https://doi.org/10.1177/1049731509358424

Freud, S. (1920). *A general introduction to psychoanalysis.* New York: Horace Liveright. https://doi.org/10.1037/10667-000

Friedkin, N. E. (2004). Social cohesion. *Annual Review of Sociology, 30*, 409–425. https://doi.org/10.1146/annurev.soc.30.012703.110625

Friedman, M., & Michael, F. (1999). *Reconsidering logical positivism* (3rd ed.). Cambridge: Cambridge University Press. Retrieved from https://books.google.co.uk/books?id=e9TjZc9wNUAC

Frohlich, K. L., & Potvin, L. (2008). Transcending the known in public health practice: The inequality paradox: The population approach and vulnerable populations. *American Journal of Public Health, 98*, 216–221. https://doi.org/10.2105/AJPH.2007.114777

Galéra, C., Bouvard, M. P., Côté, S. M., Tremblay, R. E., Pingault, J.-B., Melchior, M., … Boivin, M. (2011). Early risk factors for hyperactivity-impulsivity and inattention trajectories from age 17 months to 8 years. *Archives of General Psychiatry, 68*, 1267–1275. https://doi.org/10.1001/archgenpsychiatry.2011.138

Gardner, F., Montgomery, P., & Knerr, W. (2016). Transporting evidence-based parenting programs for child problem behavior (age 3–10) between countries: Systematic

review and meta-analysis. *Journal of Clinical Child and Adolescent Psychology*, *45*, 749–762. https://doi.org/10.1080/15374416.2015.1015134

Garrett, P., Ng'andu, N., & Ferron, J. (1994). Poverty experiences of young children and the quality of their home environments. *Child Development*, *65*, 331–345. https://doi.org/10.1111/j.1467-8624.1994.tb00754.x

Garza, G., & Fisher Smith, A. (2009). Beyond neurobiological reductionism: Recovering the intentional and expressive body. *Theory & Psychology*, *19*, 519–544. https://doi.org/10.1177/0959354309336318

Gennetian, L. A., & Miller, C. (2002). Children and welfare reform: A view from an experimental welfare program in Minnesota. *Child Development*, *73*, 601–620. https://doi.org/10.1111/1467-8624.00426

Gifford-Smith, M. E., & Brownell, C. A. (2003). Childhood peer relationships: Social acceptance, friendships, and peer networks. *Journal of School Psychology*, *41*, 235–284. https://doi.org/10.1016/S0022-4405(03)00048-7

Giraudeau, B., & Ravaud, P. (2009). Preventing bias in cluster randomised trials. *PLoS Medicine*, *6*, 15–27. https://doi.org/10.1371/journal.pmed.1000065

Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health*, *93*, 1261–1267. https://doi.org/10.2105/AJPH.93.8.1261

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, *37*(2), 479–507. https://doi.org/10.3102/00028312037002479

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, *44*, 293–307. https://doi.org/10.3102/0013189x15592622

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Chichester: Wiley. Retrieved from https://books.google.co.uk/books?id=mdwt7ibSGUYC

Gorard, S. (2012). Who is eligible for free school meals? Characterising free school meals as a measure of disadvantage in England. *British Educational Research Journal*, *38*, 1003–1017. https://doi.org/10.1080/01411926.2011.608118

Gordon, A. D. (1987). A review of hierarchical classification. *Journal of the Royal Statistical Society. Series A (General)*, *150*, 119–137. https://doi.org/10.2307/2981629

Gordon, D., Levitas, R., Pantazis, C., Patsios, D., Payne, S., Townsend, P., … Williams, J. (2000). *Poverty and social exclusion in britain*. York. Retrieved from https://www.jrf.org.uk/sites/default/files/jrf/migrated/files/185935128x.pdf

Gordon, R. (1963). An operational classification of disease prevention. *Public Health Reports*, *96*, 107–109. https://doi.org/10.1037/e659532007-006

Gottfredson, D. C. (2000). *Schools and delinquency*. Cambridge: Cambridge University Press. Retrieved from https://books.google.co.uk/books?id=jy27NQX0CWIC

Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science*, *16*, 893–926. https://doi.org/10.1007/s11121-015-0555-x

Gottfredson, G. D., Gottfredson, D. C., Payne, A. A., & Gottfredson, N. C. (2005). School climate predictors of school disorder: Results from a national study of delinquency prevention in schools. *Journal of Research in Crime and Delinquency*, *42*, 412–444. https://doi.org/10.1177/0022427804271931

Gottfredson, G. D., Jones, E. M., & Gore, T. W. (2002). Implementation and evaluation of a cognitive-behavioral intervention to prevent problem behavior in a disorganized school. *Prevention Science*, *3*, 43–56. https://doi.org/10.1023/A:1014671310038

Gough, I. (2013). *Understanding prevention policy: a theoretical approach. LSE Research Online*. London, United Kingdom. Retrieved from http://eprints.lse.ac.uk/47951/

Graetz, B., Littlefield, L., Trinder, M., Dobia, B., Souter, M., Champion, C., … Cummins, R. (2008). KidsMatter: A population health model to support student mental health and well-being in primary schools. *International Journal of Mental Health Promotion*, *10*, 13–20. https://doi.org/10.1080/14623730.2008.9721772

Greenberg, M. T. (2010). School-based prevention: Current status and future challenges. *Effective Education*, *2*, 27–52. https://doi.org/10.1080/19415531003616862

Greenberg, M. T., & Abenavoli, R. (2017). Universal interventions: fully exploring their impacts and potential to produce population-level impacts. *Journal of Research on Educational Effectiveness*, *10*, 40–67. https://doi.org/10.1080/19345747.2016.1246632

Greenhalgh, T., Robert, G., Bate, P., Kyriakidou, O., Macfarlane, F., & Peacock, R. (2004). Diffusion of innovations in service organisations: Systematic review and recommendations, *The Milbank Quarterly, 82,* 581-629. https://doi.org/10.1111/j.0887-378X.2004.00325.x

Gregg, P., Propper, C., & Washbrook, E. (2008). *Understanding the relationship between parental income and multiple child outcomes: A decomposition analysis*. London. Retrieved from http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/wp193.pdf

Gregory, A., Henry, D. B., & Schoeny, M. E. (2007). School climate and implementation of a preventive intervention. *American Journal of Community Psychology*, *40*, 250–260. https://doi.org/10.1007/s10464-007-9142-z

Grix, J. (2004). *The foundations of research: a student's guide*. London: Palgrave Macmillan. Retrieved from https://books.google.co.uk/books?id=eqZyDwAAQBAJ

Guilbeault, D., Becker, J., & Centola, D. (2018). Complex contagions: A decade in review (pp. 3–25). https://doi.org/10.1007/978-3-319-77332-2_1

Gullotta, T. P. (2015). *Understanding primary prevention. Handbook of adolescent behavioral problems*. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4899-7497-6_1

Günther, I., & Maier, J. K. (2014). Poverty, vulnerability, and reference-dependent utility. *Review of Income and Wealth*, *60*, 155–181. https://doi.org/10.1111/roiw.12081

Gvion, Y., & Apter, A. (2011). Aggression, impulsivity, and suicide behavior: A review of the literature. *Archives of Suicide Research*, *15*, 93–112. https://doi.org/10.1080/13811118.2011.565265

Haataja, A., Voeten, M., Boulton, A. J., Ahtola, A., Poskiparta, E., & Salmivalli, C. (2014). The KiVa antibullying curriculum and outcome: Does fidelity matter? *Journal of School Psychology*, *52*, 479–493. https://doi.org/10.1016/j.jsp.2014.07.001

Haezewindt, P. (2003). Investing in each other and the community: the role of social capital. *Social Trends*, *33*, 19–27. Retrieved from http://www.statistics.gov.uk/articles/social_trends/SocialTrends33Article.pdf

Hagermoser Sanetti, L. M., & Fallon, L. M. (2011). Treatment integrity assessment: How estimates of adherence, quality, and exposure influence interpretation of implementation. *Journal of Educational and Psychological Consultation*, *21*, 209–232. https://doi.org/10.1080/10474412.2011.595163

Hahn, R., Fuqua-Whitley, D., Wethington, H., Lowy, J., Crosby, A., Fullilove, M., … Dahlberg, L. (2007). Effectiveness of universal school-based programs to prevent violent and aggressive behavior. *American Journal of Preventive Medicine*, *33*, S114–S129. https://doi.org/10.1016/j.amepre.2007.04.012

Hahn, S., Puffer, S., Torgerson, D. J., & Watson, J. (2005). Methodological bias in cluster randomised trials. *BMC Medical Research Methodology*, *5*, 1–8. https://doi.org/10.1186/1471-2288-5-10

Halleröd, B., Bradshaw, J., & Holmes, H. (1997). Adapting the consensual definition of poverty. In D. Gordon & C. Pantazis (Eds.), *Breadline Britain in the 1990s* (pp. 213–234). Aldershot: Ashgate. https://doi.org/10.4324/9780429460173-11

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*, 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Hamad, R., & Rehkopf, D. H. (2016). Poverty and child development: A longitudinal study of the impact of the earned income tax credit. *American Journal of Epidemiology*, *183*, 775–784. https://doi.org/10.1093/aje/kwv317

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

Hanley, P., Bonke, J., Slavin, B., Elliott, L., & Croudace, T. (2016). *Let' s think secondary science*. london. Retrieved from https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/EEF_Project_Protocol_LetsThinkSecondaryScience.pdf

Hansen, W. B., Graham, J. W., Wolkenstein, B. H., & Rohrbach, L. A. (1991). Program integrity as a moderator of prevention program effectiveness: Results for fifth-grade students in the adolescent alcohol prevention trial. *Journal of Studies on Alcohol*, *52*, 568–579. https://doi.org/10.15288/jsa.1991.52.568

Hansen, W. B., Pankratz, M. M., & Bishop, D. C. (2014). differences in observers' and teachers' fidelity assessments. *Journal of Primary Prevention*, *35*, 297–308. https://doi.org/10.1007/s10935-014-0351-6

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools? *Exceptional Children, 79*, 181–193. https://doi.org/10.1177/001440291307900204

Hartfree, Y., & Collard, S. (2014). *Poverty, debt and credit: An expert-led review*. Bristol: The University of Bristol.

Henkel, D. (2011). Unemployment and substance use: A review of the literature (1990-2010). *Current Drug Abuse Reviewse, 4*, 4–27. https://doi.org/10.2174/1874473711104010004

Hobbs, G., & Vignoles, A. (2007). *Is free school meal status a valid proxy for socio-economic status?* London.

Hogue, A., Liddle, H. A., Singer, A., & Leckrone, J. (2005). Intervention fidelity in family-based prevention counseling for adolescent problem behaviors. *Journal of Community Psychology, 33*, 191–211. https://doi.org/10.1002/jcop.20031

Holmes, J., & Kiernan, K. (2013). Persistent poverty and children's development in the early years of childhood. *Policy and Politics, 41*, 19–42. https://doi.org/10.1332/030557312X645810

Holsen, I., Iversen, A. C., & Smith, B. H. (2009). Universal social competence promotion programme in school: Does it work for children with low socio-economic background? *Advances in School Mental Health Promotion, 2*, 51–60. https://doi.org/10.1080/1754730X.2009.9715704

Hopson, L. M., & Lee, E. (2011). Mitigating the effect of family poverty on academic and behavioral outcomes: The role of school climate in middle and high school. *Children and Youth Services Review, 33*, 2221–2229. https://doi.org/10.1016/j.childyouth.2011.07.006

Howitt, D., & Cramer, D. (2011). *Introduction to Research Methods in Psychology*. Prentice Hall. Retrieved from https://books.google.co.uk/books?id=pIqpbwAACAAJ

Hudson, S., Levickis, P., Down, K., Nicholls, R., & Wake, M. (2015). Maternal responsiveness predicts child language at ages 3 and 4 in a community-based sample of slow-to-talk toddlers. *International Journal of Language and Communication Disorders, 50*, 136–142. https://doi.org/10.1111/1460-6984.12129

Humphrey, N., Barlow, A., & Lendrum, A. (2018). Quality matters: Implementation moderates student outcomes in the PATHS curriculum. *Prevention Science, 19*, 197–208. https://doi.org/10.1007/s11121-017-0802-4

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). *Implementation and process evaluation (IPE) for interventions in educational settings: An introductory handbook. Education Endowment Foundation*. London. https://doi.org/10.1017/CBO9781107415324.004

Humphrey, N., & Wigelsworth, M. (2016). Making the case for universal school-based mental health screening. *Emotional and Behavioural Difficulties, 21*, 22–42. https://doi.org/10.1080/13632752.2015.1120051

Hussain, M. A., Elyas, T., & Nasseef, O. A. (2013). Research paradigms: A slippery slope for fresh researchers. *Life Science Journal, 10*, 2374–2381. Retrieved from http://www.lifesciencesite.com

Hynes, K., Smith, E. P., & Perkins, D. (2009). Piloting a classroom-based intervention in after-school programmes: A case study in science migration. *Journal of Child Services*, *4*, 4–20. https://doi.org/10.1108/17466660200900014.Piloting

Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence. *Journal of Emotional and Behavioral Disorders*, *9*, 146–160. https://doi.org/10.1177/106342660100900301

Janlert, U., Winefield, A. H., & Hammarström, A. (2015). Length of unemployment and health-related outcomes: A life-course analysis. *European Journal of Public Health*, *25*, 662–667. https://doi.org/10.1093/eurpub/cku186

Jennings, P. A., & Greenberg, M. T. (2009). The pro-social classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, *79*, 491–525. https://doi.org/10.3102/0034654308325693

Jepson, E., & Forrest, S. (2006). Individual contributory factors in teacher stress: The role of achievement striving and occupational commitment. *British Journal of Educational Psychology*, *76*, 183–197. https://doi.org/10.1348/000709905X37299

Jones, H. (1893). Idealism and epistemology. *Mind*, *2*(7), 289–306. https://doi.org/10.1093/mind/II.7.289

Jones, S. M., Brown, J. L., & Lawrence Aber, J. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development*, *82*, 533–554. https://doi.org/10.1111/j.1467-8624.2010.01560.x

Juni, P., Altman, D. G., & Egger, M. (2001). Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*, *323*, 42–46. https://doi.org/10.1136/bmj.323.7303.42

Kam, C., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science*, *4*, 55–63. https://doi.org/10.1023/A:1021786811186

Kaplan, A., Gheen, M., & Midgley, C. (2002). Classroom goal structure and student disruptive behaviour. *British Journal of Educational Psychology*, *72*, 191–211. https://doi.org/10.1348/000709902158847

Karelis, C. (2007). *The persistence of poverty*. New Haven: Yale University Press.

Kasper, H. (1967). The asking price of labor and the duration of unemployment. *The Review of Economics and Statistics*, *49*, 165–172. https://doi.org/10.2307/1928224

Kauder, E. (1953). Genesis of the marginal utility theory: From aristotle to the end of the eighteenth century. *The Economic Journal*, *63*, 638. https://doi.org/10.2307/2226451

Kellam, S. G., Branch, J., & Agrawal, D. C. (1975). *Mental health and going to school: The Woodlawn program of assessment, early intervention, and evaluation*. Chicago: University of Chicago Press.

Kellam, S. G., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child

Psychology and Psychiatry and Allied Disciplines. https://doi.org/10.1111/j.1469-7610.1994.tb02308.x

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*, *10*, 169–170. https://doi.org/10.1017/S0954579498001564

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

Kellam, S. G., Mackenzie, A. C. L., Brown, C. H., Poduska, J. M., Wang, W., Petras, H., & Wilcox, H. C. (2011). The Good Behavior Game and the future of prevention and treatment. *Addiction Science & Clinical Practice*, *6*, 73–84. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22003425%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3188824

Kellam, S. G., Wang, W., Mackenzie, A. C. L., Brown, C. H., Ompad, D. C., Or, F., … Windham, A. (2014). The Impact of the Good Behavior Game, a Universal Classroom-Based Preventive Intervention in First and Second Grades, on High-Risk Sexual Behaviors and Drug Abuse and Dependence Disorders into Young Adulthood. *Prevention Science*, *15*, 6–18. https://doi.org/10.1007/s11121-012-0296-z

Kelly, J. A., Murphy, D. A., Sikkema, K. J., McAuliffe, T. L., Roffman, R. A., Solomon, L. J., … Kalichman, S. C. (1997). Randomised, controlled, community-level HIV-prevention intervention for sexual-risk behaviour among homosexual men in US cities. *The Lancet*, *350*, 1500–1505. https://doi.org/10.1016/S0140-6736(97)07439-4

Keynes, J. M. (1937). The general theory of employment. *The Quarterly Journal of Economics*, *51*, 209–223. https://doi.org/10.2307/1882087

Khan, S., Combaz, E., & McAslan Fraser, E. (2015). *Social exclusion: Topic guide*. Birmingham. Retrieved from http://www.gsdrc.org/go/topic-guides/social-exclusion/definitions-and-different-understandings-of-social-exclusion

Kidger, J., Brockman, R., Tilling, K., Campbell, R., Ford, T., Araya, R., … Gunnell, D. (2016). Teachers' wellbeing and depressive symptoms, and associated risk factors: A large cross sectional study in English secondary schools. *Journal of Affective Disorders*, *192*, 76–82. https://doi.org/10.1016/j.jad.2015.11.054

Kim, P., Evans, G. W., Angstadt, M., Ho, S. S., Sripada, C. S., Swain, J. E., … Phan, K. L. (2013). Effects of childhood poverty and chronic stress on emotion regulatory brain function in adulthood. *Proceedings of the National Academy of Sciences*, *110*, 18442–18447. https://doi.org/10.1073/pnas.1308240110

Kim, S., & Kochanska, G. (2015). Mothers' power assertion; children's negative, adversarial orientation; and future behavior problems in low-income families: Early maternal responsiveness as a moderator of the developmental cascade. *Journal of Family Psychology*, *29*, 1–9. https://doi.org/10.1037/a0038430

Kim, T. J., & von dem Knesebeck, O. (2016). Perceived job insecurity, unemployment and depressive symptoms: A systematic review and meta-analysis of prospective observational studies. *International Archives of Occupational and Environmental Health*, *89*(4), 561–573. https://doi.org/10.1007/s00420-015-1107-1

King, B., & Kasim, A. (2015). *Evaluation of Rapid Phonics*. London: The Education Endowment Foundation.

King, M. B., & Bouchard, K. (2011). The capacity to build organizational capacity in schools. *Journal of Educational Administration*, *49*, 653–669. https://doi.org/10.1108/09578231111174802

Klassen, R. M., Usher, E. L., & Bong, M. (2010). Teachers' collective efficacy, job satisfaction, and job stress in a cross-cultural context. *Journal of Experimental Education*, *78*, 464–486. https://doi.org/10.1080/00220970903292975

Kleinman, K. E., & Saigh, P. A. (2011). The effects of the Good Behavior Game on the conduct of regular education New York City high school students. *Behavior Modification*, *35*, 95–105. https://doi.org/10.1177/0145445510392213

Kleiver, A., & Cash, R. (2005). Characteristics of a public health model of mental health service delivery. *NASP Communique*, *34*, 22–45.

Klimes-Dougan, B., August, G. J., Lee, C. Y. S., Realmuto, G. M., Bloomquist, M. L., Horowitz, J. L., & Eisenberg, T. L. (2009). Practitioner and site characteristics that relate to fidelity of implementation: The early risers prevention program in a going-to-scale intervention trial. *Professional Psychology: Research and Practice*, *40*, 467–475. https://doi.org/10.1037/a0014623

Knapp, M., Mcdaid, D., & Parsonage, M. (2011). *Mental health promotion and mental illness prevention: The economic case. Mental health promotion and mental illness prevention : The economic case*. London. Retrieved from http://eprints.lse.ac.uk/id/eprint/32311

Korenman, S., Miller, J. E., & Sjaastad, J. E. (1994). *Long-term poverty and child development in the United States: Results from the NLSY* (No. 1044–94). https://doi.org/10.1016/0190-7409(95)00006-X

Kornhauser, R. (1978). *Social sources of delinquency: An appraisal of analytic models*. Chicago: University of Chicago Press. Retrieved from http://www.ncjrs.gov/App/publications/abstract.aspx?ID=55611

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, *86*, 643–680. https://doi.org/10.3102/0034654315626799

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2009). Teacher observation of classroom adaptation—checklist: development and factor structure. *Measurement and Evaluation in Counseling and Development*, *42*, 15–30. https://doi.org/10.1177/0748175609333560

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: a review of studies. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, *33*, 279–299. https://doi.org/10.1177/0888406410371643

Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: A dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of Abnormal Psychology*, *114*, 537–550. https://doi.org/10.1037/0021-843X.114.4.537

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kumpfer, K., Magalhães, C., & Xie, J. (2017). Cultural adaptation and implementation of family evidence-based interventions with diverse populations. *Prevention Science*, *18*, 649–659. https://doi.org/10.1007/s11121-016-0719-3

Kyriacou, C. (1987). Teacher stress and burnout: an international review. *Educational Research*, *29*, 146–152. https://doi.org/10.1080/0013188870290207

Lakoff, G. (1987). *Women, fire and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press.

Lange, B. C. L., Dáu, A. L. B. T., Goldblum, J., Alfano, J., & Smith, M. V. (2017). A mixed methods investigation of the experience of poverty among a population of low-income parenting women. *Community Mental Health Journal*, *53*, 832–841. https://doi.org/10.1007/s10597-017-0093-z

Langley, A. K., Nadeem, E., Kataoka, S. H., Stein, B. D., & Jaycox, L. H. (2010). Evidence-based mental health programs in schools: Barriers and facilitators of successful implementation. *School Mental Health*, *2*, 105–113. https://doi.org/10.1007/s12310-010-9038-1

Lansford, J. E., Malone, P. S., Tapanya, S., Tirado, L. M. U., Zelli, A., Alampay, L. P., … Steinberg, L. (2019). Household income predicts trajectories of child internalizing and externalizing behavior in high-, middle-, and low-income countries. *International Journal of Behavioral Development*, *43*, 74–79. https://doi.org/10.1177/0165025418783272

Larsson, H., Dilshad, R., Lichtenstein, P., & Barker, E. D. (2011). Developmental trajectories of DSM-IV symptoms of attention-deficit/ hyperactivity disorder: Genetic effects, family risk and associated psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *52*(9), 954–963. https://doi.org/10.1111/j.1469-7610.2011.02379.x

Latimer, K., Wilson, P., Kemp, J., Thompson, L., Sim, F., Gillberg, C., … Minnis, H. (2012). Disruptive behaviour disorders: A systematic review of environmental antenatal and early years risk factors. *Child: Care, Health and Development*, *38*, 611–628. https://doi.org/10.1111/j.1365-2214.2012.01366.x

Latkin, C. (1995). A personal network approach to aids prevention: an experimental peer group intervention for street-injecting drug users: The SAFE study. *NIDA Research Monograph*, *151*, 1–215. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2870820&tool=pmcentrez&rendertype=abstract

Lee, C.-Y. S., August, G. J., Realmuto, G. M., Horowitz, J. L., Bloomquist, M. L., & Klimes-Dougan, B. (2008). Fidelity at a distance: Assessing implementation fidelity of the Early Risers prevention program in a going-to-scale intervention trial. *Prevention Science*, *9*, 215–229. https://doi.org/10.1007/s11121-008-0097-6

Lee, S. J., Park, S.-H., Jeong, A., Chae, H., & Oh, K. (2017). Temperamental predictors of developmental trajectories of inattention and hyperactivity–impulsivity problems in schoolchildren. *Integrative Medicine Research*, *6*, 33–40. https://doi.org/10.1016/j.imr.2016.11.002

Leflot, G., Van Lier, P. A. C. C., Onghena, P., & 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. *Journal of School Psychology*, *51*, 187–199. https://doi.org/10.1016/j.jsp.2012.12.006

Lelkes, O., & Gasior, K. (2011). *Income poverty in the EU: Situation in 2007 and trends* (Vol. 1). Vienna: European Centre for Social Welfare Policy and Research.

Lendrum, A., & Humphrey, N. (2012). The importance of studying the implementation of interventions in school settings. *Oxford Review of Education*, *38*, 635–652. https://doi.org/10.1080/03054985.2012.734800

Lillehoj, C. J., Griffin, K. W., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education and Behavior*, *31*, 242–257. https://doi.org/10.1177/1090198103260514

Little, M. A., Sussman, S., Sun, P., & Rohrbach, L. A. (2013). The effects of implementation fidelity in the Towards No Drug Abuse dissemination trial. *Health Education*, *113*, 281–296. https://doi.org/10.1108/09654281311329231

Little, R. J. A. (1988). Missing-data adjustment in large surveys. *Journal of Business and Economic Statistics*, *6*, 287–296. https://doi.org/10.1080/07350015.1988.10509663

Loeber, R., & Stouthamer-loeber, M. (1998). Development of juvenile aggression and violence: Some common misconceptions and controversies. *American Ps*, *53*, 242–259. https://doi.org/10.1037/0003-066x.53.2.242

Low, S., Van Ryzin, M. J., Brown, E. C., Smith, B. H., & Haggerty, K. P. (2014). Engagement matters: lessons from assessing classroom implementation of steps to respect: a bullying prevention program over a one-year period. *Prevention Science*, *15*, 165–176. https://doi.org/10.1007/s11121-012-0359-1

Lupton, R. (2005). Social justice and school improvement: Improving the quality of schooling in the poorest neighbourhoods. *British Educational Research Journal*, *31*, 589–604. https://doi.org/10.1080/01411920500240759

Lynne, S. (2015). *Investigating the use of a positive variation of the Good Behavior Game in a high school setting*. The University of Southern Mississsippi, Hattisburg, United States of America.

Maas, C. J., & Hox, J. (2005). Sufficient sample sizes for multilevel modelling, *Methodology*, *1*, 86–92. https://doi.org/10.1027/1614-1881.1.3.86

MacDonald, G. (1997). Social work: Beyond control. In A. Maynard & I. Chalmers (Eds.), *Non-random reflections on health services research* (pp. 34–56). London: BMJ Publications.

Mack, J. (2016). Income Threshold Approach. Retrieved on 21 April, 2010, from http://www.poverty.ac.uk/definitions-poverty/income-threshold-approach

Mack, J., & Lansley, S. (1985). *Poor Britain*. London: George Allen & Unwin. https://doi.org/10.1215/03616878-11-2-345

MacKinnon, D. P. (2011). Integrating mediators and moderators in research design. *Research on Social Work Practice*, *21*, 675–681. https://doi.org/10.1177/1049731511414148

MacKinnon, D. P., & Luecken, L. J. (2008). How and for whom? Mediation and moderation in health psychology. *Health Psychology*, *27*, S99–S100. https://doi.org/10.1037/0278-6133.27.2(Suppl.).S99

Malloy, M., Acock, A., DuBois, D. L., Vuchinich, S., Silverthorn, N., Ji, P., & Flay, B. R. (2015). Teachers' perceptions of school organizational climate as predictors of dosage and quality of implementation of a social-emotional and character development program. *Prevention Science*, *16*, 1086–1095. https://doi.org/10.1007/s11121-014-0534-7

Martin, C. S., Earleywine, M., Blackson, T. C., Vanyukov, M. M., Moss, H. B., & Tarter, R. E. (1994). Aggressivity, inattention, hyperactivity, and impulsivity in boys at high and low risk for substance abuse. *Journal of Abnormal Child Psychology*, *22*, 177–203. https://doi.org/10.1007/BF02167899

Marx, K. (1848). *The communist manifesto*. (S. Moore, Ed.) (10th ed.). London: Penguin Classics.

Mayer, S. (n.d.). *What money can't buy: Family income and children's life chances*. Cambridge, MA: Havard University Press.

Mazza, J. R. S. E., Boivin, M., Tremblay, R. E., Michel, G., Salla, J., Lambert, J., … Côté, S. M. (2016). Poverty and behavior problems trajectories from 1.5 to 8 years of age: Is the gap widening between poor and non-poor children? *Social Psychiatry and Psychiatric Epidemiology*, *51*, 1083–1092. https://doi.org/10.1007/s00127-016-1252-1

McAdam, D. (1986). Recruitment to high-risk activism: The case of freedom summer. *American Journal of Sociology*, *92*, 64–90. Retrieved from http://www.jstor.org/stable/2779717

McClelland, M. M., & Cameron, C. E. (2012). Self-regulation early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives*, *6*(2), 136–142. https://doi.org/10.1111/j.1750-8606.2011.00191.x

McCormick, L. K., Steckler, A. B., & McLeroy, K. R. (1995). Diffusion of innovations in schools: A study of adoption and implementation of school-based tobacco prevention curricula. *American Journal of Health Promotion*, *9*, 210–219. https://doi.org/10.4278/0890-1171-9.3.210

McCurdy, B. L., Lannie, A. L., & Barnabas, E. (2009). Reducing disruptive behavior in an urban school cafeteria: An extension of the Good Behavior Game. *Journal of School Psychology*, *47*, 39–54. https://doi.org/10.1016/j.jsp.2008.09.003

McEvoy, A., & Welker, R. (2000). Antisocial behavior, academic failure, and school climate. *Journal of Emotional and Behavioral Disorders*, *8*(3), 130–140. https://doi.org/10.1177/106342660000800301

McEwen, B. S., & Gianaros, P. J. (2011). Stress- and allostasis-induced brain plasticity. *Annual Review of Medicine*, *62*, 431–445. https://doi.org/10.1146/annurev-med-052209-100430

McGilloway, S., Mhaille, G. N., Bywater, T., Furlong, M., Leckey, Y., Kelly, P., … Donnelly, M. (2012). A parenting intervention for childhood behavioral problems: A randomized controlled trial in disadvantaged community-based settings. *Journal of Consulting and Clinical Psychology*, *80*, 116–127. https://doi.org/10.1037/a0026304

McGuinness, F. (2018). *Poverty in the UK: Statistics. University of California, Berkeley*. London. Retrieved from http://scholar.google.co.uk/scholar?q=townsend+poverty+in+the+uk&hl=en&as_sdt=0,5&as_ylo=1976&as_yhi=1979#0

McKay, S. (2005). Poverty or preference: what do 'consensual deprivation indicators' really mean? *Fiscal Studies*, *25*, 201–223. https://doi.org/10.1111/j.1475-5890.2004.tb00102.x

McKenna, C. S., Law, C., & Pearce, A. (2017). Financial strain, parental smoking, and the great recession: An analysis of the uk millennium cohort study. *Nicotine & Tobacco Research*, *19*, 1521–1525. https://doi.org/10.1093/ntr/ntw269

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, *53*, 185–204. https://doi.org/10.1037/0003-066x.53.2.185

Medical Research Council. (2002). *Cluster randomised trials: methodological and ethical considerations*. London: Medical Research Council.

Mihalic, S. (2002). The importance of implementation fidelity, *Emotional and Behavioral Disorders in Youth, 4*, 83-105. https://doi.org/10.1017/CBO9781107415324.004

Miller, L. D., Laye-Gindhu, A., Bennett, J. L., Liu, Y., Gold, S., March, J. S., … Waechtler, V. E. (2011). An effectiveness study of a culturally enriched school-based CBT anxiety prevention program. *Journal of Clinical Child & Adolescent Psychology*, *40*(4), 618–629. https://doi.org/10.1080/15374416.2011.581619

Milligan, K., & Stabile, M. (2011). Do child tax benefits affect the well-being of children? evidence from Canadian child benefit expansions. *American Economic Journal: Economic Policy*, *3*, 175–205. https://doi.org/10.1257/pol.3.3.175

Mitchell, R. R., Tingstrom, D. H., Dufrene, B. A., Ford, W. B., & Sterling, H. E. (2015). The effects of the Good Behavior Game with general-education high school students. *School Psychology Review*, *44*, 191–207. https://doi.org/http://dx.doi.org/10.17105/spr-14-0063.1

Mocan, N., Raschke, C., & Unel, B. (2015). The impact of mothers' earnings on health inputs and infant health. *Economics and Human Biology*, *19*, 204–223. https://doi.org/10.1016/j.ehb.2015.08.008

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., … Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ (Clinical Research Ed.)*, *350*, 1–6. https://doi.org/10.1136/bmj.h1258

Morrison, K. (2009). Randomised controlled trials for evidence-based education: Some problems in judging "what works." *Evaluation & Research in Education*, *15*, 69–83. https://doi.org/10.1080/09500790108666984

Moskowitz, J. M., Malvin, J. H., Schaeffer, G. A., & Schaps, E. (1984). An experimental evaluation of a drug education course. *Journal of Drug Education*, *14*, 9–22. https://doi.org/10.2190/R977-NCC7-9ET5-3PLX

Murali, V., & Oyebode, F. (2004). Poverty, social inequality and mental health. *Advances in Psychiatric Treatment*, *10*, 216–224. https://doi.org/10.1192/apt.10.3.216

Nagin, D., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile

delinquency. *Child Development*, *70*, 1181–1196. https://doi.org/10.1111/1467-8624.00086

Nastasi, B. K., Hitchcock, J., Sarkar, S., Burkholder, G., Varjas, K., & Jayasena, A. (2007). Mixed methods in intervention research: Theory to adaptation. *Journal of Mixed Methods Research*, *1*, 164–182. https://doi.org/10.1177/1558689806298181

Nation, M., Crusto, C., Wandersman, A., Kumpfer, K. L., Seybolt, D., Morrissey-Kane, E., & Davino, K. (2003). What works in prevention: Principles of effective prevention programs. *American Psychologist*, *58*, 449–456. https://doi.org/10.1037/0003-066X.58.6-7.449

Naylor, P.-J., Nettlefold, L., Race, D., Hoy, C., Ashe, M. C., Wharf Higgins, J., & McKay, H. a. (2015). Implementation of school based physical activity interventions: A systematic review. *Preventive Medicine*, *72*, 95–115. https://doi.org/10.1016/j.ypmed.2014.12.034

Neppl, T. K., Senia, J. M., & Donnellan, M. B. (2016). Effects of economic hardship: Testing the family stress model over time. *Journal of Family Psychology*, *30*, 12–21. https://doi.org/10.1037/fam0000168

Newmann, F. M., King, M. B., & Youngs, P. (2005). Professional development that addresses school capacity: Lessons from urban elementary schools. *American Journal of Education*, *108*, 259–299. https://doi.org/10.1086/444249

NICHD Early Child Care Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood: Predictors, correlates, and outcomes. *Monographs of the Society for Research in Child Development*, *69*, 1–143. https://doi.org/10.4135/9781452286143.n496

Nielsen, L., Meilstrup, C., Nelausen, M. K., Koushede, V., & Holstein, B. E. (2015). Promotion of social and emotional competence: Experiences from a mental health intervention applying a whole school approach. *Health Education*, *115*, 339–356. https://doi.org/10.1108/HE-03-2014-0039

Nikulina, V., Widom, C. S., & Czaja, S. (2011). The role of childhood neglect and childhood poverty in predicting mental health, academic achievement and crime in adulthood. *American Journal of Community Psychology*, *48*, 309–321. https://doi.org/10.1007/s10464-010-9385-y

Nolan, B. (2007). *A comparative perspective on the development of poverty and exclusion in european societies*. Bonn: International Policy Analysis.

Nusslock, R., & Miller, G. E. (2016). Early-life adversity and physical and emotional health across the lifespan: A neuroimmune network hypothesis. *Biological Psychiatry*, *80*, 23–32. https://doi.org/10.1016/j.biopsych.2015.05.017

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*, 33–84. https://doi.org/10.3102/0034654307313793

Oakley, A. (1998). Experimentation and social interventions: A forgotten but important history. *BMJ*, *317*, 1239–1242. https://doi.org/10.1136/bmj.317.7167.1239

Oberwittler, D. (2007). The effects of neighbourhood poverty on adolescent problem behaviours: A multi-level analysis differentiated by gender and ethnicity. *Housing Studies*, *22*, 781–803. https://doi.org/10.1080/02673030701474727

Office for National Statistics. (2015). *English Indices of Deprivation 2015 Statistical Release*. London: Office for National Statistics.

Ofsted. (2014). *Below the radar : low-level disruption in the country's classrooms*. London: Ofsted. Retrieved from www.ofsted.gov.uk/resources/140157.

Olweus, D. (1993). *Mental disorder and crime*. Thousand Oaks: SAGE Publications.

Oriana Linares, L., Montalto, D., Li, M. M., & Oza, V. S. (2006). A promising parenting intervention in foster care. *Journal of Consulting and Clinical Psychology*, *74*, 32–41. https://doi.org/10.1037/0022-006X.74.1.32

Österman, K., Björkqvist, K., Lagerspetz, K. M. J., Kaukiainen, A., Landau, S. F., Frączek, A., & Caprara, G. V. (1998). Cross-cultural evidence of female indirect aggression. *Aggressive Behavior*, *24*, 1–8. https://doi.org/10.1002/1098-2337

Ouellette, R. R., Frazier, S. L., Shernoff, E. S., Cappella, E., Mehta, T. G., Maríñez-Lora, A., … Atkins, M. S. (2018). Teacher job stress and satisfaction in urban schools: disentangling individual-, classroom-, and organizational-level influences. *Behavior Therapy*, *49*(4), 494–508. https://doi.org/10.1016/j.beth.2017.11.011

Owens, L., Shute, R., & Slee, P. (2000). "Guess what I just heard!": Indirect aggression among teenage girls in Australia. *Aggressive Behavior*, *26*, 67–83. https://doi.org/10.1002/(SICI)1098-2337(2000)26:1<67::AID-AB6>3.0.CO;2-C

Ozer, E. J. (2006). Contextual effects in school-based violence prevention programs: A conceptual framework and empirical review. *The Journal of Primary Prevention*, *27*, 315–340. https://doi.org/10.1007/s10935-006-0036-x

Ozer, E. J., Weinstein, R. S., Maslach, C., & Siegel, D. (1997). Adolescent AIDS prevention in context: The Impact of peer educator qualities and classroom environments on intervention efficacy. *American Journal of Community Psychology*, *25*, 289–323. https://doi.org/10.1023/A:1024624610117

Padley, M., & Hirsch, D. (2011). *A minimum income standard for the UK in 2011* (No. 3263 21). York: The Joseph Rowntree Foundation. Retrieved from http://www.jrf.org.uk/sites/files/jrf/minimum-income-standards-rural-full.pdf%5Cnhttps://dspace.lboro.ac.uk/dspace-jspui/handle/2134/13295

Pagani, L., Boulerice, B., Vitaro, F., & Tremblay, R. E. (1999). Effects of poverty on academic failure and delinquency in boys: A change and process model approach. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*, 1209–1219. https://doi.org/10.1017/S0021963099004783

Pager, D. (2003). The mark of a criminal record. *American Journal of Sociology*, *108*, 937–975. https://doi.org/10.1086/374403

Pampaka, M., Hutcheson, G., & Julian, W. (2016). Handling missing data: Analysis of a challenging data set using multiple imputation. *International Journal of Research & Method in Education*, *39*, 19–37. https://doi.org/10.1080/1743727X.2014.979146

Papatheodorou, T. (2005). *Behaviour problems in the early years*. London: Routledge. https://doi.org/10.4324/9780203464656

Parker, J. D., Schoendorf, K. C., & Kiely, J. L. (1994). Associations between measures of socioeconomic status and low birth weight, small for gestational age, and premature delivery in the United States. *Annals of Epidemiology*, *4*, 271–278. https://doi.org/10.1016/1047-2797(94)90082-5

Patchin, J. W., Huebner, B. M., McCluskey, J. D., Varano, S. P., & Bynum, T. S. (2006). Exposure to community violence and childhood delinquency. *Crime and Delinquency*, *52*, 307–332. https://doi.org/10.1177/0011128704267476

PATTERSON, E. B. (1991). Poverty, Income Inequality, and Community Crime Rates. *Criminology*, *29*, 755–776. https://doi.org/10.1111/j.1745-9125.1991.tb01087.x

Payne, A. A., & Eckert, R. (2010). The relative importance of provider, program, school, and community predictors of the implementation quality of school-based prevention programs. *Prevention Science*, *11*, 126–141. https://doi.org/10.1007/s11121-009-0157-6

Peace, R. (2001). Social exclusion: a concept in need of definition? *Social Policy Journal of New Zealand*, 16, 17–36. Retrieved from https://go.galegroup.com/ps/anonymous?id=GALE%7CA77930783&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=11724382&p=AONE&sw=w

Pearce, A., Hope, S., Lewis, H., & Law, C. (2014). Family structure and socio-emotional wellbeing in the early years: A life course approach. *Longitudinal and Life Course Studies*, *5*(3), 263–282. https://doi.org/10.14301/llcs.v5i3.264

Petras, H., Kellam, S. G., Brown, C. H., Muthén, B. O., Ialongo, N. S., & Poduska, J. M. (2008). Developmental epidemiological courses leading to antisocial personality disorder and violent and criminal behavior: Effects by young adulthood of a universal preventive intervention in first- and second-grade classrooms. *Drug and Alcohol Dependence*, *95*, S45–S59. https://doi.org/10.1016/j.drugalcdep.2007.10.015

Petras, H., Masyn, K., & Iaolongo, N. (2011). The developmental impact of two first grade preventive interventions on aggressive/disruptive behavior in childhood and adolescence: An application of latent transition growth mixture modeling. *Prevention Science*, *12*, 300–313. https://doi.org/10.1007/s11121-011-0216-7.The

Pettigrew, J., Graham, J. W., Miller-Day, M., Hecht, M. L., Krieger, J. L., & Shin, Y. J. (2014). Adherence and delivery: implementation quality and program outcomes for the seventh-grade Keepin' it REAL program. *Prevention Science*, *16*, 90–99. https://doi.org/10.1007/s11121-014-0459-1

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*, 525–556. https://doi.org/10.3102/00346543074004525

Pianta, R. C., la Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal*, *102*(3), 225–238. https://doi.org/10.1086/499701

Pichler, F., & Wallace, C. (2007). Patterns of formal and informal social capital in Europe. *European Sociological Review*, *23*, 423–435. https://doi.org/10.1093/esr/jcm013

Pickett, K., & Wilkinson, R. G. (2009). *The spirit level: Why greater equality makes societies stronger.* New York: Bloomsbury Press.

Piotrowska, P. J., Stride, C. B., Croft, S. E., & Rowe, R. (2015). Socioeconomic status and antisocial behaviour among children and adolescents: A systematic review and meta-analysis. *Clinical Psychology Review*, *35*, 47–55. https://doi.org/10.1016/j.cpr.2014.11.003

Poduska, J., Kellam, S., Wang, W., Brown, C. H., Ialongo, N., & Toyinbo, P. (2009). 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. https://doi.org/10.1016/j.drugalcdep.2007.10.009.Impact

Polanin, J. R., & Espelage, D. L. (2015). Using a meta-analytic technique to assess the relationship between treatment intensity and program effects in a cluster-randomized trial. *Journal of Behavioral Education*, *24*, 133–151. https://doi.org/10.1007/s10864-014-9205-9

Popper, K. (1972). *Objective knowledge*. Oxford: Clarendon Press.

Popper, K. (1963). *Conjectures and refutations*. London: Routledge.

Pradella, L. (2015). The working poor in western Europe: Labour, poverty and global capitalism. *Comparative European Politics*, *13*, 596–613. https://doi.org/10.1057/cep.2015.17

Puffer, S., Torgerson, D. J., & Watson, J. (2003). Evidence for risk of bias in cluster randomised trials: Review of recent trials published in three general medical journals. *BMJ*, *327*, 785–789. https://doi.org/10.1136/bmj.327.7418.785

Pulerwitz, J., Barker, G., & Segundo, M. (2004). *Promoting healthy relationships and hiv/sti prevention for young men: positive findings from an intervention study in Brazil*. Washington: The Population Council.

Pullen, N. A., Raulston, T. J., Lang, R., & Wendt, O. (2013). A modified preschool life skills program focused on teaching functional communication and self-control resulted in acquisition of requesting skills and the prevention of problem behavior in small groups of preschool children. *Evidence-Based Communication Assessment and Intervention*, *7*, 112–116. https://doi.org/10.1080/17489539.2013.874646

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of american community*. New York: Simon & Schuster. Retrieved from http://bowlingalone.com/

Putnam, R. D. (1995). Tuning in, tuning out: The strange disappearance of social capital in america. *PS: Political Science and Politics*, *28*, 664–686. https://doi.org/10.2307/420517

Ransford, C. R., Greenberg, M. T., Domitrovich, C. E., Small, M., & Jacobson, L. (2009). The role of teachers' psychological experiences and perceptions of curriculum supports on the implementation of a social and emotional learning curriculum. *School Psychology Review*, *38*, 510–532. Retrieved from https://search.proquest.com/docview/219656424?accountid=11528

Ravallion, M., & Chen, S. (2011). Weakly relative poverty. *Review of Economics and Statistics*, *93*, 1251–1261. https://doi.org/10.1162/REST_a_00127

Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive

behavior. *School Psychology Review*, *37*, 315–332. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2603055/

Resnicow, K. A., Davis, M., Smith, M., Lazarus-Yaroch, A., Baranowski, T., Baranowski, J., … Wang, D. T. (1998). How best to measure implementation of school health curricula: A comparison of three measures. *Health Education Research*, *13*, 239–250. https://doi.org/10.1093/her/13.2.239

Resnicow, K., Soler, R., Braithwaite, R. L., Ahluwalia, J. S., & Butler, J. (2000). Cultural sensitivity in substance use prevention. *Journal of Community Psychology*, *28*, 271–290. https://doi.org/10.1002/(SICI)1520-6629(200005)28:3<271::AID-JCOP4>3.0.CO;2-I

Ricardo, D. (1821). *On political economy and taxation*. London: J.M. Dent & Sons.

Ringen, S. (1988). Direct and Indirect Measures of Poverty. *Journal of Social Policy*, *17*, 351–365. https://doi.org/10.1017/S0047279400016858

Ringwalt, C. L., Ennett, S., Johnson, R., Rohrbach, L. A., Simons-rudolph, A., & Vincus, A. (2003). Factors associated with fidelity to substance use prevention curriculum guides in the nation' s middle schools. *Health Education and Behavior*, *30*, 375–391. https://doi.org/10.1177/1090198103253627

Rivers, C., & Barnett, R. C. (2011). *The truth about girls and boys: challenging toxic stereotypes about our children*. New York: Columbia University Press. Retrieved from http://www.jstor.org/stable/10.7312/rive15162.9

Roache, J., & Lewis, R. R. (2011). Teachers' views on the impact of classroom management on student responsibility. *Australian Journal of Education*, *55*, 132–146. https://doi.org/10.1177/000494411105500204

Rogers, E. (2003). *Diffusion of innovations* (5th ed.). New York, NY: The Free Press.

Rogers, E., & Shoemaker, F. (1971). *Communication of innovations; A cross-cultural approach*. New York: The Free Press.

Roggman, L., Boyce, L., & Innocenti, M. (2008). *Developmental parenting: A guide for early childhood practitioners*. Baltimore: Paul H. Brookes Publishing.

Rogosch, F. A., Dackis, M. N., & Cicchetti, D. (2011). Child maltreatment and allostatic load: Consequences for physical and mental health in children from low-income families. *Development and Psychopathology*, *23*, 1107–1124. https://doi.org/10.1017/s0954579411000587

Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: Predictors of program implementation. *Preventive Medicine*, *22*, 237–260. https://doi.org/10.1006/pmed.1993.1020

Rohrbach, L. A., Gunning, M., Sun, P., & Sussman, S. (2010). The project Towards No Drug Abuse (TND) dissemination trial: Implementation fidelity and immediate outcomes. *Prevention Science*, *11*, 77–88. https://doi.org/10.1007/s11121-009-0151-z

Romer, D., Betancourt, L. M., Brodsky, N. L., & Giannetta, J. M. (2012). Does adolescent risk-taking imply weak executive function?, *14*, 1119–1133. https://doi.org/10.1111/j.1467-7687.2011.01061.x.Does

Rose, G. (1992). *The Strategy of Preventive Medicine*. Oxford: Oxford University Press.

Rosenblatt, J. L., & Elias, M. J. (2008). Dosage effects of a preventive social-emotional learning intervention on achievement loss associated with middle school transition. *Journal of Primary Prevention*, *29*, 535–555. https://doi.org/10.1007/s10935-008-0153-9

Ross, S. W., Romer, N., & Horner, R. H. (2012). Teacher well-being and the implementation of school-wide positive behavior interventions and supports. *Journal of Positive Behavior Interventions*, *14*, 118–128. https://doi.org/10.1177/1098300711413820

Ruisch, I. H., Dietrich, A., Glennon, J. C., Buitelaar, J. K., & Hoekstra, P. J. (2017). Maternal substance use during pregnancy and offspring conduct problems: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *84*, 325–336. https://doi.org/10.1016/j.neubiorev.2017.08.014

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*, 1046–1058. https://doi.org/10.1002/pits.20523

Rutter, M. (1985). Resilience in the face of adversity: Protective factors and resistence to psychiatric disorder. *British Journal of Psychiatry*, *147*, 598–611. https://doi.org/10.1192/bjp.147.6.598

Sabey, C. V., Charlton, C. T., Pyle, D., Lignugaris-Kraft, B., & Ross, S. W. (2017). A review of classwide or universal social, emotional, behavioral programs for students in kindergarten. *Review of Educational Research*, *87*, 512–543. https://doi.org/10.3102/0034654316689307

Saigh, P. A., & Umar, A. M. (1983). The effects of a good behavior game on the disruptive behavior of Sudanese elementary school students. *Journal of Applied Behavior Analysis*, *16*, 339–344. https://doi.org/10.1901/jaba.1983.16-339

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent Crime: A multilevel study of collective efficacy. *Science*, *277*, 918–924. https://doi.org/10.1126/science.277.5328.918

Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: testing social-disorganization theory. *American Journal of Sociology*, *94*, 774–802. https://doi.org/10.1086/229068

Sanders, M. R. (2000). Community-based parenting and family support interventions and the prevention of drug abuse. *Addictive Behaviors*, *25*, 929–942. https://doi.org/10.1016/S0306-4603(00)00128-3

Say, J., Prinsep, C., & Biddle, C. (1821). *A treatise on political economy: The production, distribution, and consumption of wealth* (2nd ed.). Boston: Wells and Lilly.

Schein, E. H. (2010). *Organizational culture and leadership*. John Wiley & Sons. Retrieved from https://books.google.co.uk/books?id=DlGhlT34jCUC

Schneider-Rosen, K., & Cicchetti, D. (1984). The relationship between affect and cognition in maltreated infants: quality of attachment and the development of visual self-recognition. *Child Development*, *55*, 648–658. https://doi.org/10.1111/j.1467-8624.1984.tb00325.x

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine*, *152*, 726–731. https://doi.org/10.7326/0003-4819-152-11-201006010-00232

Scotland, J. (2012). Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. *English Language Teaching*, *5*, 9–16. https://doi.org/10.5539/elt.v5n9p9

Scott, S., Doolan, M., Beckett, C., Harry, S., & Cartwright, S. (2010). *Preliminary findings from the Helping Children Achieve study*. London. Retrieved from https://core.ac.uk/download/pdf/4162728.pdf

Scott, S., Sylva, K., Doolan, M., Price, J., Jacobs, B., Crook, C., & Landau, S. (2002). Evaluation of the first 3 years of the Fast Track prevention trial with children at high risk for adolescent conduct problems. *Journal of Abnormal Child Psychology*, *30*, 19–35. https://doi.org/http://dx.doi.org/10.1023/A:1014274914287

Sen, A. (2000). *The Social exclusion: concept, application, and scrutiny* (Social Development Papers No. 1). *The Social Exclusion: Concept, Application, And Scrutiny*. Manila: Asian Development Bank.

Shafter, P. (1995). Beneath the poverty debate: Some Issues. *IDS Bulletin*, *27*, 23–35.

Shaw, C., & McKay, H. (1942). *Juvenile delinquency and urban areas*. Chicago: University of Chicago Press.

Shaw, D. S., & Shelleby, E. C. (2014). Early-starting conduct problems: Intersection of conduct problems and poverty. *Annual Review of Clinical Psychology*, *10*, 503–528. https://doi.org/10.1146/annurev-clinpsy-032813-153650

Sheard, M., Chambers, B., & Elliott, L. (2015). *Units of Sound: Evaluation report and Executive summary*. London: The Education Endowment Foundation. https://doi.org/10.13140/RG.2.1.5064.4729

Shenkin, S. D., Starr, J. M., & Deary, I. J. (2004). Birth weight and cognitive ability in childhood: A systematic review. *Psychological Bulletin*, *130*, 989–1013. https://doi.org/10.1037/0033-2909.130.6.989

Shonkoff, J., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: The National Academies Press. https://doi.org/10.17226/9824

Silver. (1994). Social exclusion and social solidarity: Three paradigms. *International Labour Review*, *133*, 532–578. Retrieved from https://search.proquest.com/docview/224008056?accountid=11528

Simpson, J., & Weiner, E. (Eds.). (1989). *The oxford english dictionary*. Oxford: Clarendon Press.

Sims, S., & Allen, R. (2018). Identifying schools with high usage and high loss of newly qualified teachers. *National Institute Economic Review*, *243*, R27–R36. https://doi.org/10.1177/002795011824300112

Singal, A. G., Higgins, P. D. R., & Waljee, A. K. (2014). A primer on effectiveness and efficacy trials. *Clinical and Translational Gastroenterology*, *5*, 45-4. https://doi.org/10.1038/ctg.2013.13

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*, 417–453. https://doi.org/10.3102/00346543075003417

Skaalvik, E. M., & Skaalvik, S. (2016). teacher stress and teacher self-efficacy as predictors of engagement, emotional exhaustion, and motivation to leave the teaching profession. *Creative Education*, *07*, 1785–1799. https://doi.org/10.4236/ce.2016.713182

Skinner, B. (1948). 'Superstition' in the pigeon. *Journal of Experimental Psychology*, *38*, 168–172. https://doi.org/10.1037/h0055873

Smith, A., & Krueger, A. B. (2003). *The wealth of nations*. New York: Bantam Classics.

Smith, J., Brooks–Gunn, J., & Klebanov, P. K. (n.d.). Consequences of living in poverty for young children's cognitive and verbal ability and early school achievement. In G. Duncan & J. Brooks–Gunn (Eds.), *Consequences of growing up poor* (pp. 132–189). New York: Russell Sage Foundation. Retrieved from http://www.jstor.org/stable/10.7758/9781610448260

Smith, J. D., Schneider, B. H., Smith, P. K., & Ananiadou, K. (2004). The effectiveness of whole-school antibullying programs: A synthesis of evaluation research. *School Psychology Review*, *33*, 547–560. Retrieved from http://discovery.ucl.ac.uk/id/eprint/10008732

Sosu, E. M., & Schmidt, P. (2017). Economic deprivation and its effects on childhood conduct problems: The mediating role of family stress and investment factors. *Frontiers in Psychology*, *8*(SEP), 1–12. https://doi.org/10.3389/fpsyg.2017.01580

Spoth, R., Guyll, M., Lillehoj, C. J., Redmond, C., & Greenberg, M. (2007). Prosper study of evidence-based intervention implementation quality by community-university partnerships. *Journal of Community Psychology*, *35*, 981–999. https://doi.org/10.1002/jcop.20207

Spoth, R., Rohrbach, L. A., Greenberg, M., Leaf, P., Brown, C. H., Fagan, A., … Hawkins, J. D. (2013). Addressing core challenges for the next generation of type 2 translation research and systems: The translation science to population impact framework. *Prevention Science*, *14*, 319–351. https://doi.org/10.1007/s11121-012-0362-6

Stallard, P., & Buck, R. (2013). Preventing depression and promoting resilience: Feasibility study of a school-based cognitive-behavioural intervention. *British Journal of Psychiatry*, *202*, s18–s23. https://doi.org/10.1192/bjp.bp.112.119172

Stein, M. L., Berends, M., Fuchs, D., McMaster, K., Sáenz, L., Yen, L., … Compton, D. L. (2008). Scaling up an early reading program: relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Educational Evaluation and Policy Analysis*, *30*, 368–388. https://doi.org/10.3102/0162373708322738

Stephanou, G., Gkavras, G., & Doulkeridou, M. (2013). The role of teachers' self- and collective-efficacy beliefs on their job satisfaction and experienced emotions in school. *Psychology*, *04*, 268–278. https://doi.org/10.4236/psych.2013.43a040

Stephenson, J., Linfoot, K., & Martin, A. (2000). Behaviours of concern to teachers in the early years of school. *International Journal of Disability, Development and Education*, *47*, 225–235. https://doi.org/10.1080/713671118

Stewart, E. A. (2003). School social bonds, school climate, and school misbehavior: A multilevel analysis. *Justice Quarterly*, *20*, 575–604. https://doi.org/10.1080/07418820300095621

Stigler, G. J. (1950). The development of utility theory. *Journal of Political Economy*, *58*(5), 373–396. Retrieved from https://www.jstor.org/stable/1825710

Stormont, M., Reinke, W. M., Newcomer, L., Marchese, D., & Lewis, C. (2015). Coaching teachers' use of social behavior interventions to improve children's outcomes: A review of the literature. *Journal of Positive Behavior Interventions*, *17*, 69–82. https://doi.org/10.1177/1098300714550657

Stosich, E. L. (2016). Building teacher and school capacity to teach to ambitious standards in high-poverty schools. *Teaching and Teacher Education*, *58*, 43–53. https://doi.org/10.1016/j.tate.2016.04.010

Streiner, D. L., & Norman, G. R. (2009). Efficacy and effectiveness trials. *Community Oncology*, *6*, 472–474. https://doi.org/10.1016/S1548-5315(11)70286-8

Sweizy, N., Matson, J. L., & Box, P. (1992). The Good Behavior Game: A token reinforcement system for preschoolers. *Child & Family Behavior Therapy*, *14*, 21–32. https://doi.org/10.1300/J019v14n03_02

Swiezy, N. B., Matson, J. L., & Box, P. (2008). The Good Behaviour Game. *Child & Family Behavior Therapy*, *14*, 21–32. https://doi.org/10.1300/J019v14n03_02

Swift, L. E., Hubbard, J. A., Bookhout, M. K., Grassetti, S. N., Smith, M. A., & Morrow, M. T. (2017). Teacher factors contributing to dosage of the KiVa anti-bullying program. *Journal of School Psychology*, *65*, 102–115. https://doi.org/10.1016/j.jsp.2017.07.005

Tagiuri, R. (1968). The concept of organizational climate. In R. Tagiuri & L. Litwin (Eds.), *Organizational climate: Explorations of a concept* (pp. 43–75). Boston: Havard University Press.

Tan, S., & Karpatne, K. (2011). *Introduction to data mining* (2nd ed.). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118029145

Tangen, S. (2004). Demystifying productivity and performance. *International Journal of Productivity and Performance Management*, *54*, 34–46. https://doi.org/10.1108/17410400510571437

Taylor, A., Wilson, C., Slater, A., & Mohr, P. (2011). Parent- and child-reported parenting. associations with child weight-related outcomes. *Appetite*, *57*, 700–706. https://doi.org/10.1016/j.appet.2011.08.014

Taylor, R. B. (1996). Neighborhood responses to disorder and local attachments: The systemic model of attachment, social disorganization, and neighborhood use value. *Sociological Forum*, *11*, 41–74. https://doi.org/10.1007/BF02408301

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting Positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, *88*, 1156–1171. https://doi.org/10.1111/cdev.12864

Thaker, S., Steckler, A., Sanchez, V., Khatapoush, S., Rose, J., & Hallfors, D. D. (2007). Program characteristics and organizational factors affecting the implementation of a

school-based indicated prevention program. *Health Education Research*, *23*, 238–248. https://doi.org/10.1093/her/cym025

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*, 357–385. https://doi.org/10.3102/0034654313483907

The British Psychological Society. (2009). *Code of ethics and conduct*. Leicester. Retrieved from https://www.bps.org.uk/psychologists/standards-and-guidelines

The World Bank. (2017). *The world bank annual report 2017*. The World Bank. https://doi.org/10.1596/978-1-4648-1119-7

Thurow, L. (1972). Education and economic equality. *The Public Interest*, *28*, 69–81. Retrieved from https://search.proquest.com/docview/1298108098?accountid=11528

Tibbits, M. K., & Perkins, D. F. (2011). Conditions: results from a large-scale diffusion project. *Prevention Science*, *11*, 252–262. https://doi.org/10.1007/s11121-010-0170-9.Sustaining

Tine, M. (2017). Growing up in rural vs. urban poverty: Contextual, academic, and cognitive differences. *Poverty, Inequality and Policy*. https://doi.org/10.5772/intechopen.68581

Tine, M. (2014). working memory differences between children living in rural and urban poverty. *Journal of Cognition and Development*, *15*, 599–613. https://doi.org/10.1080/15248372.2013.797906

Torgerson, C. J., & Torgerson, D. J. (2001). The need for randomised controlled trials in educational research. *British Journal of Educational Studies*, *49*, 316–328. https://doi.org/10.1111/1467-8527.t01-1-00178

Torgerson, C., Ainsworth, H., Buckley, H., Hampden-Thompson, G., Hewitt, C., Humphry, D., … Torgerson, D. (2016). *Affordable online maths tuition: Evaluation report and executive summary*. London. Retrieved from https://educationendowmentfoundation.org.uk

Torgerson, C., & Torgerson, D. (2005). *Randomised trials in education : An introductory handbook Director , York Trials Unit*. York. Retrieved from http://educationendowmentfoundation.org.uk/uploads/pdf/Randomised_trials_in_education_revised.pdf

Treasure, T., & MacRae, K. D. (1998). Minimisation: The platinum standard for trials? *BMJ*, *317*(7155), 362–363. https://doi.org/10.1136/bmj.317.7155.362

Tremblay, M., Duchesne, S., Vitaro, F., & Tremblay, R. E. (2013). Developmental trajectories of oppositional behavior during elementary school and their risk factors. *Journal of Educational and Developmental Psychology*, *3*, 234–243. https://doi.org/10.5539/jedp.v3n1p234

Tremblay, R. E. (2010). Developmental origins of disruptive behaviour problems: The 'original sin' hypothesis, epigenetics and their consequences for prevention. *Child Pscyhology and Pscyhiatry*, *51*, 341–367. https://doi.org/10.1111/j.1469-7610.2010.02211.x

Trone, K. (2005). The effects of poverty on children, youth, and families. *Family Court Review*, *36*, 355–359. https://doi.org/10.1111/j.174-1617.1998.tb00517.x

Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology*, *7*, 27–56. https://doi.org/10.1007/s11292-010-9109-1

Twisk, J. W. R. (2006). *Applied multilevel analysis*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511610806

Underwood, M. K., Galen, B. R., & Paquette, J. A. (2002). Hopes rather than fears, admirations rather than hostilities: A response to Archer and Bjorqkvist. *Social Development*, *10*, 275–280. https://doi.org/10.1111/1467-9507.00165

Underwood, M. K., Galenand, B. R., & Paquette, J. A. (2004). Top ten challenges for understanding gender and aggression in children: Why can't we all just get along? *Social Development*, *10*, 248–266. https://doi.org/10.1111/1467-9507.00162

United Nations. (1995). *The Copenhagen declaration and programme of action, world summit for social development, 6-12 March*. New York: United Nations.

United States Census Bureau. (2017). Quick facts: Baltimore, Maryland. Retrieved August 12, 2018, from https://www.census.gov/quickfacts/fact/table/baltimorecitymaryland,US/PST045218

VadenBos, G. . (Ed.). (2007). *APA dictionary of psychology*. Washington, DC: American Psychological Association.

Vaillancourt, T., Miller, J. L., Fagbemi, J., Côté, S., & Tremblay, R. E. (2007). Trajectories and predictors of indirect aggression: Results from a nationally representative longitudinal study of Canadian children aged 2–10. *Aggressive Behavior*, *33*, 314–326. https://doi.org/10.1002/ab.20202

van der Sar, A. M., & Goudswaard, M. (2001). *Docenthandleiding taakspel voor basisonderwijs*. Rotterdam: Pedologisch Instituut.

van Lier, P. (2002). *Preventing disruptive behavior in the urban classroom*. University of Rotterdam, Rotterdam, The Netherlands.

van Lier, P. a C., Muthén, B. O., van der Sar, R. M., & Crijnen, A. a M. (2004). Preventing disruptive behavior in elementary schoolchildren: Impact of a universal classroom-based intervention. *Journal of Consulting and Clinical Psychology*, *72*, 467–478. https://doi.org/10.1037/0022-006X.72.3.467

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

Varadhan, R., Segal, J. B., Boyd, C. M., Wu, A. W., & Weiss, C. O. (2013). A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. *Journal of Clinical Epidemiology*, *66*, 818–825. https://doi.org/10.1016/j.jclinepi.2013.02.009

Veit-Wilson, J. H. (1987). Consensual approaches to poverty lines and social security. *Journal of Social Policy*, *16*, 183–211. https://doi.org/10.1017/s0047279400020377

Violato, M., Petrou, S., Gray, R., & Redshaw, M. (2011). Family income and child cognitive and behavioural development in the United Kingdom: Does money matter? *Health Economics*, *20*, 1201–1225. https://doi.org/10.1002/hec

Voils, C. I., Chang, Y. K., Crandell, J., Leeman, J., Sandelowski, M., & Maciejewski, M. L. (2012). Informing the dosing of interventions in randomized trials. *Contemporary Clinical Trials*, *33*, 1225–1230. https://doi.org/10.1016/j.cct.2012.07.011

von Stumm, S. (2017). Socioeconomic status amplifies the achievement gap throughout compulsory education independent of intelligence. *Intelligence*, *60*, 57–62. https://doi.org/10.1016/j.intell.2016.11.006

Votruba-Drzal, E. (2006). Economic disparities in middle childhood development: Does income matter? *Developmental Psychology*, *42*, 1154–1167. https://doi.org/10.1037/0012-1649.42.6.1154

Walberg, H. J., & Tsai, S.-L. (1983). "Matthew" effects in education. *American Educational Research Journal*, *20*, 359. https://doi.org/10.2307/1162605

Wang, M.-T., & Degol, J. L. (2016). School climate: a review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*, 315–352. https://doi.org/10.1007/s10648-015-9319-1

Wanless, S. B., & Domitrovich, C. E. (2015). Readiness to implement school-based social-emotional learning interventions: Using research on factors related to implementation to maximize quality. *Prevention Science*, *16*, 1037–1043. https://doi.org/10.1007/s11121-015-0612-5

Warren, E. J., & Font, S. A. (2015). Housing insecurity, maternal stress, and child maltreatment: An application of the family stress model. *Social Service Review*, *89*, 9–39. https://doi.org/10.1086/680043

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*, 70–77. https://doi.org/10.1002/mrdd

Weber, B., Jensen, L., Miller, K., Mosley, J., & Fisher, M. (2005). A critical review of rural poverty literature: Is there truly a rural effect? *International Regional Science Review*, *28*, 381–414. https://doi.org/10.1177/0160017605278996

Webster-Stratton, C., Jamila Reid, M., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: evaluation of the Incredible Years teacher and child training programs in high-risk schools. *Journal of Child Psychology and Psychiatry*, *49*, 471–488. https://doi.org/10.1111/j.1469-7610.2007.01861.x

Webster-Stratton, C., & Reid, M. J. (2003). Treating conduct problems and strengthening social and emotional competence in young children. *Journal of Emotional and Behavioral Disorders*, *11*, 130–143. https://doi.org/10.1177/10634266030110030101

Wehby, J. H., Maggin, D. M., Moore Partin, T. C., & Robertson, R. (2012). The impact of working alliance, social validity, and teacher burnout on implementation fidelity of the Good Behavior Game. *School Mental Health*, *4*, 22–33. https://doi.org/10.1007/s12310-011-9067-4

Weinstein, A., & Dannon, P. (2015). Is impulsivity a male trait rather than female trait? Exploring the sex difference in impulsivity. *Current Behavioral Neuroscience Reports*, *2*, 9–14. https://doi.org/10.1007/s40473-015-0031-8

Weiss, C. (2000). Theory-based evaluation: Theories of change for poverty reduction programs. In O. N. Feinstein & R. Picciotto (Eds.), *Evaluation and poverty reduction:*

proceedings from a world bank conference (pp. 102–134). Washington, DC: World Bank. Retrieved from https://books.google.co.uk/books?id=e8kiRvx7L-AC

Weissberg, R. P., Kumpfer, K. L., & Seligman, M. E. P. (2003). Prevention that works for children and youth: An introduction. *American Psychologist*, *58*, 425–432. https://doi.org/10.1037/0003-066X.58.6-7.425

Wells, M., Williams, B., Treweek, S., Coyle, J., & Taylor, J. (2012). Intervention description is not enough: Evidence from an in-depth multiple case study on the untold role and impact of context in randomised controlled trials of seven complex interventions. *Trials*, *13*, 1–17. https://doi.org/10.1186/1745-6215-13-95

Welsh, J. A., Beirman, K., & Mathis, E. (2014). Parenting programs that promote school readiness. In K. Bolvin & K. Bierman (Eds.), *Promoting school readiness and early learning: Implications of developmental research for practice* (pp. 253–278). New York: The Guilford Press.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology*, *102*, 43–53. https://doi.org/10.1037/a0016738

WELSH, W. N., GREENE, J. R., & JENKINS, P. H. (2006). School disorder: The influence of individual, institutional, and community factors. *Criminology*, *37*, 73–116. https://doi.org/10.1111/j.1745-9125.1999.tb00480.x

Werthamer-Larsson, L., Kellam, S., & Wheeler, L. (1991). Effect of first-grade classroom environment on shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology*, *19*, 585–602. https://doi.org/10.1007/BF00937993

Whear, R., Thompson-Coon, J., Boddy, K., Ford, T., Racey, D., & Stein, K. (2012). The effect of teacher-led interventions on social and emotional behaviour in primary school children: a systematic review. *British Educational Research Journal*, 1–38. https://doi.org/10.1080/01411926.2011.650680

Whelan, C. T., Layte, R., & Maôtre, B. (2004). Understanding the mismatch between income poverty and deprivation: A dynamic comparative analysis. *European Sociological Review*, *20*, 287–302. https://doi.org/10.1093/esr/jch029

Wheldall, K., & Merrett, F. (1988). Which classroom behaviours do primary school teachers say they find most troublesome? *Educational Review*, *40*, 13–27. https://doi.org/10.1080/0013191880400102

Wickham, S., Whitehead, M., Taylor-Robinson, D., & Barr, B. (2017). The effect of a transition into poverty on child and maternal mental health: A longitudinal analysis of the UK Millennium Cohort Study. *The Lancet Public Health*, *2*, 141–148. https://doi.org/10.1016/S2468-2667(17)30011-7

Wigelsworth, M., Lendrum, a., Oldfield, J., Scott, a., ten Bokkel, I., Tate, K., & Emery, C. (2016). The impact of trial stage, developer involvement and international transferability on universal social and emotional learning programme outcomes: a meta-analysis. *Cambridge Journal of Education*, *46*, 347–376. https://doi.org/10.1080/0305764X.2016.1195791

Wigelsworth, M., Humphrey, N., & Lendrum, A. (2013). Evaluation of a school-wide preventive intervention for adolescents: the secondary Social and Emotional Aspects

of Learning (SEAL) Programme. *School of Mental Health*, *52*, 96–109. https://doi.org/10.1007/s12310-012-9085-x

Wilcox, H. C., Kellam, S. G., Brown, C. H., Poduska, J., Ialongo, S., Wang, W., & Anthony, J. C. (2009). The impact of two universal randomized first- and second-grade classroom interventions on young adult suicide ideation and attempt. *Alcohol*, *95*, 1–27. https://doi.org/10.1016/j.drugalcdep.2008.01.005.

Wilkinson, R. G. (1992). Income distribution and life expectancy. *BMJ*, *304*, 165–168. https://doi.org/10.1136/bmj.304.6820.165

Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior. update of a meta-analysis. *American Journal of Preventive Medicine*, *33*(1), 130–143. https://doi.org/10.1016/j.amepre.2007.04.011

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*, 203–214. https://doi.org/10.1901/jaba.1978.11-203

Woolf, S. H. (2008). The meaning of translational research and why it matters. *Jama*, *299*, 211–213. https://doi.org/10.1001/jama.2007.26

Yeung, W. J., Linver, M. R., & Brooks–Gunn, J. (2002). how money matters for young children's development: Parental investment and family processes. *Child Development*, *73*, 1861–1879. https://doi.org/10.1111/1467-8624.t01-1-00511

Zachrisson, H. D., & Dearing, E. (2015). Family income dynamics, early childhood education and care, and early child behavior problems in Norway. *Child Development*, *86*, 425–440. https://doi.org/10.1111/cdev.12306

Zhueng, B. (1997). *Statistical inferences for poverty measures with realitive poverty rates* (LIS Working Paper No. 187). Luxembourg. Retrieved from http://hdl.handle.net/10419/160839%0AStandard-Nutzungsbedingungen

Appendix A

1. Observation Record (A)

A1 Date (dd/mm/yy)	
A2 School code	
A3 Teacher code	
A4 Start time (hh:mm)	
A5 End time (hh:mm)	
A6 Observer name	
A7 Location	Classroom/PE/playground etc

2. Classroom

B. Reach			
B1 Number of children			
B2 Number of absences and withdrawals (note reasons)			
B3 Number of adults present other than teacher (e.g. teaching assistants). Note their roles (if known) in the space below			
B4 Notes			
C. Physical Artefacts			
C1 Rules poster	PA Present & accessible	P Present	No
C2 Voice levels poster			
C3 Team assignment wall-chart this includes space fornames of pupils in teams and recording weekly winners			

C4 Scoreboard poster for recording infractions			
Rules on desk			
C5 GBG booklets record in notes if these are the standard version or if they have been adapted			
C6 Timer			
C7 Stamps/stickers for booklets			
C8 Reinforcers			
C9 Notes "Present & accessible" refers to the artefact being in the room and visible to the pupils regarding posters; and within easy reach of the pupils regarding the desk artefacts. "Present" refers to the artefacts being in the room but not easily visible to all.			

3.   Pre-Game

D1 Lesson	
D2 Task/activity	
D3 Independent/group/pair work	
D4 Voice level	
D5 Game length (mins)	
D6 N pupils per team (range)	
D7 N teams in the class	

E. Descriptor	**Procedural Fidelity**		**Quality**	
Activity				
E1Teacher explains the task/activity	Y	N	1 Unlcear instructions	2 clear and detailed explanation
E2 Teacher checks understanding of the task/activity	Y	N	1 briefly checks understanding/asks if children understand	2 asks children to explain instructions/give examples
E3 Teacher reminds pupils that they cannot ask for help	Y	N	1 teacher reminds pupils	2 teacher reminds pupils and provides example

				of alternative strategies
Teams				
E4 Pupils are in teams of between 3 and 7 (except for special circumstances, e.g. team of 1) add any observations to notes	Y	N		
E5 Pupils are in clear teams	Y	N	1 Some/not always distnict	2 All distinct
E6 Teams are gender balanced if class is unbalanced but teams are balanced as much as possible record in notes	Y	N	1 Some – at least 1 team unbalanced	2 all
Rules				
E7 Rules verbally reviewed/discussed with class	Y	N	1 Rote repitition of rules	2 *discussion* of how the applies to the activity – *always* elicits input from pupils
E8 Exemplars modelled/discussed by teacher and/or students	Y	N	1 rote statement /inappropriate examples	2 acting out or further description/detailing of appropriate examples.
E9 Infractions modelled/discusessed by teacher	Y	N	1 rote statement of examples	2 acting out or further description/detailing of examples
E10 Infractions only described by students Students should not be acting these out	Y	N	1 rote statement of examples	2 further descriptions/ explanations of examples
E11 Voice level given by teacher	Y	N	1 voice level stated	2 voice level

238

				discussed/explained
Game Specifics				
E12 Teacher states when the game begins	Y	N		
E13 Teacher states how long the game will be played for	Y	N		
E14 Teachers sets timer	Y	N		
E15 Teacher states that they will monitor infractions	Y	N		
E16 Teacher states that 4 infractions are permitted	Y	N		
E17 Teacher reminds pupils that they are competing against each other	Y	N		

- Note any adaptations (additions or omissions) made, whether surface/low risk or deep/high risk. Describe them and then interpret them later (e.g. if the teacher misses something out).
- Surface level adaptations that enhance quality, engagement etc. tend to be thought of as positive (additions or enhancements); however, if the addition results in the omission of a core activity / element of the lesson, then it may potentially be seen as a high risk adaptation. If an adaptation meant that something else was missed out then the adherence rating (1b) should also be reduced. Add in notes of any supplementary activities used within the main lesson.

F. Adaptations	**Notes**	**Interpretation**		
F1 Teams				
F2 Rules				
F3 Activity				
F4 Game Specifics				
G. Participant Responsiveness mark "n/a" if behaviours are not observed/previously scored "0"				
G1 How do pupils respond to the	N/A	0 most pupils react	1 some pupils react	2 most pupils react

					4. During Game
announcement of the game?		negatively	negatively/some positively/general indifference/no reaction (clarify in notes).	positively	
G2 How attentive are pupils to the teacher's instructions and examples regarding the game?	N/A	0 pupils clearly not listening/engaging with teacher	1 some pupils engaging/all pupils not fully focused (clarify in notes)	2 Most are engaged	
G3 How enthusiastic/willing to participate are pupils when discussing the rules?	N/A	0 no enthusiasm/ nobody volunteers an answer	1 some children are enthusiastic	2 most participate/ display enthusiasm	
G4 Notes			**G4 interpretation**		

H. Descriptor	Procedural Fidelity		Quality	
Check, Comment, Redirect				
H1 Teacher records majority of infractions on scoreboard	Y	N	1 records most infractions explain any justifications in notes	2 records all infractions
H2 Teacher identifies majority rule broken	Y	N	1 identifies most infractions	2 always identifies rule.
H3 Teacher **discreetly** indicates rules broken to specific pupil most of the time	Y	N	1 Draws attention to specific child - mostly is discreet when doing so	2 always indicates rule and does so discreetly
H4 Teacher frequently identifies rule breaking team Only focus on identification of	Y	N	1 mostly identifies team	2 always identifies team

teams, not individual pupils				
H5 Rest of team frequently praised for adhering to rules	Y	N	1 mostly praises other members/generic praise for behaving	2 always praises rest of team for specific rule
H6 Other teams frequently praised for adhering to rules	Y	N	1 mostly praises other teams/generic praise for behaving	2 always praises teams for specific rule
H7 Teacher does not punish pupils/teams for infractions	Y	N		
Game Management				
H8 Teacher monitors behaviour	Y	N	1 sometimes monitors behaviour	2 always monitors behaviour
H9 Teacher does not interact with pupils Teachers should not interact with pupils during the game	Y	N		
H10 Teacher adheres to time limit	Y	N		
H11 Teacher announces the end of the game	Y	N		

H12 Notes		**Interpretation**	

I. Adaptations	**Notes**		**Interpretation**

I1 Check, Comment, Redirect					
I2 Game Management					

Team	1	2	3	4	5
Rule 1 We will work quietly					
Rule 2 We will be polite to others					
Rule 3 We will get out of our seats with permission					
Rule 4 We will follow directions					
Total Infractions					

J. Participant Responsiveness   mark "n/a" if behaviours are not observed/previously scored "0"

J1 Generally, do rule breaking pupils correct their behaviour following an	N/A	0 pupils continue to repeatedly break	1 some pupils do break the rule	2 pupils do not break the rule again

infraction? Do not focus on behaviour of 1 child		the rules	again	
J2 Generally, how well do pupils respond to a member of their team getting a check?	N/A	0 pupils always display a negative response	1 some negative responses/mostly ignore the check	2 always ignore the check/supportive towards pupil if voice level allows
J3 Are there any externalising responses from pupils when they receive a check?	Note any externalising responses and potential reasons (e.g. SEN)			
J4 Notes	**Interpretation**			

5.   **Post-Game**

K. Descriptor	Procedural Fidelity		Quality	
Game management				
K1 Teacher repeats 4 checks or less criterion	Y	N		
K2 Teacher announces winning teams **only**	Y	N	1 generic acknowledgement of winners with praise	2 announces winning teams individually with praise
Reinforcers				
K3 Members of winning team receive stamp	Y	N	1	2

				most members team receive stamp	all members receive stamp
(or marker etc) even if they broke a rule they should still receive a stamp				most members team receive stamp	all members receive stamp
K4 Star placed on wall-chart	Y	N	1 teacher places marks	2 only team leaders place marks	

K5 Notes	Interpretation

L. Adaptations	Notes	Interpretation
L1 Game management		
L2 Reinforcers		

M. Participant Responsiveness   mark "n/a" if behaviours are not observed/previously scored "0"

M1 Generally, what is the level of interest and attentiveness to the reinforcers?	N/A	0 no interest	1 some are interested	2 all interested /engaged by the receipt of a reward
M2 Generally, how do team leaders respond to being called up and sticking items on the board?	N/A	0 no interest	1 most are interested	2 all interested /engaged by the receipt of a reward
M3 Generally, how do pupils respond if they do not win the game?	N/A	0 Mostly negative responses	1 Some negative responses	2 No negative responses

M4 Notes		Interpretation	

N. Reinforcers			
N1 Type	Tangible	Intangible	Token
N2 Prize(s) given			
N3 Delay of gratification	Immediate	Delayed make any notes on length of delay	
N4 Notes Reinforcers should be additional to stamps in booklets		**Intepretation**	

6. Overall (O)

O1 Interest and enthusiasm	0 not present	1 some interest/enthusiasm	2 very enthusiastic
O2 Clarity of expression	0 always unclear	1 sometimes unclear	2 always clear
O3 Preparedness	0 teacher unprepared	1 teacher has most items prepared	2 teacher prepared and has materials to hand
O4 Consistency of behaviour Parameters can be different for different children e.g. SEN, but need to be maintained consistently	0 is inconsistent within the parameters set	1 sometimes consistent	2 always consistent once parameters are set. *Record any differences in adaptations (e.g. if due to SEN)*
O5 Engagement of pupils	0 teacher makes no effort to	1 Teacher uses some positive and/or	2 teacher always engages pupils using positive

	engage/uses negative strategies to engage	negative strategies	strategies
O6 Notes		**Interpretation**	

The Good Behaviour Game

INFORMATION SHEET FOR PARENTS

Your child's school is involved in a project about the Good Behaviour Game. The Good Behaviour Game is a way to help children to concentrate on their school work and improve their behaviour. It has been shown to be very helpful in other countries in the world. We want to find out if it can help children in England too. The project is funded by The Education Endowment Foundation and The National Institute for Health Research.

We are writing to you because your child's school is involved in the project. We will ask your child's teacher to complete a survey about your child's behaviour once a year starting summer (May-July) 2015. From the summer of 2017 onwards we will also ask your child to complete a brief annual survey about their wellbeing (see below for more details). Our surveys will conclude in summer 2019.

Please take time to read the following information carefully and decide whether or not your child would like to take part. If you would like any more information or have any questions about the research project, please telephone Dr. Alexandra Barlow on 0161 275 3504 or email her at alexandra.barlow@manchester.ac.uk.

Who will conduct the research?
The research will be conducted by Professor Neil Humphrey and his research team at the Manchester Institute of Education, The University of Manchester, Oxford Road, Manchester M13 9PL.

Title of the research
"The Good Behaviour Game"

What is the aim of the research?
Our main aim is to examine the impact of the Good Behaviour game on reading and behaviour.

Where will the research be conducted?
Primary schools in Greater Manchester, West Yorkshire, South Yorkshire and East Midlands.

What is the duration of the research?
The project itself runs from September 2014 until March 2020. The schools that implement the Good Behaviour Game (see below) will do so from September 2015 to July 2017.

Why have I been chosen?
We are writing to you because your child's school is taking part in the Good Behaviour Game project. Schools will be randomly chosen to (a) implement the Good Behaviour Game over a two-year period (Good Behaviour Game schools), or (b) continue as normal (comparison schools). We will be collecting data in both Good Behaviour Game and

comparison schools. After two years, all schools will be free to decide whether they wish to start/continue using the Good Behaviour Game.

What would my child be asked to do if he/she took part?
Your child's class teacher will be asked to complete a brief online survey about your child's behaviour. These surveys will be completed annually – in May/-July 2015, 2016, 2017, 2018 and 2019.

Your child will be asked to complete both a short reading assessment and a short survey about wellbeing at the end of the main trial in summer (May-July) 2017, and again in May-July 2018 and 2019. The survey will take approximately 20 minutes to complete and the reading assessment will take approximately 30 minutes to complete.

If you agree, you will be saying that your child can take the tests and fill in the questionnaires. You will also be saying that his/her teacher can complete surveys about him/her.

What happens to the data collected?
The data will be downloaded from our secure online survey site so that it can be analysed by our research team at the University of Manchester. We will write a report based on our analyses for our funders, the Education Endowment Foundation and the National Institute for Health Research. It is also likely that we will write articles for academic journals based on what we find out in the project. The data may also be used as part of a doctoral thesis. Finally, it is possible that we will write a book about the research. Your child's name will not be used in any of the reports that we write.

How is confidentiality maintained?
All data provided will be treated as confidential and will be completely anonymous. Identifying information (e.g. your child's name) will only be used in order to match responses about the same individual from different respondents (e.g. teacher and pupil surveys) and across different times (e.g. May-July 2015, 2016, and 2017). After this matching process is complete, all identifying information will be destroyed.

The website that houses these surveys will be completely secure and password protected. All survey data will be stored on a secure, password protected computer to which only senior members of the research team have access.

What happens if I do not want my child to take part or I change my mind later?
It is up to you if you want your child to take part in the data collection.
If you decide your child and his/her teacher can take part in the data collection you do not need to do anything – your child's school will be sent further details about when and how to complete the survey in the near future.

If you decide not to take part then you need to either complete the opt-out consent form enclosed and return it to our research team or contact Dr. Alexandra Barlow by telephone or email (details below).

If you decide to take part and then change your mind, you are free to withdraw without needing to give a reason by contacting Dr. Alexandra Barlow by telephone or email (details below). We will send annual reminders about the study, but you can opt your child out at any time up until the end of the study, in summer 2017. If you do this please rest assured that we will destroy any data collected about your child as part of the study.

Will I be paid for participating in the research?
We are not able to offer any payment or incentive for participating in this study.

Disclosure and Barring Service (DBS) Check

248

Every member of our research team has undergone a Disclosure and Barring Service (formerly 'Criminal Records Bureau') check at the Enhanced Disclosure level. This means that they have permission to work with and do research with children.

Contact for further information
Dr. Alexandra Barlow
Educational Support and Inclusion
School of Education
University of Manchester
Oxford Road
Manchester
M13 9PL

Tel: 0161 275 3504
Email: alexandra.barlow@manchester.ac.uk

Also, please see our website for further details about the Good Behaviour Game and background, the project design and project team. The website can be found at: http://www.goodbehaviourgame.info

What if something goes wrong?
If your child or your child's teacher completing the survey makes you worry about your child's wellbeing then you should contact the school in the first instance and ask to speak to his/her teacher.
You can also get independent support and advice from a charity called Young Minds. Their parent helpline number is 0808 802 5544.

What if I want to complain?
If you have any concerns or wish to complain, you should contact the researcher Alexandra Barlow in the first instance (contact details above).

If you remain dissatisfied, or if the research team is unable to address the issues you raise you should contact the Head of School, Prof Tim Allott (School of Environment, Education and Development), at Tim.Allott@manchester.ac.uk or on 0161 275 3662.

If there are any issues regarding this research that you would prefer not to discuss with members of the research team or Head of School, please contact the Research Practice and Governance Co-ordinator by either writing to 'The Research Practice and Governance Co-ordinator, Research Office, Christie Building, The University of Manchester, Oxford Road, Manchester M13 9PL', by emailing: Research-Governance@manchester.ac.uk, or by telephoning 0161 275 7583 or 275 8093

The Good Behaviour Game

CONSENT FORM FOR PARENTS

An information sheet is attached to this form. Please read it carefully before making a decision about taking part.

If you are willing to let your child take part and for his/her teacher to give information about him/her then you do not need to do anything at the moment.

If you decide not to let your child take part, then you need to complete the opt-out consent form below and use the freepost code below to return it to us:

FREEPOST RLYU-KAAB-AXRC
Dr. Alexandra Barlow,
Manchester Institute of Education
The University of Manchester,
Ellen Wilkinson Building
Oxford Road,
Manchester,
M13 9PL.

Alternatively, Dr. Barlow can be contacted by telephone on 0161 275 3504 or email at alexandra.barlow@manchester.ac.uk. If you do not want your child to participate please let us know by Friday 2nd October 2015.
Finally, please also remember that if you do decide he/she can take part, you are free to change your mind at any point in the study.

I do not wish my child to participate in the Good Behaviour Game project. My details are as follows:

My name	
My child's name	
Name of my child's school	

Signed: _____   Date: _____

The Good Behaviour Game

INFORMATION SHEET FOR TEACHERS

Your school is involved in an exciting project about the Good Behaviour Game. The Good Behaviour Game is an approach to classroom management designed to improve children's pro-social behaviour and concentration, while reducing disruptive behaviour. The GBG has an extensive international evidence base, but we need to know if it is effective in English primary schools. The project is funded by The Education Endowment Foundation and The National Institute for Health Research.

We are writing to you to explain what your role as a teacher would be in the research project if you agree to participate. Teachers are asked to complete a survey on behalf of the pupils in their class once a year starting summer (May-July) 2015. Specific information about this will be sent nearer the time.Please take time to read the following information carefully and decide whether or not you would like to take part.

If you would like any more information or have any questions about the research project, please telephone Dr. Alexandra Barlow on 0161 275 3504 or email her at alexandra.barlow@manchester.ac.uk.

Who will conduct the research?
The research will be conducted by Professor Neil Humphrey and his research team at the Manchester Institute of Education, The University of Manchester, Oxford Road, Manchester M13 9PL.

Title of the research
"The Good Behaviour Game".

What is the aim of the research?
Our main aim is to examine the impact of the Good Behaviour game on reading and behaviour.

Where will the research be conducted?
Primary schools in Greater Manchester, West Yorkshire, and South Yorkshire.

What is the duration of the research?
The project itself runs from September 2014 until March 2020. The schools that implement the Good Behaviour Game (see below) will do so from September 2015 to July 2017.

Why have I been chosen?
We are writing to you because your school is taking part in the Good Behaviour Game Project. Schools will be randomly chosen to (a) implement the Good Behaviour Game over a two-year period (Good Behaviour Game schools), or (b) continue as normal (comparison schools). We will be collecting data in both the Good Behaviour Game and comparison schools. After two years, all schools will be free to decide whether they wish to start/continue using the Good Behaviour Game.

What would I be asked to do?
All participating schools (i.e. both GBG and comparison schools)

Teachers (and/or support staff, such as teaching assistants) of participating classes will be asked to complete a short online survey about each pupil in their class that focuses on their behaviour. These
surveys will be completed annually between May-July in each year from 2015 to 2019. Each survey should take approximately 3-5 minutes to complete for each child.
In addition to the above, teachers of participating classes will also complete a short survey about themselves which will cover behaviour management practices and well-being. This survey will be completed annually – in May-July 2015-2017– and will take approximately 15 minutes to complete.

GBG schools only
Finally, in our research visits, we will observe teachers of participating classes delivering the Good Behaviour Game once each school year. Teachers will also complete a brief implementation survey. This survey will be completed annually – in February-April 2015-2017– and will take approximately 15 minutes to complete.

What happens to the data collected?
The online data will be downloaded to a secure server and analysed by our research team at the University of Manchester. We will write a report based on our analyses for our funders, the Education Endowment Foundation and the National Institute for Health Research. It is also likely that we will write articles for academic journals based on what we find out in the project. The data may also be used as part of a doctoral theses. Finally, it is possible that we will write a book about the research. Your name will not be used in any of the reports that we write.

How is confidentiality maintained?
All data provided will be treated as confidential and will be completely anonymous. Identifying information (e.g. names) will only be used in order to match responses about the same individual from different respondents (e.g. teacher and pupil surveys) and across different times (e.g. May-July 2015, 2016, and 2017). After this matching process is complete, all identifying information will be destroyed.
The website that houses these surveys will be completely secure and password protected. All survey data will be stored on a secure, password protected computer to which only senior members of the research team have access.

What happens if I do not want to take part or I change my mind?
It is up to you if you want to take part. If you decide to take part you do not need to do anything – you will be sent further details about when and how to complete the survey in the near future. If you decide not to take part then you need to either complete the opt-out consent form enclosed and return it to our research team at the address above or contact Dr. Alexandra Barlow by telephone or email (details above).

If you decide to take part and then change your mind, you are free to withdraw without needing to give a reason by contacting Dr. Alexandra Barlow by telephone or email (details below). We will send annual reminders about the study, but you can opt out at any time up until the end of the study, in summer 2017. If you do this please rest assured that we will destroy any data collected about your child as part of the study.

Will I be paid for participating in the research?
We are not able to offer any payment or incentive for participating in this study.

Disclosure and Barring Service (DBS) Check
Every member of our research team has undergone a Disclosure and Barring Service (formerly 'Criminal Records Bureau') check at the Enhanced Disclosure level. This means that they have permission to work with and do research with children.

Contact for further information
Dr. Alexandra Barlow
Educational Support and Inclusion
School of Education

University of Manchester
Oxford Road
Manchester
M13 9PL

Tel: 0161 275 3504
Email: alexandra.barlow@manchester.ac.uk

Also, please see our website for further details about the Good Behaviour Game and background, the project design and project team.
The website can be found at: http://www.goodbehaviourgame.info

What if I want to complain?
If you have any concerns or wish to complain, you should contact the researcher Alexandra Barlow in the first instance (contact details above).

If you remain dissatisfied, or if the research team is unable to address the issues you raise you should contact the Head of School, Prof Tim Allott (School of Environment, Education and Development), at Tim.Allott@manchester.ac.uk or on 0161 275 3662.

If there are any issues regarding this research that you would prefer not to discuss with members of the research team or Head of School, please contact the Research Practice and Governance Co-ordinator by either writing to 'The Research Practice and Governance Co-ordinator, Research Office, Christie Building, The University of Manchester, Oxford Road, Manchester M13 9PL', by emailing: Research-Governance@manchester.ac.uk, or by telephoning 0161 275 7583 or 275 8093

The Good Behaviour Game

CONSENT FORM FOR TEACHERS

An information sheet is attached to this form. Please read it carefully before making a decision about taking part.

Please complete the slip below to indicate if you do not wish to participate in the research strand of the Good Behaviour Game project, which involves the completion of teacher-pupil surveys and if allocated to the Good Behaviour Game strand, an observation of a Good Behaviour lesson and annual teacher surveys.

Finally, please also remember that if you do decide to take part, you are free to change your mind at any point in the study.

I do not to participate in the Good Behaviour Game project. My details are as follows:

My name	
School name	

Signed: _____   Date: _____

Return this slip to:
FREEPOST RLYU-KAAB-AXRC
Dr. Alexandra Barlow,
Manchester Institute of Education
The University of Manchester,
Ellen Wilkinson Building
Oxford Road,
Manchester,
M13 9PL

Appendix D
Data Screening for Research Question One

D.1
Scatterplot showing the relationship between the standardised residual against the residual predicted value for the outcome variable, DB T2.
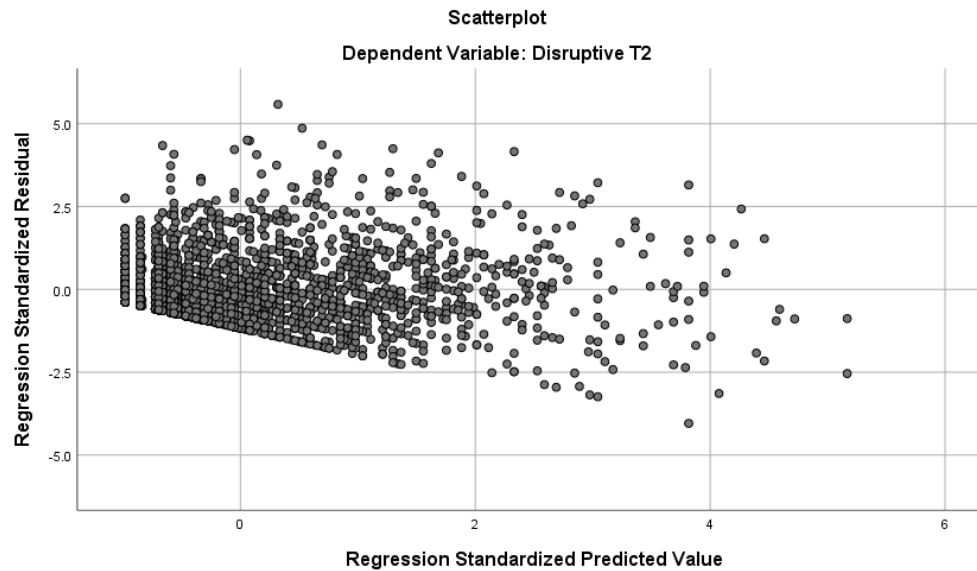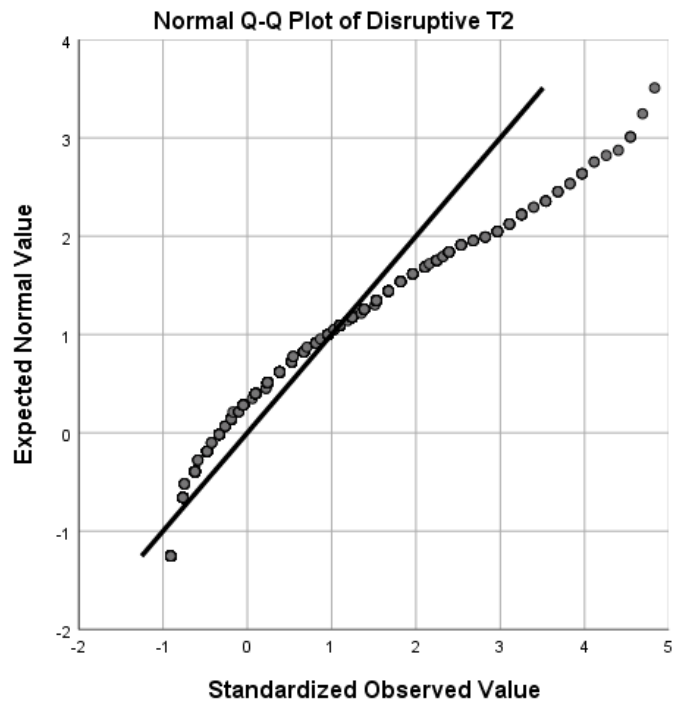

Scatterplot
Dependent Variable: Disruptive T2

D.2
Scatterplot showing the relationship between the standardised residual against the residual predicted value for the outcome variable, DB T3.
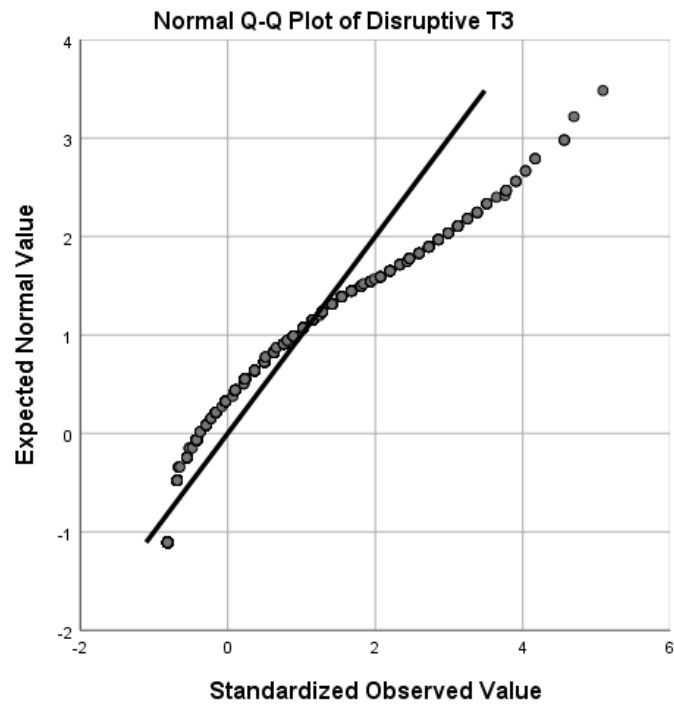

Scatterplot
Dependent variable: Disruptive T3
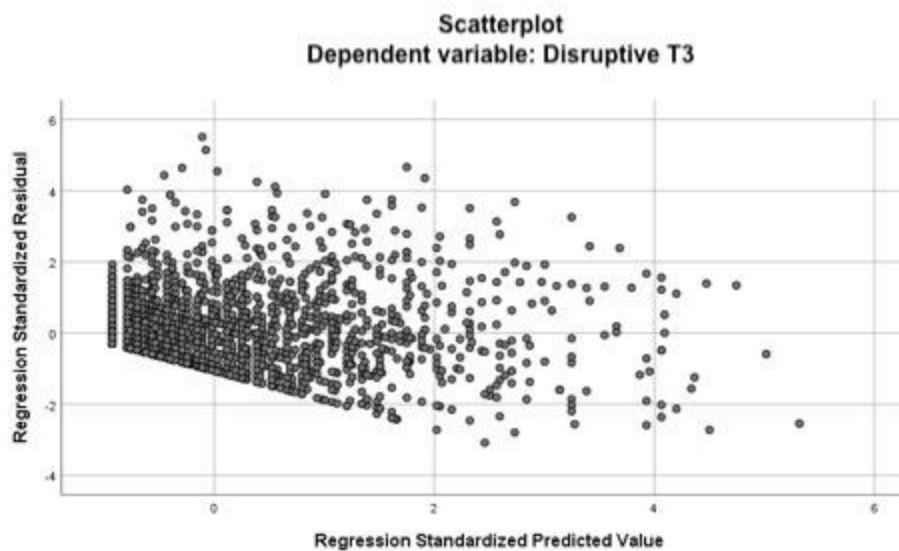
D.3

Q-Q plot for DB T2 residuals to establish normal distribution.

D.4

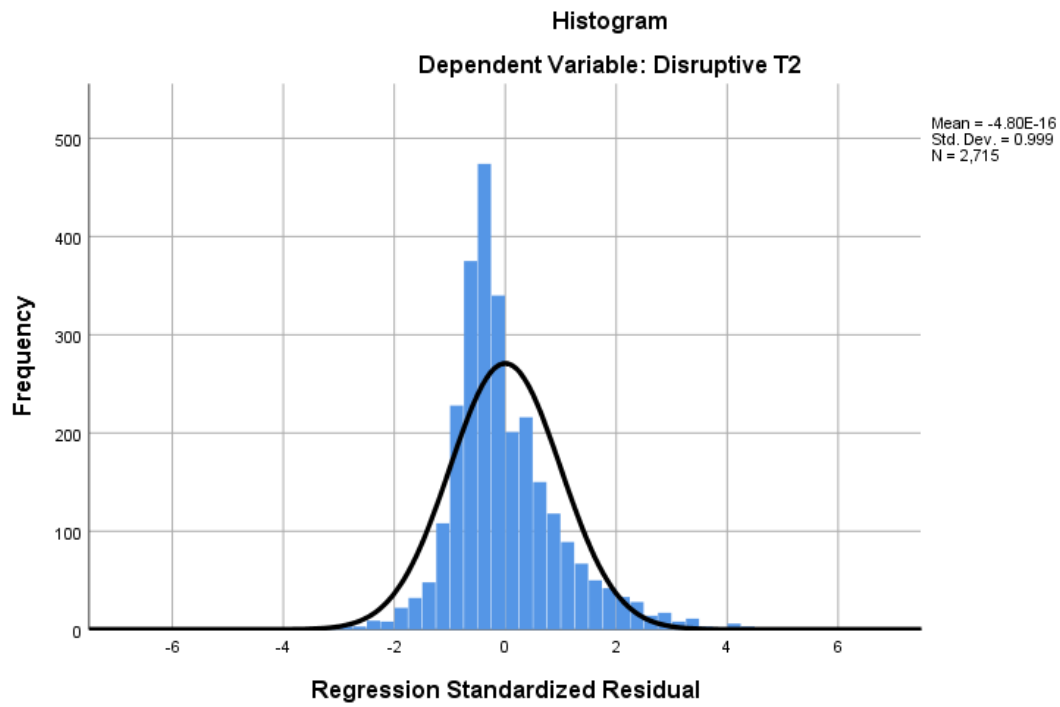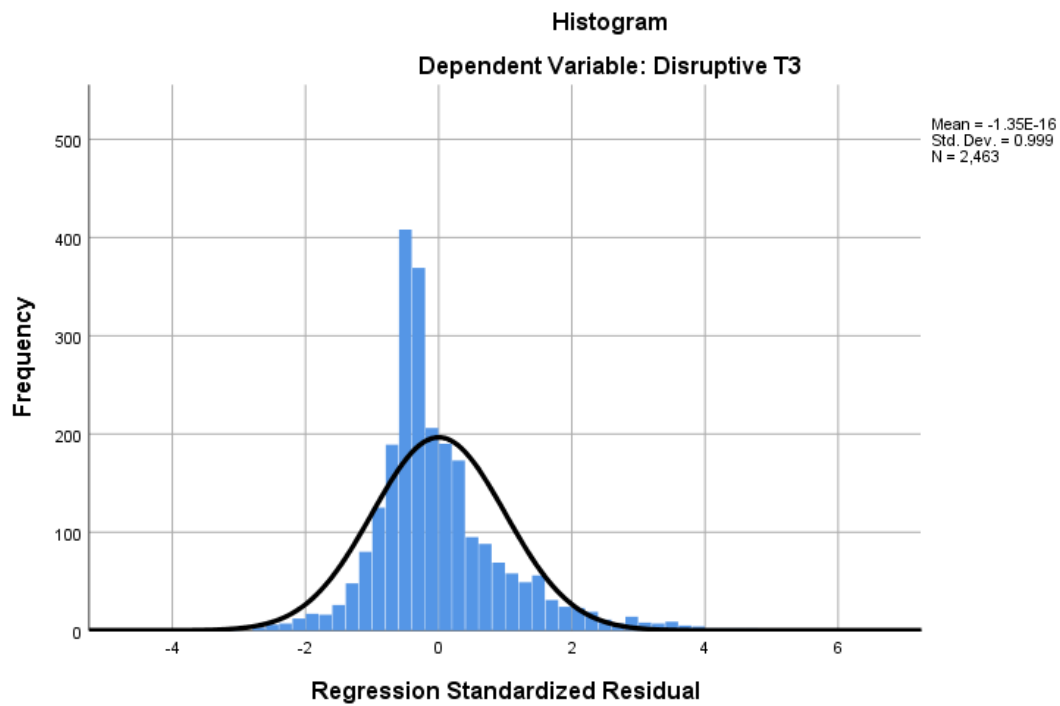*Q-Q plot for DB T3 residuals to establish normal distribution.*

D.5
Histogram for DB T2 residuals to establish normal distribution.

Histogram
Dependent Variable: Disruptive T2

Mean = -4.80E-16
Std. Dev. = 0.999
N = 2,715

D.6
Histogram for DB T3 residuals to establish normal distribution.

Histogram
Dependent Variable: Disruptive T3

Mean = -1.35E-16
Std. Dev. = 0.999
N = 2,463

D.7
Histogram depicting the cut off point to determine school-level poverty

% of pupils known to be eligible for free school meals (Performance Tables)
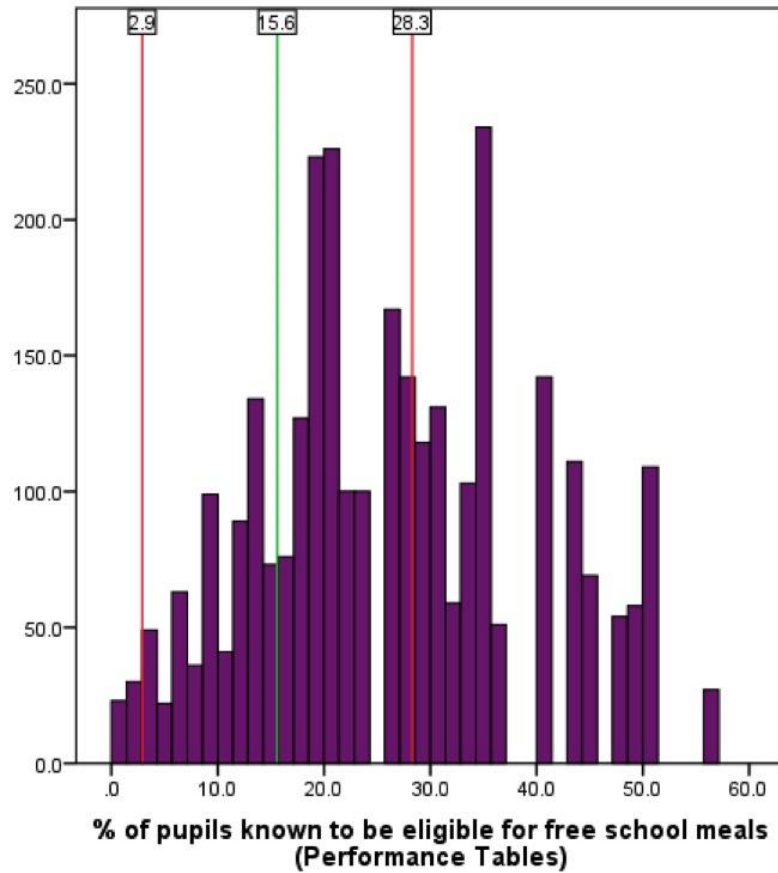
The green line represents the national average of FSM for the first year of the study (DfE, 2016). The red cut off point to the right shows 1SD above the national average and was used for the cut-off point to group schools into poverty categories. The red line to the left is 1SD below the national average.

Appendix E
Preliminary Analyses for Research Question One

E.1
ANOVA plot showing the DB T2 means for gender by trial group
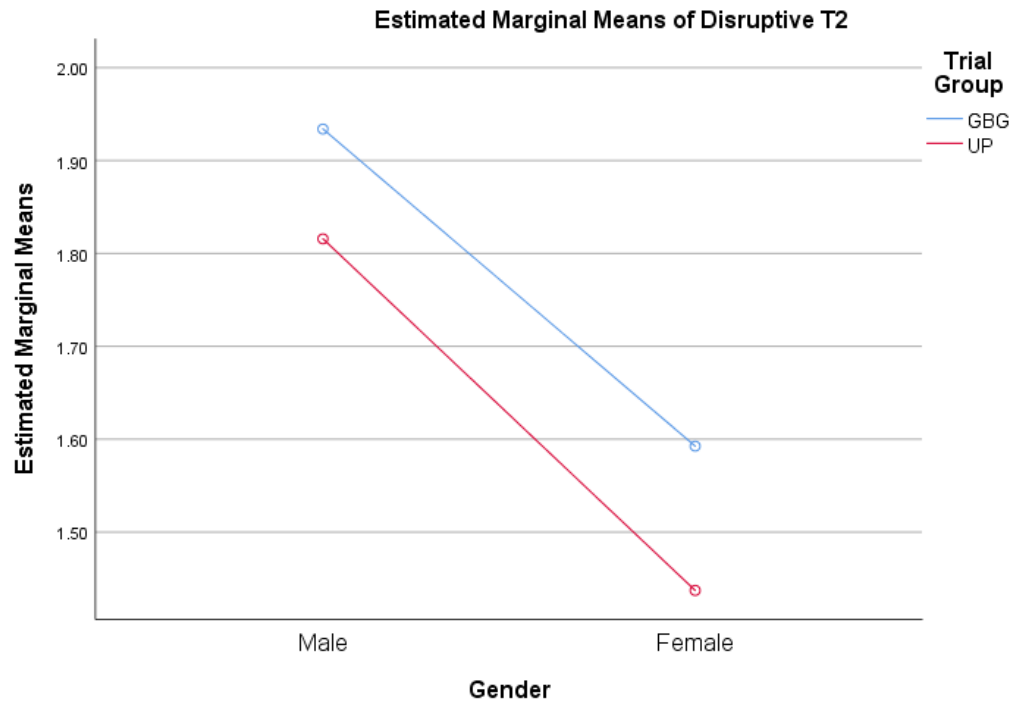
E.2
ANOVA plot showing the DB T3 means for gender by trial group.

Estimated marginal means for DBT3

E.3

ANOVA plot showing the DB T2 means for FSM eligibility by trial group.

Estimated Marginal Means of Disruptive T2
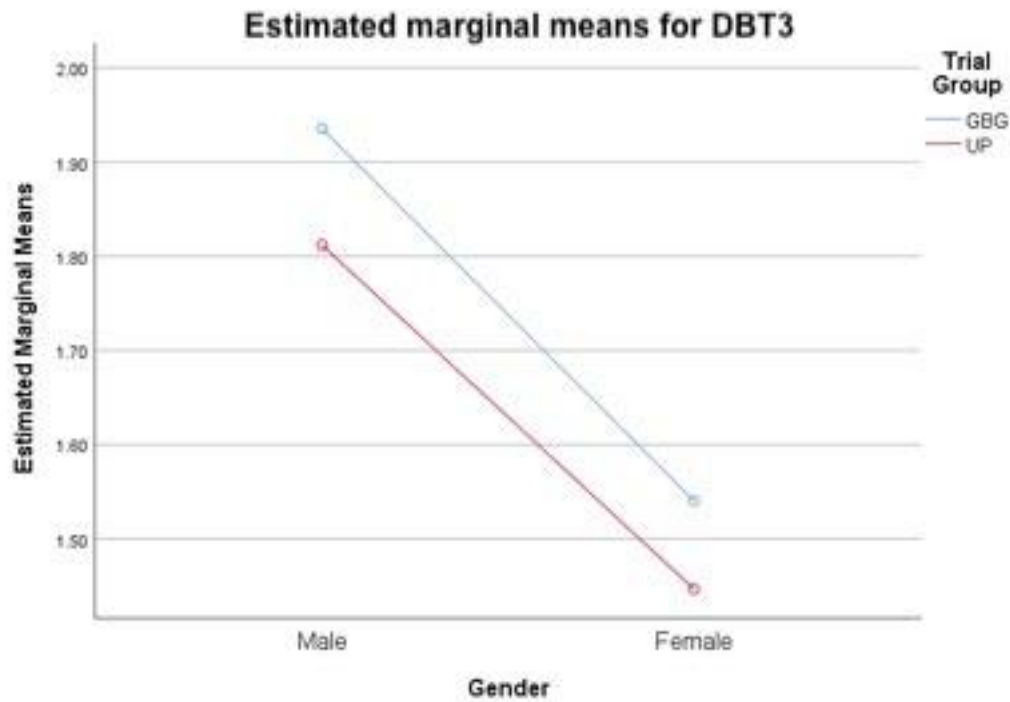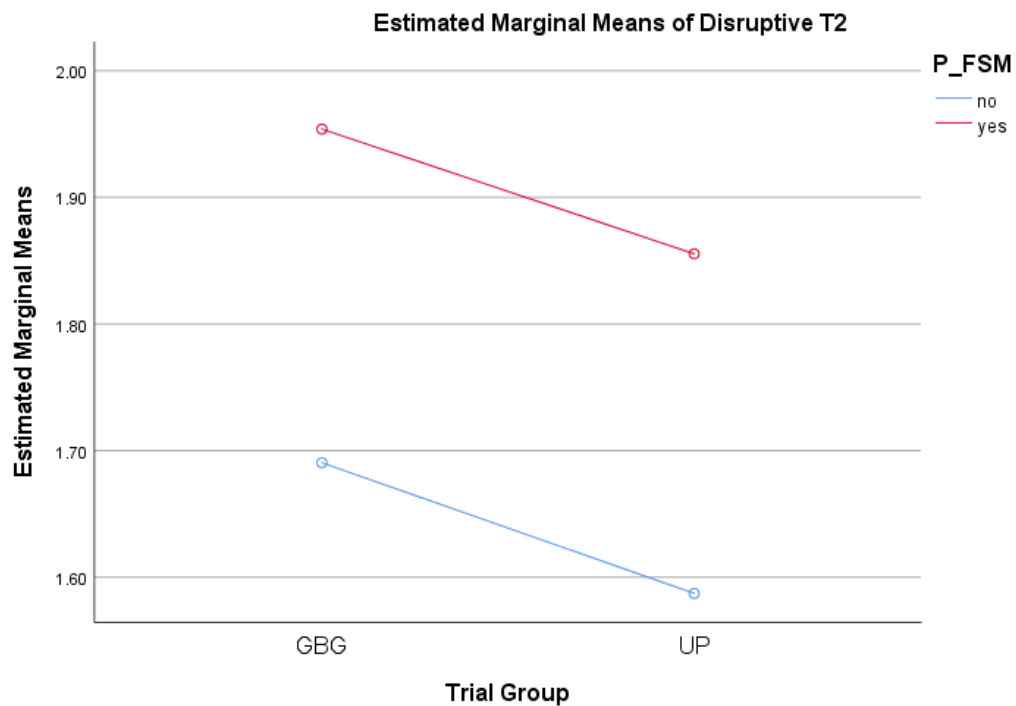
E.4

ANOVA plot showing the DB T2 means for FSM eligibility by trial group.

Estimated marginal means for DBT3
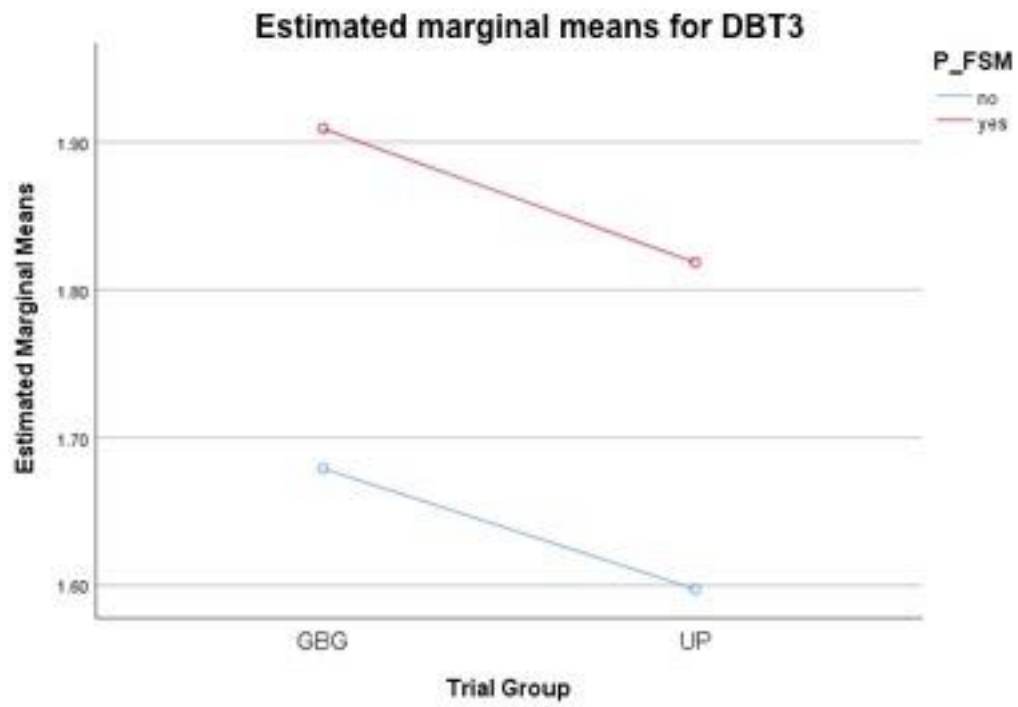
Appendix F
Data Suitability for Hierarchical Cluster Analysis

F1
Box plot for teachers' average minutes played per week.

F2
Box plot for teachers' dosage.

F3
Box plot for teachers' fidelity and quality scores.
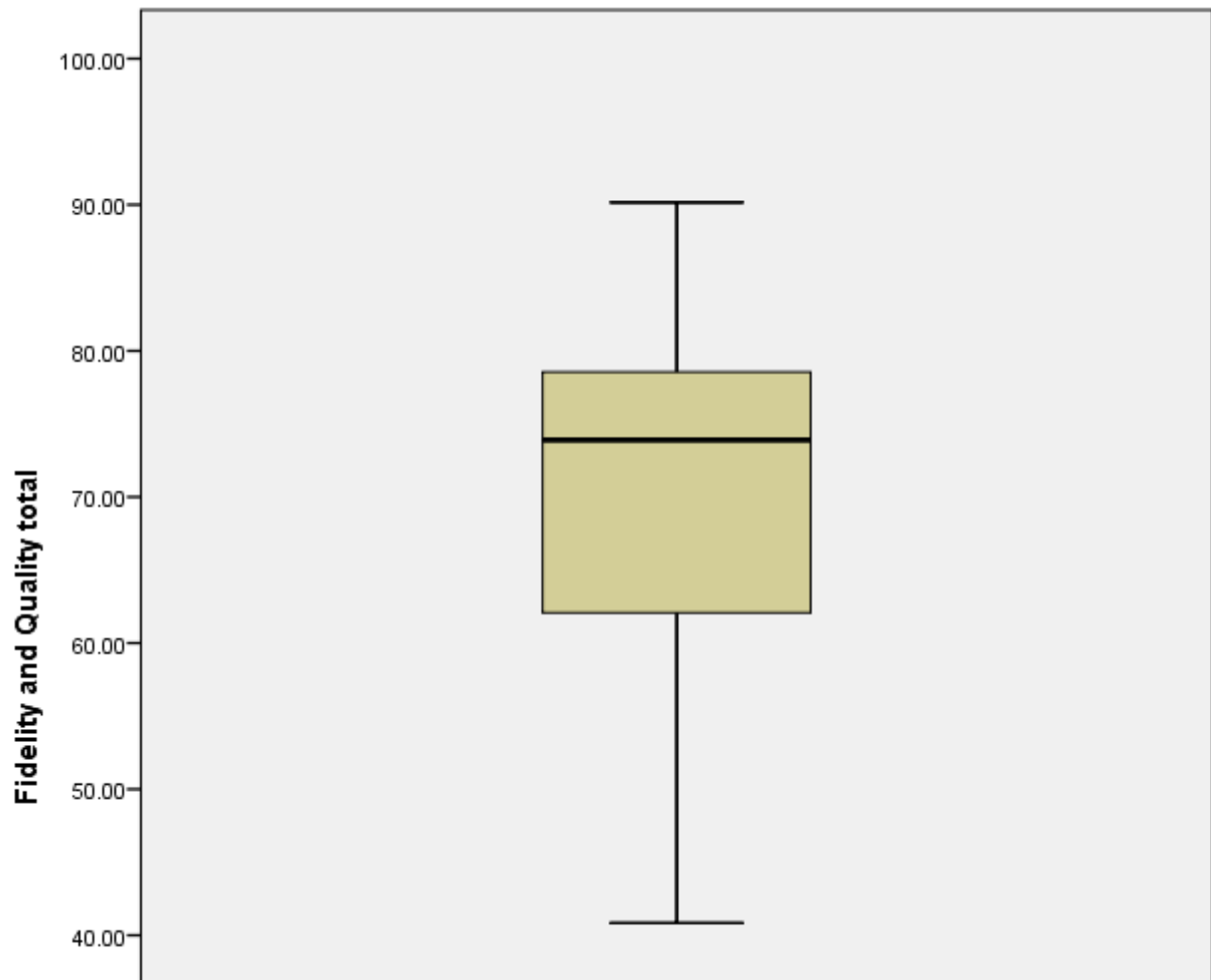
Appendix G
Hierarchical Cluster Analysis

Table G.1

Agglomeration schedule for implementation profiles using hierarchical cluster analysis

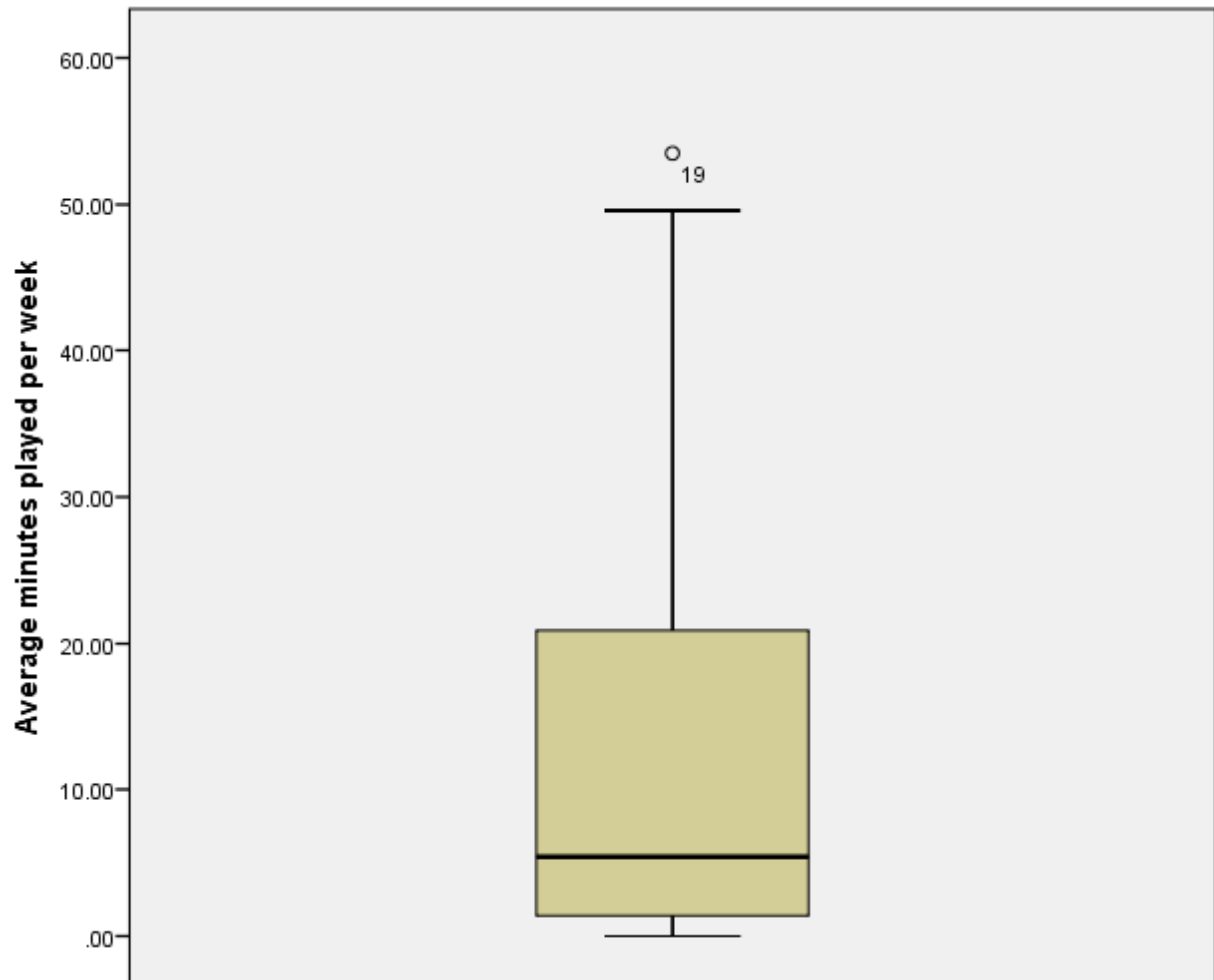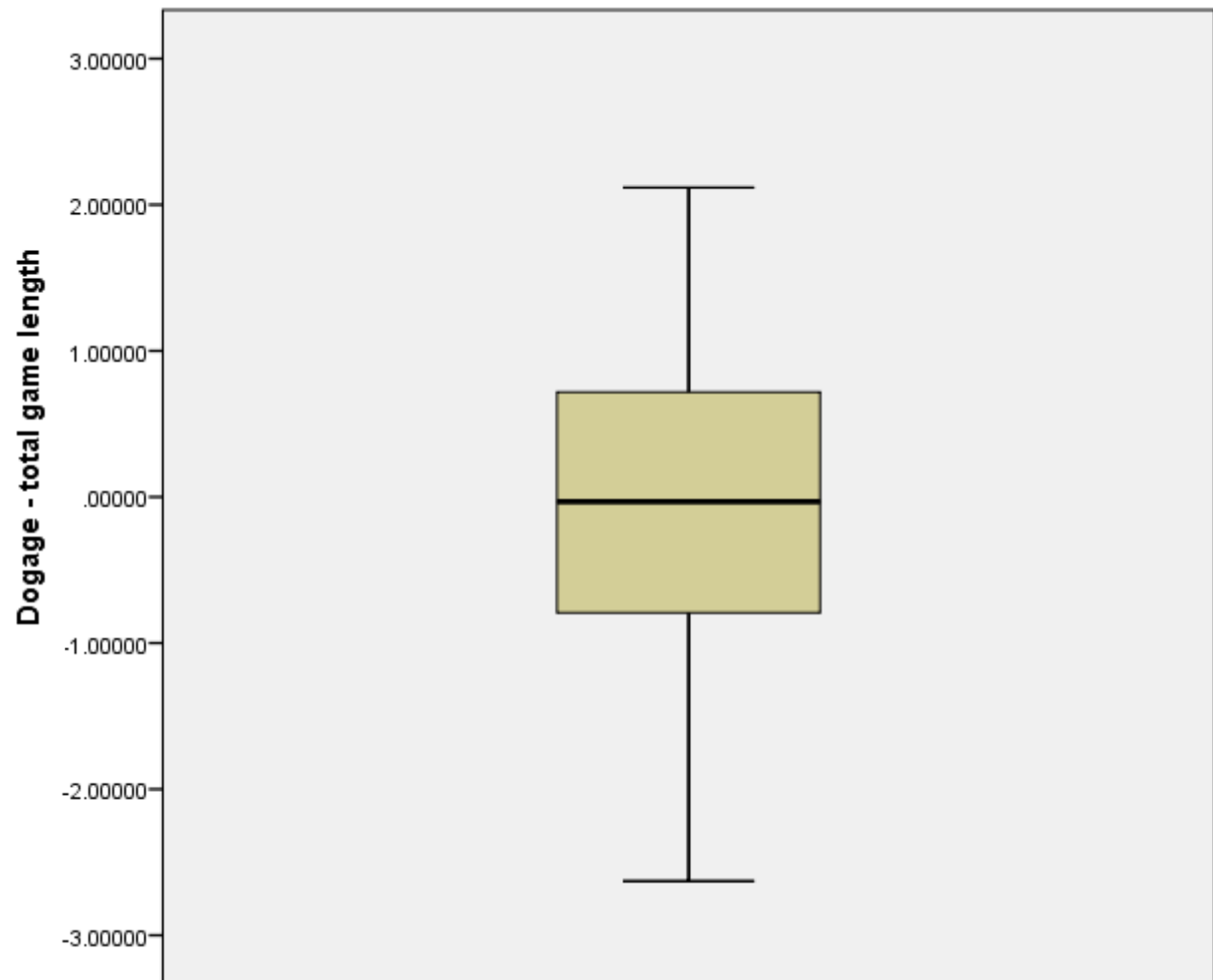|	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	81	92	.000	0	0	10
2	7	26	.000	0	0	4
3	21	85	.000	0	0	8
4	7	74	.001	2	0	9
5	5	14	.001	0	0	22
6	16	78	.001	0	0	12
7	87	89	.002	0	0	9
8	21	71	.002	3	0	36
9	7	87	.005	4	7	14
10	52	81	.005	0	1	47
11	20	66	.005	0	0	25
12	16	58	.006	6	0	24
13	3	61	.009	0	0	32
14	2	7	.010	0	9	23
15	15	65	.010	0	0	32
16	51	68	.013	0	0	27
17	4	32	.013	0	0	29
18	73	77	.013	0	0	68
19	50	62	.018	0	0	64
20	63	75	.019	0	0	70
21	31	55	.021	0	0	42
22	5	33	.022	5	0	47
23	2	56	.022	14	0	43
24	16	79	.022	12	0	54
25	20	54	.022	11	0	44
26	1	27	.029	0	0	39
27	42	51	.031	0	16	31
28	19	47	.032	0	0	38
29	4	12	.033	17	0	61
30	36	53	.035	0	0	53
31	42	67	.035	27	0	71

32	3	15	.036	13	15	50
33	57	80	.036	0	0	46
34	11	44	.038	0	0	62
35	30	48	.043	0	0	44
36	21	43	.044	8	0	51
37	41	59	.046	0	0	52
38	19	29	.050	28	0	56
39	1	69	.060	26	0	60
40	34	84	.060	0	0	58
41	10	35	.063	0	0	63
42	31	90	.067	21	0	50
43	2	39	.071	23	0	57
44	20	30	.084	25	35	59
45	6	45	.090	0	0	62
46	57	83	.105	33	0	58
47	5	52	.109	22	10	65
48	37	40	.115	0	0	69
49	49	88	.117	0	0	64
50	3	31	.117	32	42	61
51	9	21	.119	0	36	66
52	17	41	.149	0	37	72
53	36	72	.161	30	0	66
54	16	24	.162	24	0	67
55	23	28	.171	0	0	74
56	19	76	.211	38	0	63
57	2	46	.229	43	0	71
58	34	57	.237	40	46	65
59	20	25	.284	44	0	73
60	1	13	.286	39	0	70
61	3	4	.299	50	29	76
62	6	11	.305	45	34	81
63	10	19	.306	41	56	68
64	49	50	.311	49	19	67
65	5	34	.319	47	58	74
66	9	36	.336	51	53	75
67	16	49	.374	54	64	83
68	10	73	.395	63	18	79
69	18	37	.452	0	48	79

70	1	63	.453	60	20	77
71	2	42	.511	57	31	73
72	8	17	.588	0	52	75
73	2	20	.694	71	59	80
74	5	23	.732	65	55	77
75	8	9	.791	72	66	80
76	3	60	.838	61	0	81
77	1	5	.845	70	74	85
78	64	70	1.081	0	0	82
79	10	18	1.211	68	69	84
80	2	8	1.326	73	75	84
81	3	6	1.542	76	62	87
82	64	91	1.698	78	0	85
83	16	86	1.871	67	0	86
84	2	10	2.366	80	79	86
85	1	64	2.754	77	82	87
86	2	16	3.225	84	83	88
87	1	3	3.973	85	81	88
88	1	2	5.626	87	86	0

Table G.2
Coefficients schedule used to ascertain viable cluster solutions

No. of clusters	Coefficients		
	Last step	This step	Change
2	5.62	3.97	1.65
3	3.97	3.23	0.74
4	3.23	2.75	0.48
5	2.75	2.37	0.38
6	2.37	1.87	0.50
7	1.87	1.70	0.17
8	1.70	1.54	0.16
9	1.54	1.33	0.22
10	1.33	1.21	0.12

The last eleven stages were used to create a table where the differences in coefficient changes were used to determine which cluster solutions would be most suitable for answering RQ2a. The last step

column is the coefficient of the previous cluster solution, so in cluster two, the previous solution would have been a one cluster solution. As a one cluster solution is not a useful solution, the coefficient is only used to help calculate the difference for the two-cluster solution. As solutions two and three had the largest change in coefficient, these are considered the most viable solutions.