Differential effects of the Good Behaviour Game on pupils’ school functioning: Cumulative risk exposure as a moderator of intervention outcomes

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

2018

Emma L Ashworth

School of Environment, Education and Development
Contents

List of Tables ............................................................................................................ 8
List of Figures .......................................................................................................... 11
List of Abbreviations ............................................................................................. 12
Abstract .................................................................................................................... 14
Declaration ............................................................................................................... 15
Copyright Statement ............................................................................................... 15
Acknowledgements .................................................................................................. 16

Chapter 1

1. Introduction ........................................................................................................... 17
   1.1. Chapter Overview ......................................................................................... 17
   1.2 School Functioning and Risk .......................................................................... 17
   1.3 Cumulative Risk ............................................................................................ 18
      1.3.1 Ecological Systems Theory (Bronfenbrenner) ....................................... 19
   1.4 Prevention Science ......................................................................................... 22
      1.4.1 The Good Behaviour Game (GBG) ......................................................... 24
   1.5 Implementation .............................................................................................. 25
   1.6 Aims of Study and Research Questions ....................................................... 26
      1.6.1 Aims ...................................................................................................... 26
      1.6.2 Research Questions .............................................................................. 27

Chapter 2

2. Risk ...................................................................................................................... 28
   2.1 Chapter Overview ........................................................................................... 28
   2.2 Definitions of Risk ......................................................................................... 28
   2.3 Risk Factors ................................................................................................... 28
      2.3.1 Types of Risk Factor ............................................................................. 30
   2.4 Resilience (Protective and Promotive Factors) ............................................ 31
   2.5 Linking Risk and Pupils’ Outcomes ............................................................... 33
      2.5.1 Microsystem – Individual-Level Risk Variables .................................... 33
      2.5.2 Microsystem – Familial and Neighbourhood-Level Risk Variables ....... 39
2.5.3 Mesosystem – School-Level Risk Variables .................................................. 42
2.6 Chapter Summary .................................................................................................. 50

Chapter 3
3. Cumulative Risk ...................................................................................................... 51
   3.1 Chapter Overview .............................................................................................. 51
   3.2 Cumulative Risk Model .................................................................................. 51
   3.3 Theories of Cumulative Risk .......................................................................... 53
   3.4 Functional Form ............................................................................................. 55
       3.4.1 Linear ...................................................................................................... 55
       3.4.2 Quadratic ............................................................................................... 56
   3.5 Measuring Cumulative Risk ........................................................................... 58
   3.6 Chapter Summary ............................................................................................ 59

Chapter 4
4. The Good Behaviour Game .................................................................................. 60
   4.1 Chapter Overview ............................................................................................ 60
   4.2 The Origins of the GBG ................................................................................ 60
       4.2.1 Overview of the GBG ......................................................................... 60
       4.2.2 History of the GBG ........................................................................... 62
   4.3 Theoretical Underpinnings of the GBG ........................................................... 64
       4.3.1 Operant Conditioning .......................................................................... 64
       4.3.2 Social Learning Theory (SLT) .............................................................. 65
       4.3.3 Life Course/Social Field Theory ........................................................... 66
       4.3.4 Objections to the Theoretical Underpinnings of the GBG ................. 67
   4.4 GBG Logic Model ............................................................................................ 68
   4.5 Research Outcomes ......................................................................................... 70
       4.5.1 Behavioural Outcomes ....................................................................... 70
       4.5.2 Academic Outcomes .......................................................................... 75
       4.5.3 Differential Effects ............................................................................... 77
   4.6 Chapter Summary ............................................................................................ 80

Chapter 5
5. Implementation ....................................................................................................... 81
   5.1 Chapter Overview ............................................................................................ 81
5.2 The Importance of Implementation ......................................................... 81
5.3 Aspects of Implementation .................................................................... 83
  5.3.1 Fidelity, Quality and Dosage .......................................................... 85
5.4 Factors Affecting Implementation ......................................................... 87
5.5 Measuring Implementation .................................................................... 89
5.6 Linking Implementation and Outcomes ............................................... 91
  5.6.1 Linking Implementation and the GBG .............................................. 93
5.7 Chapter Summary .................................................................................. 97

Chapter 6
6. Methodology ............................................................................................ 98
  6.1 Chapter Overview .................................................................................. 98
6.2 Context of the Study .............................................................................. 98
  6.2.1 The GBG Trial .................................................................................. 98
  6.2.2 National Evaluation of the GBG ...................................................... 99
  6.2.3 Independence of the Current Study ................................................. 100
6.3 Epistemology .......................................................................................... 102
  6.3.1 Pragmatism .................................................................................... 103
  6.3.2 Pragmatic Mixed Methods .............................................................. 105
6.4 Design ...................................................................................................... 107
  6.4.1 Quantitative Approach ................................................................. 108
  6.4.2 Qualitative Approach .................................................................... 108
  6.4.3 Integration of the Quantitative and Qualitative Strands ................. 109
  6.4.4 Variables in the Study .................................................................... 109
6.5 Participants .............................................................................................. 112
  6.5.1 Schools .......................................................................................... 112
  6.5.2 Pupils ............................................................................................ 114
  6.5.3 Quantitative Strand ....................................................................... 116
  6.5.4 Qualitative Strand ......................................................................... 116
  6.5.5 Attrition ........................................................................................ 117
6.6 Materials .................................................................................................. 119
  6.6.1 Quantitative Strand ....................................................................... 119
  6.6.2 Qualitative Strand ......................................................................... 126
7.6.2 Descriptive Statistics ........................................... 182
7.6.3 Research Question 2 .......................................... 184
7.6.4 Summary of RQ2 ................................................. 189
7.7 Implementation ....................................................... 190
7.7.1 Descriptive Statistics .......................................... 190
7.7.2 Research Question 3 .......................................... 195
7.7.3 Summary of RQ3 ................................................. 203
7.8 Sensitivity Analyses .................................................. 204

Chapter 8
8. Qualitative Findings – RQ2a ............................................. 206
8.1 Chapter Overview ..................................................... 206
8.2 Perceptions of Impact ............................................... 207
8.3 Perceptions of Differential Impact ................................. 209
  8.3.1 Pupils with Low Academic Attainment ....................... 210
  8.3.2 Pupils with Behavioural Difficulties .......................... 214
  8.3.3 Gender (Male) .................................................... 217
  8.3.4 SEND Status ..................................................... 218
  8.3.5 White EAL ....................................................... 222
  8.3.6 School-Level Behaviour ....................................... 224
  8.3.7 Deprivation (Familial and Neighbourhood) .................. 225
8.4 Chapter Summary .................................................... 228

Chapter 9
9. Discussion ..................................................................... 230
9.1 Chapter Overview ..................................................... 230
9.2 Summary of Findings ............................................... 230
  9.2.1 RQ1 ............................................................... 230
  9.2.2 RQ2 ............................................................... 231
  9.2.3 RQ3 ............................................................... 232
9.3 Research Question 1: Risk .......................................... 233
  9.3.1 Findings Consistent Across Both Models ..................... 233
  9.3.2 Unique Predictors of Disruptive Behaviour .................. 236
  9.3.3 Unique Predictors of Reading Attainment ................... 238
List of Tables

Table 5.1: Aspects of implementation 84
Table 6.1: Data collection time points 100
Table 6.2: Description of differences between the current study and the GBG trial 101
Table 6.3: Purposes for conducting MMR 105
Table 6.4: Potential risk variables investigated in the study 110
Table 6.5: School-level socio-demographic data 113
Table 6.6: Pupil-level socio-demographic data 115
Table 6.7: Case study school demographics 116
Table 6.8: Attrition by risk group 118
Table 6.9: Case study data collection 119
Table 6.10: IRR scores for the implementation observation schedule 123
Table 6.11: GBG observations 2-factor EFA 124
Table 6.12: Details of data collection at each case study visit 133
Table 6.13: MDESs by behaviour risk group 139
Table 6.14: MDESs by reading attainment risk group 139
Table 6.15: Six-phase guide to thematic analysis 141
Table 7.1: Summary of data assumptions 150
Table 7.2: Means and standard deviations for outcome variables at baseline 152
Table 7.3: Correlations between predictor variables and outcome variables at baseline 153
Table 7.4: Empty MLM for disruptive behaviour outcome variable 156
Table 7.5: Empty MLM for reading attainment outcome variable 157
Table 7.6: Full MLM for disruptive behaviour outcome variable 158
Table 7.25: Mean scores for reading variable, by risk and trial group allocation

Table 7.26: Empty MLM for disruptive behaviour outcome variable at T2

Table 7.27: Empty MLM for reading attainment outcome variable at T2

Table 7.28: Imputed MLM for differential effects on pupils’ disruptive behaviour

Table 7.29: Imputed MLM for differential effects on pupils’ reading attainment

Table 7.30: Number of classes and mean (SD) scores for implementation

Table 7.31: Sample size (classes) and mean scores (SE) for implementation categories

Table 7.32: Sample sizes for implementation groups, by risk group for behaviour

Table 7.33: Mean T2 behaviour scores (SE) by risk group for implementation groups

Table 7.34: Sample sizes for implementation groups, by risk group for reading

Table 7.35: Mean T2 reading point scores (SE) by risk group for implementation groups

Table 7.36: Empty MLM for disruptive behaviour outcome variable

Table 7.37: Empty MLM for reading attainment outcome variable

Table 7.38: Imputed year 3 MLM for implementation and behaviour

Table 7.39: Imputed year 4 MLM for implementation and behaviour

Table 7.40: Imputed year 3 MLM for implementation and reading

Table 7.41: Imputed year 4 MLM for implementation and reading

Table 9.1: Discussion of adaptations
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Bronfenbrenner’s Ecological Systems Theory</td>
<td>21</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>GBG logic model</td>
<td>69</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>Overview of a paradigm</td>
<td>103</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>Overview of the study procedure</td>
<td>131</td>
</tr>
<tr>
<td>Figure 7.1</td>
<td>Significant risk variables for school functioning</td>
<td>164</td>
</tr>
<tr>
<td>Figure 7.2</td>
<td>Functional form of risk-outcome relationship for disruptive behaviour</td>
<td>173</td>
</tr>
<tr>
<td>Figure 7.3</td>
<td>Functional form of risk-outcome relationship for reading attainment</td>
<td>174</td>
</tr>
<tr>
<td>Figure 7.4</td>
<td>Cumulative risk graph for disruptive behaviour, with elbow points</td>
<td>179</td>
</tr>
<tr>
<td>Figure 7.5</td>
<td>Cumulative risk graph for reading attainment, with elbow points</td>
<td>181</td>
</tr>
<tr>
<td>Figure 8.1</td>
<td>Thematic map of global and organising themes</td>
<td>206</td>
</tr>
<tr>
<td>Figure 8.2</td>
<td>Thematic map</td>
<td>229</td>
</tr>
</tbody>
</table>
List of Abbreviations

**ADHD:** Attention Deficit Hyperactivity Disorder

**AIR:** American Institutes for Research

**CC:** Classroom-Centred

**CCR:** Check-Comment-Redirect

**CLA:** Children Looked After

**CD:** Conduct Disorder

**DfE:** Department for Education

**EAL:** English as an Additional Language

**EEF:** Education Endowment Foundation

**EF:** Effect Size

**EFA:** Exploratory Factor Analysis

**EST:** Ecological Systems Theory

**FSM:** Free School Meals

**GBG:** Good Behaviour Game

**GCSE:** General Certificate of Secondary Education

**HGRT:** Hodder Group Reading Test

**ICC:** Intra-Class Correlation Coefficient

**IDACI:** Income Deprivation Affecting Children Index

**IQ:** Intelligence Quotient

**ITT:** Intention-to-Treat

**IRR:** Inter-Rater Reliability

**KS1:** Key Stage 1
**KS2:** Key Stage 2

**MAR:** Missing at Random

**MCAR:** Missing Completely at Random

**MDES:** Minimum Detectable Effect Size

**MI:** Multiple Imputation

**MLM:** Multi-Level Modelling

**MMR:** Mixed Methods Research

**MNAR:** Missing Not At Random

**NC:** National Curriculum

**NPD:** National Pupil Database

**ODD:** Oppositional Defiant Disorder

**PATHS:** Promoting Alternative Thinking Strategies

**PWB:** Psychological Wellbeing

**RCT:** Randomised Controlled Trial

**SAS:** Social Adaptational Status

**SDQ:** Strengths and Difficulties Questionnaire

**SCPT:** School and College Performance Tables

**SEND:** Special Educational Needs and Disabilities

**SES:** Socio-Economic Status

**SLT:** Social Learning Theory

**TA:** Teaching Assistant

**TOCA:** Teacher Observation of Classroom Adaptations

**UP:** Usual Practice
Abstract

Risk factors that increase the likelihood of children experiencing poorer school functioning (academic and behavioural difficulties) have been widely researched. However, risk factors do not occur in isolation, but instead cluster together across different ecological domains (Oldfield, Humphrey, & Hebron, 2015). As they accumulate, the likelihood of negative outcomes is proposed to increase, often disproportionately. This is referred to as cumulative risk theory (Evans, Li, & Whipple, 2013). The Good Behaviour Game (GBG), a universal, classroom-based intervention, has the potential to successfully improve academic and behavioural outcomes for at-risk pupils (Flower et al., 2014). However, “at-risk” is often used as a proxy for “highly aggressive” in GBG research, thus failing to account for pupils exposed to multiple risks. Indeed, the impact of the GBG on pupils at varying levels of risk is poorly researched, and so any potential differential effects are unknown. Furthermore, research suggests that the way the GBG is implemented can have an impact on the achievement of its intended outcomes (Ialongo et al., 1999), although the evidence base is extremely limited. However, the research carried out to date on other school-based interventions does appear to suggest an interaction between high implementation and gains for pupils with higher levels of risk exposure (Abbott et al., 1998).

The study aimed to contribute to the knowledge base on cumulative risk exposure and child development, assess the impact of the GBG on pupils at different levels of risk exposure, and determine the extent to which differential intervention gains varied among at-risk pupils as a function of implementation. A hybrid concurrent embedded mixed methods design was employed to allow for a deeper exploration of the perceived ways in which the GBG influenced at-risk pupils’ outcomes. Quantitative data were collected from 3,084 pupils in 77 primary schools as part of a longitudinal randomised controlled trial of the GBG. Demographic and attainment data were collected from the National Pupil Database, while teachers completed surveys regarding pupils’ disruptive behaviour. A reading test was administered to pupils to obtain attainment data. Qualitative data were collected from six self-selecting case study utilising semi-structured interviews and focus groups with teachers and pupils. Data analysis was conducted using multi-level modelling and thematic analysis.

Analyses revealed six risk variables for disruptive behaviour, and seven risk variables for reading attainment, operating at pupil- and school-levels. A cumulative risk effect was identified, and a quadratic risk-outcome relationship was determined to be present, indicative of an exponential increase in negative outcomes as risk exposure increased. However, there were no statistically significant main or subgroup effects of the GBG on at-risk pupils’ outcomes. Higher levels of implementation fidelity and quality were significantly associated with worsening disruptive behaviour scores for high-risk pupils. Qualitative analysis identified perceived differential gains for pupils exposed to six different risk variables; 10 proximal outcomes also emerged, along with five common GBG elements considered to be the mechanisms through which these outcomes were influenced. The implications of these findings and directions for future research are discussed.
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

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University’s policy on Presentation of Theses.
Acknowledgements

I would firstly like to thank the pupils, teachers and schools involved in the trial, whose participation made this research possible.

Thanks also to my first and second supervisors, Professor Neil Humphrey and Dr Ann Lendrum for their invaluable guidance, support and encouragement over the last three years. I am also grateful to Dr Alex Hennessey for her endless statistical advice, and the wider MIE team for their help and friendship.

Finally, thanks to my family and friends for their unconditional support and belief in the completion of this research.
1. Introduction

1.1. Chapter Overview

This chapter aims to provide an overview of the key concepts and terminology utilised in the current study. The chapter begins by discussing the concept of ‘school functioning’, and explores the role that risk plays in pupils’ outcomes, with Bronfenbrenner’s ecological systems theory providing a conceptual framework for this. The fields of prevention and implementation science are then introduced, with a particular focus on one universal, school-based intervention, the Good Behaviour Game. Finally, the aims and research questions are outlined.

1.2 School Functioning and Risk

School functioning, as defined by Dick and colleagues (2010), refers to a “wide range of factors including school attendance, academic achievement and social relationships” (p. 238). Previous research on good school functioning has also extended this definition to include behavioural factors such as focused attention on the teacher (Meijer, Habekothe, & Van Den Wittenboer, 2000) and the absence of frequent discipline problems (Attar-schwartz, 2009). In the context of this study, two important aspects of school functioning are considered: academic attainment and disruptive behaviour. There is currently conjecture in the literature as to what types of behaviours are classed as disruptive (DfE, 2012). For the purposes of the current study, disruptive behaviour is considered to be comprised of: aggressive behaviour (e.g. hitting, pushing, abusive language), physically disruptive behaviour (e.g. damaging objects, throwing objects), socially disruptive behaviour (e.g. screaming, temper tantrums), and authority-challenging behaviour (e.g. defiant verbal or non-verbal behaviour) (Cameron, 1998).

Factors that increase the likelihood of children experiencing greater behavioural and academic difficulties have been widely researched and are frequently referred to as “risk factors”. In the literature, risk factors are generally grouped into four
categories: individual, school, family and community factors (Deater-Deckard et al., 1998). A range of child-level risk factors have been identified including: male gender, being a looked-after child, medical problems, low intrinsic motivation, poor mental health, negative relationships with peers and teacher, and belonging to an ethnic minority group (Huffman, Mehlinger, & Kerivan, 2000). Furthermore, early academic and behavioural problems such as poor attendance (Christenson et al., 2004; Christiansen, Christiansen, & Howard, 1997), low academic achievement, and high levels of aggression and anti-social behaviour are highly predictive of later academic and behavioural problems (Campbell & Ewing, 1990; Patterson, Reid, & Dishion, 1992). Hansen and colleagues (2010) reported that behavioural problems at age seven are strongly associated with behavioural problems at ages three and five, suggesting stability over time; whilst Webster-Stratton, Reid and Stoolmiller (2008) found that children with early onset behaviour problems are at high-risk for academic failure, school absences, eventual conduct disorders (CD) and school dropout, indicating that the two outcomes may be inter-related.

School-level factors can also increase a pupil’s risk of low school functioning, although less research has been conducted in this area. Larger schools, situated in urban areas, with low average SES have been found to increase the chances of a child experiencing academic and behavioural difficulties (Huffman et al., 2000). Family-level factors including changes to family composition, low maternal education and poor mental health, and insecure attachments have also been shown to have an impact on outcomes. The present study contributes to the field of prevention science by identifying which factors influence children’s school functioning in academic and behavioural domains. Whilst also considering individual-level factors, this study extends understanding of risk by identifying school-level factors that can increase the probability of low school functioning.

1.3 Cumulative Risk

In addition to investigating the predictive utility of individual risk factors on school functioning, cumulative risk exposure can also be examined. Research into
cumulative risk indicates that as the number of contextual risk factors accumulate, the problems associated with them increase (Trentacosta et al., 2008), often disproportionately (Oldfield et al., 2015). The cumulative risk effect for poor school functioning is well established, with research consistently showing that as the number of risk factors increase, language and mathematic achievement scores decrease, whilst problem behaviour increases (Deater-Deckard et al., 1998; Gutman, Sameroff, & Eccles, 2002; Prelow & Loukas, 2003). Pungello and colleagues (2010) found that exposure to each additional risk factor was associated with 0.28 fewer years of education, being 1.45 times less likely to graduate from high school, and being 1.61 times less likely to be employed. These findings have been replicated several times with children from low SES backgrounds (Prelow & Loukas, 2003), those that have been diagnosed with a special educational need or disability (SEND; Oldfield et al., 2015), those in ethnic minority groups (Gutman et al., 2002), and those who have been exposed to multiple familial and community risk factors (Pungello et al., 1996, 2010).

The current study aims to identify whether the influence of established risk factors can be replicated, in addition to investigating cumulative effects of exposure to risk factors on school functioning. The study also extends the existing knowledge base by investigating the functional form of the risk-outcome relationship.

1.3.1 Ecological Systems Theory (Bronfenbrenner)

One of the models that best accommodates cumulative risk theory is Bronfenbrenner’s ecological systems theory (EST), as it accounts for the clustered nature of risk factors nested in a variety of contexts. EST suggests that in addition to being directed by internal characteristics, human development is also shaped by the immediate and broader social and psychical environments, and the interrelationships between them (Atzaba-Poria, Pike, & Deater-Deckard, 2004). According to Bronfenbrenner, human development occurs within a progressively complex system of continuous and reciprocal relationships, consisting of the child’s environment, which comprises several layers. EST posits that there are both
proximal and distal systems influencing the child, known as ecological domains, which never function independently of one another. These domains include (Bronfenbrenner, 1986; Evans et al., 2013; Powers, 2010):

The microsystem: The innermost level, consisting of a child’s immediate surroundings, including the home, family and neighbourhood. The child’s relationships with their microsystems can have a direct influence on their behaviour.

The mesosystem: Another proximal system, primarily comprising the interrelations between a child’s microsystems, although it also includes the relationships between a child’s microsystem and more distal ecological domains. The mesosystem is thus succinctly described by Bronfenbrenner (ibid) as “a system of microsystems” (p.515) and includes interrelationships such as home-school links.

The exosystem: A more distal system, considered to be extension of the mesosystem, including social structures that do not directly influence the developing child, but instead impact on the immediate settings they inhabit, thereby indirectly influencing their development. Examples of these structures include the school’s board of Governors, the media and parents’ place of work.

The macrosystem: This differs from the other systems in that it refers to more abstract, overarching constructs including the broader societal and cultural values. Macrosystems therefore set the pattern for the more specific activities and structures occurring at the more concrete levels.

The chronosystem: This involves the temporal dimensions of the model, referring to the timing of an individual’s development and how this interacts with historical time. This also involves changes occurring during the child’s lifetime, and how this influences the child’s development.

Although it is beyond the scope of the study to investigate the wider factors in children’s chronosystems, they may still indirectly be influencing children’s
outcomes. For example, children’s mental health problems are currently a focus in research literature, Government agendas and the media (e.g. the mental health Green Paper; DfH & DfE, 2017). As one in five children have reported experiencing a mental health problem (Deighton et al., 2018), and 70% of children do not receive appropriate interventions at a sufficiently early age (Children’s society, 2008), much more is currently being done in schools to combat these problems; hence the Department for Education’s (DfE) recent report (2016b) offering advice to schools regarding the promotion of positive mental health in pupils. This report outlines the cumulative detrimental effect of risk factors and ways to address these at an early age in order to promote resilience. Therefore, whilst this is not directly investigated in the present study, the social and historical context in which this research was carried out is clearly important.

Figure 1.1: Bronfenbrenner’s Ecological Systems Theory
1.4 Prevention Science

Prevention science is a relatively new field, having only begun to emerge in the last few decades. It draws from multiple disciplines and a wide variety of research including studies investigating epidemiology, risk and protective factors, and the development of interventions to prevent risk, disease, disorder or injury. Prevention science also includes research on the dissemination and scaling-up of a range of preventive public health interventions targeting biological, psychological, social and environmental risk factors that increase the likelihood of reduced wellbeing, disorders and illnesses (Sloboda & Petras, 2014).

In relation to the current study, prevention science is the study of hypothetical precursors (commonly known as risk factors) to academic failure and behavioural difficulties. The aim of prevention science is to preclude or moderate these negative outcomes in children by implementing interventions before the problems are fully manifested. Preventive interventions aim to counteract the risk factors, whilst also reinforcing the protective factors that disrupt the processes contributing to the negative outcomes. It is thought that for a preventive intervention to be successful, a developmental theory is required, connecting the risk factors, mediating processes, and maladaptive patterns of behaviour. The intervention should then target those risk factors and associated mediating mechanisms, in order to enhance children’s protective factors, thus reducing the likelihood of maladaptive outcomes (Coie et al., 1993). Similarly to Bronfenbrenner’s theory outlined above, Coie and colleagues (ibid) emphasise that exposure to risk can occur in diverse and dynamic ways, and in a number of settings. An area of prevention science that they therefore highlighted as a priority for future research was the acknowledgement of the complex transaction between individuals and their environments, and between systems of influence, across different periods of time.

Schools are considered to be one of the most important locations for the implementation of preventive interventions, namely because they are the only...
setting with universal access to children. Childhood is also a time when many antecedents to later problems occur, and so it is important to intervene early to promote healthy development, before these issues become more established (Greenberg et al., 2005). 75% of schools reported using preventive interventions to deal with behaviour problems, and more than three quarters of schools offer mental health, social service, and prevention services to pupils (Wilson & Lipsey, 2007). The analysis of these school-based preventive interventions is an area of research that is rapidly expanding in the literature, with Durlak and colleagues’ (2011) meta-analysis including 213 studies of universal intervention programs. Indeed, Wilson and Lipsey (2007) reported that the most common and most effective programmes are universal interventions; these are interventions that are delivered to all the pupils in a classroom or school, as opposed to selective (targeted to subpopulations identified as being at increased risk) or indicated (targeted to individuals who are identified as having an increased vulnerability based on an individual assessment) interventions (O’Connell, Boat, & Warner, 2009). They also reported that larger effect sizes are evidenced if the intervention has a behavioural component, and that larger treatment effects are found for higher-risk children, with economically disadvantaged children demonstrating the greatest benefits of universal interventions.

Universal school-based preventive interventions seem to be particularly effective for those deemed to be at-risk (Hutchings et al., 2013; Jones, Brown, & Aber, 2011). This is congruent with the idea that “influencing a risk factor or a set of risk factors though a preventive intervention would lower the proportion of the population needing to be served through more intensive levels of interventions and services for problems attributable to that risk factor” (Poduska, Kellam, Wang, Hendricks Brown, et al., 2008, p. s30). Whilst these interventions, such as the Good Behaviour Game (discussed in more detail below), have generally been shown to be successful (Flower et al., 2014), they tend to focus on one aspect of school functioning, or only use a single variable as a measure of “at-risk” (e.g. Kellam et al., 1998). Therefore, the current study incorporates advice from Coie and colleagues (1993) by examining the accumulation of risk factors in various ecological domains, and the
differential effects that a universal preventive intervention has on children exposed to multiple risk factors.

1.4.1 The Good Behaviour Game (GBG)

The GBG is a universal classroom behaviour management strategy described by Tingstrom, Sterling-Turner and Wilczynski (2006) as an “interdependent group-oriented contingency management procedure”. Previous research suggests that the game is successful in reducing disruptive and aggressive behaviours, and in increasing concentration and on-task behaviours. These factors are considered to be “early risk behaviours” and have been associated with later conduct disorders and anti-social behaviour (Ialongo et al., 2001), as well as substance abuse, depression (Ialongo et al., 1999) and juvenile delinquency (Reid et al., 1999). Research into the GBG has also looked at the effects on pupils exposed to the underlying factors associated with low school functioning. For instance, there is substantial research supporting the positive effects of the GBG on the behaviour of children from low socio-economic status (SES) backgrounds (Dolan et al., 1993; Kellam et al., 1994, 1998, 2011), and boys have consistently benefited from the game significantly more than girls.

Whilst current GBG research typically suggests that at-risk children will benefit from the intervention, they generally use the term “at-risk” as a proxy for “highly aggressive”. This implies that risk is binary, when evidence suggests that risks occur in clusters, as implied by cumulative risk theory (e.g. Evans et al., 2013). However, the impact of the game on pupils exposed to different levels of cumulative risk has not been researched, and so the knowledge base is limited, although there is conjecture. For example, Muthén and colleagues (2002, p. 461) hypothesised: “GBG may have its largest effect for those who are in the middle trajectory class, showing milder forms of problems, while not being strong enough to affect the most seriously aggressive children and not needed for members of the stable non-aggressive group”. The present study addresses this issue by adopting a cumulative risk approach.
1.5 Implementation

Research suggests that the way the GBG is implemented can have an impact on the achievement of its intended outcomes. Indeed, Derzon and colleagues (2005) argue that the implementation process shapes the quality of the participants’ experience. Implementation can be assessed in a number of ways, although the most common method is the measure of fidelity; this is the extent to which teachers adhere to the GBG guidance. Whilst Flower and colleagues’ GBG meta-analysis (2014) identified eight studies that reported fidelity, only three GBG studies have looked at the relationship between fidelity and student outcomes. The largest of these studies (Ialongo et al., 1999) found that implementing the GBG poorly produced similar outcomes to not implementing it at all. Thus it appears that implementation fidelity can act as a moderator.

However, while studies of other universal interventions (Abbott et al., 1998; Kam, Greenberg, & Walls, 2003) have found that high implementation results in greater reductions in externalising behaviours, other aspects of implementation are often neglected in studies of the GBG. Whilst a couple of other studies mention measures of quality (how well the GBG is delivered) and dosage (how frequently the game is played and for how long), and report their findings (Pas et al., 2015; Tanol, 2010), they do not go on to assess the effects of these as moderators of outcomes. The present study addresses this issue by examining the associations between three aspects of implementation (fidelity, quality, and dosage) and the outcomes of the GBG.

Whilst the limited research carried out to date appears to suggest an interaction between high implementation and gains for children, there is currently no research available that focuses on the interaction between implementation and differential gains for children at different levels of risk exposure, either in the GBG or in other interventions. Thus it is unclear whether the level of implementation will moderate the cumulative risk effect. The current study addresses this critical gap in the extant research base.
1.6 Aims of Study and Research Questions

1.6.1 Aims

Given the above, the aims of the proposed study are to:

1. **Increase knowledge and understanding of the risk variables for poor school functioning in academic and behavioural domains among children in primary school**

   The fulfilment of aim 1 contributes to the knowledge base on risk exposure and child development (Evans et al., 2013) by, for example, determining the functional form of the risk-outcome relationship in academic and behavioural domains.

2. **Assess the impact of the GBG on children at different levels of risk exposure**

   The impact of the GBG on at-risk children has been researched relatively infrequently (approximately 5% of published GBG studies), and risk variables are generally treated as independent, whereas in reality they are clustered together (Appleyard et al., 2005). The study aims to address this issue by investigating the effects of the GBG as a function of cumulative risk exposure.

3. **Determine the extent to which differential benefits among at-risk children vary as a function of implementation**

   The study aims to address a major gap in the literature regarding the possible interaction between implementation variability and cumulative risk exposure in determining the achievement of intervention outcomes.

Overall, the study represents an attempt to consider not just “what works”, but what works for whom, under what conditions and circumstances (Bonell et al., 2012).
1.6.2 Research Questions

Taking into account the above literature, the research questions for the current thesis are as follows:

1. What are the risk variables associated with lower school functioning?
   a. For those variables identified, what is the magnitude of their association with behavioural (disruptive behaviours) and/or academic (reading) outcomes?
   b. Is there evidence of a cumulative effect of risk exposure on these outcomes?
   c. What is the functional form of the risk-outcome relationship?
   d. Is the number of risk variables a better predictor of outcomes than their relative independent strength?

2. Are there differential intervention gains in behavioural and/or academic outcomes among children at different levels of cumulative risk exposure?
   a. In what way does exposure to the GBG influence behavioural and/or academic outcomes for at-risk children?

3. Are there any differential associations between implementation of the GBG (specifically fidelity, quality, dosage) and behavioural and/or academic outcomes for at-risk children?
2. Risk

2.1 Chapter Overview

This chapter aims to summarise issues surrounding the terminology of risk in the literature, before providing a working definition of risk for the purposes of the present study. An overview of the risk factors explored in the present study is provided, along with a summary of previous research surrounding their associations with pupils’ academic and behavioural outcomes. These risk factors are categorised according to the conceptual framework outlined in the previous chapter. Whilst a multitude of risk factors exist, and the research regarding some of these factors is vast, only a brief discussion of the factors included in the present study is provided here due to the limited space available.

2.2 Definitions of Risk

Terms such as “risk”, “risk factors”, and “correlates” are inconsistently and imprecisely used; Kraemer and colleagues have attempted to clarify these terms, suggesting that “risk” means “a probability of an outcome” within a population of subjects, whilst “risk factor” is a “correlate shown to precede the outcome”. A correlate is defined as a “measure somehow associated with the outcome” (Kraemer, 1997; Kraemer et al., 2001, p. 849).

2.3 Risk Factors

Risk factors have varying definitions, although they are generally considered to be “characteristics, variables, or hazards, that, if present for a given individual, make it more likely that this individual, rather than someone selected at random from the general population, will develop a disorder” (Arthur et al., 2002, p. 576). Emphasis in the majority of definitions is placed on the increased likelihood or probability of a maladaptive or poor developmental outcome for certain groups of individuals, relative to the rest of the population; although it is often falsely suggested that risk
factors can predict negative outcomes (Fraser & Terzian, 2005; Gutman et al., 2002; Masten, 2004; Smith & Carlson, 1997). In order for a variable to be defined as a risk factor, a high- and low-risk group must be identifiable (Kraemer, 1997).

Zolkoski and Bullock (2012) stressed that “risk factors are not black boxes to fit children where they can be neatly labelled and safely stored away” (p2295). Instead, their effects are dependent on the developmental systems assessed, such as children’s micro and mesosystems, individual variations in the responses of children, and the time and place of assessment. Therefore, it is suggested that risk factors can have different effects at different stages in children’s lives. As many risk factors are considered to be fairly stable over time, it is thought that older children, who are exposed to them for longer, may demonstrate more significant negative outcomes (Stouthamer-Loeber et al., 2002). As the current study only follows children for two years, between the ages of seven and nine, different effects may be found in older or younger samples.

Ultimately, risk factors are measurable characteristics or traits that can be categorised into one of several ecological domains. Variables can be identified as child characteristics (age, gender, SEND), school-level factors (size, average achievement, urbanicity), social- and famililial-factors (peers, relationships with caregiver, maternal education), or wider socio-demographic characteristics (neighbourhood SES, delinquency). The model that best accommodates this wide variety of risk factors is Bronfenbrenner’s EST (1986), whereby varying characteristics are located in nested contexts, with child characteristics having the most proximal influences on development. The more distal variables (e.g. socio-demographic characteristics) are thought to influence the child by exerting their effect through the more proximal variables (Stouthamer-Loeber et al., 2002). However, risk factor research does not “take into account the meaning of a factor to the child, focusing instead on statistical associations between certain factors and adverse outcomes” (Smith & Carlson, 1997, p. 234). Thus the current study adopts a mixed methods approach, utilising qualitative research methods in order to account
for this criticism of risk research, meaning that the associations between risk factors and outcomes can be explored in greater detail.

2.3.1 Types of Risk Factor

As no limitations are placed on the level of measurement of a factor (i.e. they can be nominal or ordinal), on the number of items used to define the factor (i.e. univariate or multivariate), or the distribution of the factor measures, there are varying types of risk factor that are explored in the literature (Kraemer, 1997).

Firstly, risk factors can be categorised as fixed or variable markers, depending on whether or not they are malleable. If a risk factor cannot be changed, thus holding a trait-like stability, it is considered to be a fixed marker (e.g. race, gender); whilst a risk factor that can change spontaneously within a subject, or through intervention, is classified as a variable marker (e.g. age, weight) (Kraemer, 1997; Leyro et al., 2010). However, Leyro and colleagues (2010) argue that only a variable that can be changed can be considered a risk factor; when they are fixed, they are better characterised as “risk markers”. Therefore, for the purposes of this study, the variables utilised are referred to as “risk variables”, thus including both malleable risk factors and fixed risk markers (Hebron, Oldfield, & Humphrey, 2016).

A variable risk factor that can be manipulated, or when manipulated can change the likelihood of a negative outcome, is termed a causal risk factor. When these variables are modified, they produce a systematic change in the outcome variable in individuals who did not previously display the identified problem (Kraemer, 1997; Leyro et al., 2010). However, proxy risk factors occur when they are only related to the outcome variable through their relationship with another causal risk factor, or they are a small part of a more global risk factor. Therefore, manipulating a proxy factor would not evidence a corresponding systematic change in the outcome variable; hence a proxy factor may “mark” risk, but not explain it (Kraemer et al., 2001; Leyro et al., 2010). For example, low SES has been shown to be a risk variable for attention deficit hyperactivity disorder (ADHD); however, various aspects of SES
(e.g. inadequate access to healthcare, poor parental occupational status, low income) are also likely to be risk variables for ADHD, and so all would be considered to be proxy risk variables (Kraemer et al., 2001).

2.4 Resilience (Protective and Promotive Factors)

Whilst there is an abundance of evidence in the literature suggesting that risk variables increase the likelihood of a child experiencing a negative outcome, they cannot guarantee it. In spite of the presence of risk variables in a child’s life, some individuals appear to overcome the odds, adapt to adversity, and achieve better than expected outcomes; these children are deemed to be resilient (Alvord & Grados, 2005; Benard, 1995; Brooks, 2006; Gutman et al., 2002). Resilience is broadly defined as the “skills, attributes and abilities that enable individuals to adapt to hardships, difficulties, and challenges” (Alvord & Grados, 2005, p. 238). For example, a child who has experienced life events such as trauma, poverty or abuse, yet still interacts with the environment in a positive way, and with relative success, would be considered to be displaying attributes and skills commonly characterised by resilience (Christiansen et al., 1997).

In order to be defined as resilient, an individual must show positive outcomes across several domains over different periods of time (Cicchetti & Rogosch, 1997). Furthermore, resilience is thought to be made up of multiple skills and qualities, and is not a one-dimensional, dichotomous attribute (Reivich & Shatte, 2003). Instead, resilience focuses on promotive and protective factors that can contribute to positive outcomes despite exposure to greater than average risk (Brooks, 2006); these factors are commonly cited in the literature, with research suggesting that they can counteract the effects of exposure to risk variables.

However, the definitions for these are far from clear, with the terms often being used interchangeably (Stouthamer-Loeber et al., 2002). Generally, promotive factors refer to the “the positive end of an independent variable’s distribution” (ibid, p.112). Commonly, variables are considered in the literature to be either risk or
promotive; although Stouthamer-Loeber and colleagues (ibid) argue that by viewing risk and promotive variables as separate, we can never consider that a variable can act for one person as a risk variable, for another as a promotive factor, and for a third be neutral. Thus, it is suggested that promotive and risk factors are represented as opposite poles of the same variable, and so the effect the variable has is dependent on which pole an individual scores closest to. Conversely, it is thought that protective factors are “those attributes of persons, environments, situations, and events that appear to temper predictions of psychopathology based upon an individual’s at-risk status” (Garmezy, 1993, p. 73); in other words, they interact with risk variables to reduce the probability of a negative outcome.

Protective factors are considered to be the “building blocks of resilience” (Powers, 2010, p. 447); these factors have been repeatedly investigated and corroborated over the years, resulting in a “shortlist” of correlates produced by Masten that indicate better adaptation among children at-risk, in each ecological domain (Wright, Masten, & Narayan, 2013). This shortlist is extensive, although Benard’s summary of resilience research (1995) grouped those protective factors most commonly associated with resilience into three main categories: caring and supportive relationships, positive and high expectations, and opportunities for meaningful participation. Similarly to risk variables, protective factors can operate at multiple ecological levels, both in an individual’s immediate environment, and the larger social and cultural contexts in which they are embedded (Olson & Goddard, 2010).

Whilst it is beyond the scope of this study to investigate both risk and promotive or protective factors, and the relationships between them, the reasons for some children’s resilience to their risk status may help to explain any unanticipated findings following analysis of the data.
2.5 Linking Risk and Pupils’ Outcomes

There are an abundance of variables identified in the extant literature as risk variables for children’s academic and behavioural outcomes. Whilst risk variables exist in all ecological domains, based on the data available for this study, only risk variables associated with children’s proximal domains will be investigated (Powers, 2010). This is a common constraint of risk research; a comprehensive evaluation of all risk variables would be unfeasible due to data and resource limitations. The risk variables being investigated as part of this study are explained in more detail below.

2.5.1 Microsystem – Individual-Level Risk Variables

A plethora of risk variables exist within a child’s microsystem, at various levels. Within the current study, risk variables at the individual, familial and neighbourhood levels will be investigated, in order to establish their associations with pupils’ school functioning. The list of variables being examined is not an exhaustive one, although they are amongst the most salient risk variables in the literature. Whilst it is beyond the scope of the study to explore all significant risk variables, they are acknowledged here to emphasise the diversity of research available. Previously identified risk variables in children’s microsystems include adverse temperament, medical problems, low intelligence, conflict or violence in the home, harsh parental discipline, poor parent-child relationships, instable peer relationships and social rejection. Biological factors such as genetic differences, hormonal influences and neurotransmitter processes have also been found at the child-level (Deater-Deckard et al., 1998; Lösel & Farrington, 2012).

2.5.1.i Relative Age in the School Year

Research shows that a pupil’s relative age is a risk variable for school functioning; this is often referred to as a month-of-birth effect. Studies have found that the younger the child in the class, the more likely they are to be described as having an SEND linked to literacy attainment (Squires et al., 2012). Boys’ late birthdays are
also associated with early grade failure and lower reading test scores (Fowler & Cross, 1986; Jones & Mandeville, 1990). These differences appear to be sustained in the long-term, with similar results still being found at GCSE (Menet et al., 2000).

Additionally, a larger proportion of children are diagnosed with behavioural and emotional disorders if they are born later in the school year (Polizzi, Martin, & Dombrowski, 2007). Suggested explanations for this include that younger children’s central nervous systems are less mature, specifically regarding self-regulation and inhibitory control, making them cognitively less equipped to deal with classroom demands. They are also thought to have poorer social skills which can lower self-esteem, thus reducing on-task behaviour and achievement (ibid). However, some studies suggest that the relative age effect is only present in the earlier years of schooling (Lien et al., 2005; McPhillips & Jordan-Black, 2009; Menet et al., 2000). The types of behavioural problems being investigated also appear to influence outcomes, with some studies finding that only internalising symptoms are significantly related to relative age (Patalay et al., 2015).

These issues notwithstanding, relative age is an established risk variable for school functioning, and so is included in the current study.

2.5.1.ii Gender

The “gender gap” is a long-standing issue whereby boys appear to consistently underachieve relative to girls in examinations throughout schooling (Smith, 2003; Strand, Deary, & Smith, 2006), with almost a five point difference in Attainment 8 scores at GCSE (DfE, 2017b). It is thought that gender socialisation practices (Dempster, 2011; Francis, 2006), and the education system utilising styles of teaching and assessment that favour girls (Department for Education & Skills, 2007; Machin & McNally, 2005) account for the gap. However, the gap is significantly reduced by higher education (Gorard, Rees, & Salisbury, 2001), there is evidence to suggest that it may be subject-specific (Dee, 2007), and there are no apparent differences in IQ between genders (Strand et al., 2006). Thus, the gap may not be as
wide as it appears; in fact, the social class attainment gap at Key Stage 4 is three times wider than the gender gap (Department for Education & Skills, 2007).

Regarding behaviour, evidence suggests that boys are significantly more at-risk for behavioural problems than girls; they are more likely to be aggressive, have conduct problems, and develop anti-social behaviour (Department for Education & Skills, 2007; Kellam et al., 1998; Stovoll & Wichstrom, 2002). Indeed, boys are twice as likely to display serious behaviour problems in the UK (Hansen et al., 2010), and are most commonly excluded for disruptive behaviour or physical assault (Department for Education & Skills, 2007). The reasons cited for the gender differences in behaviour include biological or hormonal differences (Craig & Halton, 2009), pressures to conform to stereotypes (Jackson, 2010) and societal expectations (Banerjee & Lintern, 2000).

Based on the available evidence, being male is an established risk variable for school functioning, and so is included in the current study.

2.5.1.iii Special Educational Needs and Disabilities

There is substantial evidence to suggest that SEND status is associated with poorer school functioning (Morrison & Cosden, 1997). Over the last several decades, pupils with an SEND have consistently been more likely to dropout or be excluded from school (Levin, Zigmond, & Birch, 1985; Stamou et al., 2014). However, it is unclear whether the cause is behavioural or academic; whilst one of the most at-risk groups for behaviour difficulties is SEND pupils (Oldfield et al., 2015), students with SEND are also considered to be the most vulnerable group of learners academically (Humphrey et al., 2013), thus school functioning may be influenced by both factors.

The type of SEND diagnosis also influences successful school functioning. Pupils diagnosed with social, emotional and mental health difficulties display “challenging, disruptive or disturbing” behaviours (DfE, 2015c, p. 98), and so are at increased risk of behavioural difficulties by definition. Pupils with specific learning difficulties such
as dyslexia also have lower academic outcomes in certain areas (Miles, Haslum, & Wheeler, 2001), while ADHD is consistently linked to both decreased academic achievement and increased disruptive behaviour problems (Biederman et al., 2004; Loeber et al., 1995; Loeber & Keenan, 1994; van Lier et al., 2004).

Some research suggests that there are certain underlying issues that make school functioning more difficult for SEND pupils. These include a lack of protective factors (Finn & Rock, 1997; McMillan & Reed, 1994) due to lower self-esteem (Tracey, Marsh, & Craven, 2003), fewer positive relationships with peers (Martlew & Hodson, 1991), and increased bullying and victimisation (Norwich & Kelly, 2004). Furthermore, learning disabilities are associated with higher levels of anxiety (Nelson & Harwood, 2011) which can have “deleterious effects on performance on cognitive and academic tasks” (p.4), disrupting attentional focus and the information processing system, and decreasing the use of metacognitive skills, thus negatively affecting academic outcomes. By compounding struggles with learning, learned helplessness or task avoidance strategies may be utilised, resulting in poorer behavioural outcomes (Turtura, Anderson, & Boyd, 2014).

Thus, SEND status is an established risk variable for school functioning, and so is included in the current study.

2.5.1.iv Minority Ethnic Group

Differences between ethnic groups have consistently been identified in the literature regarding behavioural problems, with Black Caribbean pupils being four times more likely to be excluded due to misbehaviour (Hansen et al., 2010; Stamou et al., 2014). Furthermore, whilst White pupils have below average levels of behavioural problems, Bangladeshi and Black Caribbean pupils’ behavioural difficulties have been found to be significantly higher (Brown & Schoon, 2008). However, other research evidences limited associations between ethnicity and behaviour (Dekovic, Wissink, & Meijer, 2004), although the age of onset may
influence outcomes. Indeed, one study found that differences did not emerge until adolescence (Greenberger et al., 2000).

Significant differences between ethnic groups are also evident regarding academic achievement. Whilst Chinese pupils make up the highest performing group at GCSE, with 74.4% achieving the Government’s benchmark, only 53.1% of pupils from a Black background achieve the same level (DfE, 2015b). However, research has shown that other factors, such as gender, may be moderating ethnicity effects (Leadbeater & Bishop, 1994). Indeed, Black boys are the group least likely to achieve the benchmark, but Black girls are not disadvantaged to the same extent. However, White British boys eligible for FSM are also a group with particularly low attainment (Department for Education & Skills, 2007). Reasons posited for the differences in outcomes between ethnic groups include parenting practices, family structure, familial values about education, and beliefs about the occupational rewards of academic success (Steinberg, Dornbusch, & Bradford Brown, 1992). Expectations of acceptable behaviour are also thought to differ between cultures (Guttmannova, Szanyi, & Cali, 2007).

The somewhat inconsistent results detailed above, and the other proxy variables evident in some research, indicate that it is unclear whether ethnic minority status is significantly associated with school functioning. Therefore, it is included in the current study in an attempt to bring clarity to this area.

2.5.1.5 Linguistic Diversity (Speaking English as an Additional Language; EAL)

First language is referred to as “the language to which a child was initially exposed during early development and continues to be exposed to...in the home or in the community” (DfE, 2014c, p.9).

Pupils from homes where English is the first and only language spoken have significantly fewer behavioural problems, while EAL pupils have been found to have increased hyperactivity, and peer and conduct problems (Hansen et al., 2010;
Sammons et al., 2007). Regarding attainment, a greater proportion of pupils whose first language is English achieve the GCSE benchmark, relative to EAL pupils (DfE, 2015b). While this gap is evident across all stages of education, it does decrease over time (Strand, Malmberg, & Hall, 2015); this attenuation with age is also evident with other risk variables, such as season-of-birth, suggesting that continued exposure to a risk variable does not necessarily lead to poorer outcomes (see section 3.3). Furthermore, Strand and colleagues (2015) argue that ethnicity and EAL are closely related, implying that EAL may act as a proxy for minority ethnic status, and that the binary measure of EAL does not allow for differences between ethnic groups.

The infrequency with which English is heard at home is considered to be one reason for differences in academic outcomes. This explains the larger gap during the early years of schooling; as pupils progress through school they are exposed to more English and thus catch up to their peers (Strand et al., 2015). Furthermore, they are “learning in and through another language” and come from backgrounds and communities with different expectations of education (South, 1999, p. 1). In addition, the frustration of struggling with school work may increase behaviour problems (Seifert & O’Keefe, 2001).

Based on the results detailed above, including the effects of potential proxy variables such as ethnicity, it is unclear whether EAL status is significantly associated with school functioning. Therefore, it is included in the current study in an attempt to bring clarity to this area.

2.5.1. vi Looked-After Status

According to the Children Act 1989, a child is defined as “looked-after” (CLA; children looked after) if they are provided with accommodation for more than 24 hours, subject to a care order, or subject to a placement order (DfE, 2015d).
62% of children in residential care have at least one difficulty with school functioning including behaviour problems and low achievement (Attar-schwartz, 2009), and the attainment gap between CLA and non-CLA pupils is persistent throughout schooling (DfE, 2014a). Reasons for CLA pupils’ difficulties in education include instability, withdrawal from school, lack of sufficient help with school work, primary carers being unequipped to provide sufficient support, and unmet emotional, mental and physical health needs (SEU, 2003). Furthermore, looked-after children experience high levels of abuse and neglect (Zayed & Harker, 2015), with those that have experienced this being more likely to have hyperactivity and conduct problems (Minnis & Devine, 2001). As abuse and neglect are also risk variables in their own right (Christiansen et al., 1997; Luntz & Spatz Widom, 1994; OPRE, 2012; Sullivan, 2011), it may be that these children are at an even greater risk of poor school functioning.

Of pupils with SEND statements, looked-after children are more likely to have problems relating to behaviour, emotion and social functioning (DfE, 2017a). As two-thirds of CLA pupils have an SEND, which is also considered to be a risk variable for school functioning (see section 2.5.1.iii), it may be the CLA status, the SEND statement, or the cumulative effect of exposure to both that contributes to pupils’ lower school functioning; the present study adds clarity in this area.

Based on the available evidence, CLA status is an established risk variable for school functioning, particularly regarding academic attainment, and so is included in the current study.

2.5.2 Microsystem – Familial and Neighbourhood-Level Risk Variables

2.5.2.i Socio-Economic Deprivation (Free School Meal (FSM) Eligibility)

Poverty, or low SES, is often represented as FSM eligibility, which is available when parents’ income is below 60% of the national median (Department for Work & Pensions, 2013). Although other proxy measures exist, FSM is still the measure
most frequently utilised. Alternative methods include ever6FSM, a measure of previous poverty. While this has been found to be a better predictor of academic outcomes than current eligibility (Sutherland, Illie, & Vignoles, 2015), the younger pupils in the present study have not been attending school for long enough to utilise this variable.

Pupils’ FSM eligibility and family income are significant predictors of a range of behavioural problems including hyperactivity, CD, aggression, and anti-social behaviour (DfE, 2012; Green et al., 2005; Kellam et al., 1998), with issues evident as early as pre-school (Huaqing Qi & Kaiser, 2003). Similar issues also exist regarding academic attainment throughout schooling (Smith, 2003). Indeed, the DfE has frequently expressed their concerns regarding the social class attainment gap, with the difference in attainment at KS4 at 26.7 percentage points; this is almost three times wider than the gender gap (Strand, 2015).

A number of factors, such as the higher levels of family life stressors and family instability factors, limited social support available to parents, harsh parental discipline, and exposure to violence, are commonly experienced by pupils with behavioural problems from low-income backgrounds (Ackerman et al., 1999; Dawkins, Fullilove, & Dawkins, 1995; Duncan, Brooks-Gunn, & Klebanov, 1994; Huaqing Qi & Kaiser, 2003). Thus, it may be the accumulation of these factors that results in increased behavioural difficulties (Evans, 2004), or they may themselves be proxy risk variables (see section 2.3.1). Regarding attainment, factors such as poor attendance, lower levels of parental involvement, and lower self-esteem have all been linked to FSM eligibility (Smith, 2003). Low-income pupils also experience significantly less cognitive stimulation; parents engage them in fewer literary activities and speak to them less often, and in less sophisticated ways (Evans, 2004). Indeed, the relationships between SES and poor productive vocabulary in two year olds are largely accounted for by shorter utterances of parental speech in lower class families (Hoff, 2003).
Based on the available evidence, socio-economic deprivation is an established risk variable for school functioning, and so will be included in the current study.

2.5.2.ii Neighbourhood Deprivation (Income Deprivation Affecting Children Index; IDACI)

IDACI is a score based on the percentage of children in an area whose families are considered to be income deprived (DfE, 2014b), with higher scores indicated greater neighbourhood deprivation. Thus IDACI is commonly used as a proxy for neighbourhood deprivation (Department for Communities & Local Government, 2015). Whilst the predictive power of IDACI has previously been called into question (Sutherland et al., 2015), investigations have found that associations between neighbourhood deprivation and pupil behaviour remain after adjustment for familial SES (Kalff et al., 2001), as will be done in the present study.

There is currently an attainment gap of 30.9 percentage points at GCSE between pupils from the most and least deprived areas (DfE, 2015a), while White British boys do five percentage points worse if they also live in deprived neighbourhoods (Department for Education & Skills, 2007). However, this pattern is not evident in other ethnic groups, and thus may be influenced by other proxy risk variables associated with ethnicity (see section 2.5.1.iv). Neighbourhood poverty also yields more externalising behaviour problems (Kohen, Oliver, & Pierre, 2009; McCulloch, 2006).

The reasons posited for the differences in outcomes for children living in deprived neighbourhoods include the epidemic or contagion model (whereby behaviours are learned or copied), collective socialisation models (which emphasise the importance of other adult role models in the neighbourhood who help to socialise children), and strain theory (whereby disorder is related to the gap between culturally-induced aspirations for success, and the structurally distributed possibilities of achievement) (Barnes et al., 2006). This means that the fewer opportunities for employment decrease the value placed on education, and hence
cause the rejection of school by children. Pupils living in deprived city areas are also more likely to have parents with low educational qualifications, to have low family incomes, or to live in social housing, which may mediate the effects of neighbourhood deprivation (McCulloch, 2006).

Based on the results detailed above, including the effects of potential proxy variables such as FSM, it is unclear whether IDACI will have a significant association with school functioning. Therefore, it is included in the current study in an attempt to bring clarity to this area.

2.5.3 Mesosystem – School-Level Risk Variables

Whilst there is an abundance of research focusing on risk variables in the child’s microsystem, significantly less has been done to identify the school-level variables in a child’s mesosystem that interact with the child to influence their academic and behavioural outcomes. Furthermore, “only a handful of studies have examined school characteristics and no studies... have analysed student and school factors together” (Theriot, Craun, & Dupper, 2010, p. 14). Fenning and Rose’s research (2007) into the factors affecting exclusion amongst minority groups also recognised the limited studies evaluating school-level variables, arguing that a “shift in attention is needed” (p.542). The current study therefore contributes to the literature by examining the impact of both school-level risk variables, and the combined cumulative effects of school- and pupil-level variables.

2.5.3.i Attendance

Whilst poor attendance and truancy are risk variables in their own right (Theriot et al., 2010), it appears that attending a school where overall absence is high is also associated with individual pupils’ outcomes. One qualitative study found that being in a class with low attendance was perceived to be associated with five key issues: loss of work partners, disruption in class when absentees return, resentment
among good attenders, good attenders being disappointed and confused, and poor attenders becoming role models (Malcolm et al., 2003).

Quantitative research has found school-level attendance to be significantly associated with overall suspension rate and rate of total suspension days (Bruns et al., 2005), suggesting that attending a school with high absence puts pupils at greater risk of being suspended. However, the reasons for suspensions were not cited, and so the specific impact that school-level attendance had on pupils’ school functioning is unknown. It has been suggested (ibid) that the pupils most frequently suspended are also more likely to have learning disabilities, be in foster care, live in poverty, or be from a minority ethnic group. As all of these factors have previously been found to be risk variables for school functioning, they may be influencing the effects of school-level attendance previously identified in the literature.

Based on the inconsistent and sparse results available in the literature, including the effects of potential proxy variables, it is unclear whether school-level attendance is significantly associated with school functioning. Therefore, it is included in the current study in an attempt to bring clarity to this area.

2.5.3.ii Achievement

According to the DfE (2016a), at the end of primary school, pupils are expected to achieve a scaled score of 100 (previously Level 4) in their Key Stage two (KS2) National Curriculum (NC) tests, and so the proportion of pupils attaining this benchmark is frequently used as a measure of school achievement.

Attending a low-achieving school has been found to be associated with higher levels of disruptive behaviour and violence in the classroom (Lavy, Paserman, & Schlosser, 2012; Mooij, 1998), whilst detrimental relationships between low class-level average achievement and the academic performance of regular students have also been identified. It has been suggested that this is due to the high proportion of low-ability pupils negatively impacting on teachers’ pedagogical practices. It is
thought that they divert attention from regular students and that teachers teach “to the median student in their classrooms” (Dobbie & Fryer, 2013, p. 5); thus they adapt their pace to suit the weakest students (Lavy et al., 2012). Other theories emphasise the importance of peer groups and social interactions in a school with high-achieving peers in the formation of skills and values (Dobbie & Fryer, 2013).

Conversely, other studies have found a link between attending a school with high academic results and poorer pupil behaviour (Felson et al., 1994). It has been hypothesised that this is due to the competitive nature of the classroom, and the higher test scores of peers negatively affecting pupils’ self-appraisals of their own academic performance (ibid). However, others have failed to find a causal effect of attending a high-achieving school on longer term academic outcomes (Dobbie & Fryer, 2013).

The somewhat inconsistent results detailed above indicate that it is unclear whether school-level achievement will be significantly associated with school functioning. Therefore, it is included in the current study in an attempt to bring clarity to this area.

2.5.3.iii SEND Pupils

Whilst the impact of inclusion of SEND pupils has been investigated extensively, the majority of the research has focused on the outcomes for pupils with an SEND, thus failing to examine the impact on all pupils in the class (Farrell et al., 2007). However, the studies that have investigated this have found a relationship between attending a school with a high proportion of SEND pupils and a reduction in school functioning. A higher percentage of SEND pupils has been found to be associated with an increase in behaviour disorders at the pupil-level (Barnes et al., 2006), lower overall GCSE scores (Lunt & Norwich, 1999), and poorer reading outcomes (Fletcher, 2009).

It has been suggested that this negative impact on non-SEND pupils is due to the higher levels of disruption that SEND pupils can cause (Fletcher, 2009), and teachers
lowering the general education levels to cater for the SEND pupils, or using ability grouping (Ruijs, Van der Veen, & Peetsma, 2010). There is also thought to be insufficient training available to teachers regarding effective provision for SEND pupils, which may hinder the learning of other pupils (Goodman & Burton, 2010). However, a large proportion of studies have disputed these findings, concluding that being part of an inclusive classroom has no detrimental relationship with students’ outcomes (Farrell et al., 2007; Kalambouka et al., 2007; Ruijs & Peetsma, 2009; Ruijs et al., 2010), or that any detrimental relationships are due to the fact that schools with higher levels of inclusion tend to be schools serving more disadvantaged pupils (Dyson et al., 2004).

The somewhat inconsistent results detailed above indicate that it is unclear whether the proportion of pupils with an SEND is significantly association with school functioning. Therefore, it is included in the current study in an attempt to bring clarity to this area.

2.5.3.iv Deprivation (Average SES)

Similarly to pupil-level measures (see section 2.5.2.i), school-level SES is often calculated based on pupil FSM eligibility, and has been found in the literature to be a risk variable for school functioning, regardless of whether the individual pupil is eligible.

Research suggests that attending a high-poverty school is associated with poorer behaviour and higher levels of school disorder, including verbal and physical aggression (Barnes et al., 2006; Sellström & Bremberg, 2006). Pupils from high-poverty schools also perform worse academically (ibid); school-level poverty has been found to explain 46% of the variation in GCSE achievement (Shuttleworth, 1995). Indeed, it appears that average school-level SES has an impact on pupils’ academic attainment comparable to their own SES, although the magnitude of the effect varies by subject area (Rumberger & Palardy, 2005).
It has been suggested that the effects of school-level SES can be explained by four school characteristics (ibid): amount of homework completed, teacher expectations, number of rigorous courses taken by students, and students’ feelings about safety. Schools with a high proportion of pupils from low-SES backgrounds are also thought to be organised and operate differently than high-SES schools in terms of their policies and practices, which may explain the importance of school-level SES as a predictor of pupils’ outcomes.

Based on the available evidence, school-level SES is an established risk variable for school functioning, and so is included in the current study.

2.5.3.v Linguistic Diversity (Proportion EAL)

Whilst evidence exists to suggest that pupil EAL status may act as a risk variable for poor school functioning (see section 2.5.1.iv), the effects of school-level EAL composition on non-EAL pupils’ outcomes is an extremely neglected area of research (Cho, 2012).

However, one major study (ibid) did investigate this, finding that having EAL classmates was associated with lower test score gains for non-EAL students, particularly in reading. It was suggested that this was due to the inadequate levels of training offered to teachers, and teachers struggling to align the different needs of the class with the curriculum, thus leading to poor educational opportunities for all students (Haworth, 2009; Reeves, 2006). It is also thought that EAL students slow class progression (Schmidt, 2000) and increase teachers’ workload, affecting their ability to provide individualised instruction to students (Karabenick & Noda, 2004). However, even less research has been conducted into the effects on behavioural outcomes, although the literature suggests that high class-level EAL may fracture a classroom environment, creating a negative learning atmosphere, thus fostering problem behaviours. The additional time required by teachers to aid EAL students means less time is spent monitoring the rest of the class, thus allowing more space for disorder to arise (Gottfried, 2014).
Conversely, some studies have found that high-EAL classes have no negative impact on attainment, and that pupils in these classes have fewer externalising behaviours, including higher self-control and improved approaches to learning (ibid; Strand, Malmberg, & Hall, 2015). It is theorised that high numbers of EAL pupils improve interpersonal skills and foster patience, thus increasing self-regulation and control. Non-EAL pupils will also benefit from the educational resources allocated to high-EAL classrooms, while the additional support provided will free up teachers’ time, thus allowing them more opportunities to monitor and assist other pupils (Gottfried, 2014).

Due to the limited amount of evidence available in this area, school-level EAL is investigated as a risk variable in the current study, in an attempt to address this gap in the literature.

2.5.3.vi Behaviour (Proportion of Pupils Borderline or At-Risk for Conduct Problems)

Attending a school with high levels of behavioural difficulties has previously been found to predict poor school functioning amongst pupils, and appears to be sustained over time. Schools with high levels of aggression and temporary exclusions have been shown to have lower school-level achievement across subjects (Barnes et al., 2006). Classroom-level academic focus also appears to significantly predict individual academic focus (Barth et al., 2004), suggesting a possible indirect relationship, whereby poorer academic outcomes result from a lack of focus. Regarding behaviour, classroom-level aggression has consistently been found to increase the likelihood of a pupil displaying persistent aggressive problem behaviours (ibid; Kellam et al., 1998; Thomas & Bierman, 2006; Thomas, Bierman, & Powers, 2011). However, the majority of the research has explored aggression, and so the effects of other kinds of behavioural problems are currently unclear.
It has been suggested that social learning mechanisms including rewards, punishments and modelling may account for some of the relationships between school-level behaviour on students’ outcomes (Barth et al., 2004; Thomas & Bierman, 2006), although this has not been measured and so the underlying mechanisms cannot be ascertained. However, it has been hypothesised that high proportions of pupils displaying aggressive behaviours may create a social milieu normalising these behaviours, making them socially acceptable (ibid). Alternatively, social field theory (Kellam & Anthony, 1998; see section 4.3.3) may explain these findings, as in classrooms where aggressive behaviours are the norm, pupils must conform to these behaviours in order to be rated adequately by their classmates.

Due to the limited amount of evidence available in this area, school-level behaviour is investigated as a risk variable in the current study, in an attempt to address the gap in the literature.

2.5.3.vii Size

Several researchers have hypothesised that smaller school size leads to positive behavioural and academic outcomes for children (Stewart, 2008): “both theory and evidence suggest that large schools are more disorderly than small ones” (Haller, 1992, p. 145). Research consistently shows that attending a large school increases the likelihood of behaviour problems (Bruns et al., 2005; Haller, 1992; Stewart, 2003, 2008), while a number of studies have found a correlation between smaller school size and higher achievement, particularly for poor and minority students (McRobbie, 2001). However, there is evidence to suggest that this is not significant once individual-level predictors are taken into account (Stewart, 2008). This may mean that in the current study, where both individual- and school-level risk variables are investigated, the relationship between school size and pupils’ academic outcomes may not be significant.

It has been suggested that stronger personal bonds such as greater engagement, belonging and personal value, increased parental and community involvement, and
improved instructional quality in small schools may account for the increased levels of school functioning (McRobbie, 2001). Smaller schools are often associated with more personal attention, more opportunities for involvement, less anonymity, and a more caring environment (Stewart, 2008), while larger schools are perceived as more threatening, and students can feel alienated and lonely (Barnes et al., 2006).

Based on the results detailed above, it is unclear whether school size will have a significant association with school functioning in the current sample. Therefore, it is included in the current study in an attempt to bring clarity to this area.

2.5.3.viii Location

School location is often investigated as a risk variable based on whether a school is in a rural or urban area, and has previously been found to be associated with pupils’ outcomes (Hope & Bierman, 1998; Stewart, 2003, 2008). Urbanicity has previously been found to be correlated with misbehaviour, with increases of almost 20% in urban schools (Haller, 1992; Stewart, 2003). However, the correlations between school size and discipline are larger, suggesting that size may be more important; although, rural schools are often also smaller than urban ones (Haller, 1992).

Regarding attainment, less has been done to investigate any impact of school location, although one study found that students from urban schools compared less favourably to their rural school peers on various academic measures (Lippman, Burns, & McArthur, 1996). However, several of these studies are somewhat dated and so whether their findings are still applicable is questionable. More recent research (Stewart, 2003) has indeed contradicted previous findings, suggesting that school location has no association with academic outcomes once individual-level variables are accounted for. Therefore in the present study, other variables may account for the variance in pupils’ outcomes, thus meaning that urbanicity may not be a significant predictor.
It is thought that a school’s urbanicity has a detrimental effect as it tends to reflect larger community-level processes. These schools are often situated in poorer and disorganised communities that experience more violence, and there is a denser population of high-risk pupils (Stewart, 2003). Furthermore, the use of ability tracking in large urban schools can result in the grouping of low-achieving and highly aggressive children, which can exacerbate problems (Hope & Bierman, 1998). It has also been argued that differences in academic achievement between urban and rural schools are due to variation in the individual SES of the students (Coleman et al., 1966).

Due to the limited and inconsistent evidence available in this area, including the effects of potential proxy variables, school size is investigated as a risk variable in the current study, in an attempt to address the gap in the literature.

2.6 Chapter Summary

- Terms relating to risk and risk factors have been inconsistently and imprecisely used in the previous literature; thus the present study utilises the term “risk variable” to incorporate both fixed and malleable markers identified at a single time point.

- Promotive and protective factors, and their relationships with risk variables, play an important role in children’s outcomes; whilst it is beyond the scope of the study to examine both risk and resilience, the role that these protective and promotive factors play is acknowledged.

- 16 risk variables for poor school functioning at the individual- and school-level have been summarised, and a rationale for examining them in the present study has been presented.
3. Cumulative Risk

3.1 Chapter Overview

This chapter aims to introduce cumulative risk theory and highlight the importance of investigating multiple as opposed to single risks when examining children’s outcomes. Previous research in the field is presented to support this. Following this, theoretical explanations for the superior predictive power of cumulative risk are discussed as possible mechanisms through which children’s school functioning is influenced. Inconsistencies in the literature surrounding the functional form of risk-outcome relationships are also explored. Finally, methods of measuring cumulative risk are presented.

3.2 Cumulative Risk Model

Whilst risk variables have frequently been investigated, and a range of research exists supporting the negative effects of exposure to single risk variables, there is also evidence to suggest that risk variables rarely occur in isolation. Instead, they cluster together and thus are not independent of one another (Flouri & Kallis, 2007). Therefore, researching the effects of single risk variables does not allow for the complex and interactional relationships between them that can account for a pupil’s outcomes (Gerard & Buehler, 1999). By not considering the other variables co-existing with a particular risk variable, the importance of that specific variable can be over-estimated (Sameroff, Gutman, & Peck, 2003).

For example, children growing up in poor communities are much more likely to be exposed to multiple, overlapping stressors (Morales & Guerra, 2006): children in low SES households often also live in sub-standard, single-parent, housing in high-crime neighbourhoods, and attend larger, less-prepared schools with less-experienced teachers (Evans et al., 2013). A single risk study would focus on one of these factors in isolation without considering the constellation of both proximal and distal risk variables in the child’s life. In other words, they would bias the estimates
of one variable’s impact by not accounting for the other variables present. This argument has strong links to Bronfenbrenner’s EST (1986), suggesting that all aspects of a child’s environment interact to influence development, both directly and indirectly (see section 1.3.1). Therefore, much of the recent research looking at risk variables has turned to focus instead on multiple risk (Evans et al., 2013), as is the case with the current study.

Evidence from Sameroff and colleagues’ Rochester Longitudinal Study (1987) supports this. Their study investigated the development of a group of children through to adolescence, living in a socially heterogeneous set of family circumstances. A group of 10 risk variables were identified that significantly predicted children’s IQ and social-emotional competence; however, when looked at in isolation, the effects could not be accounted for by any particular subset of the variables. This evidence suggests that no single risk variable is sufficient to account for the variance in outcomes, but rather it is the combination of risk variables that is important (Greenberg et al., 1999; Gutman, Sameroff, & Cole, 2003).

To account for the drawbacks of single risk research, Rutter (1979) developed a model of multiple risk called cumulative risk theory. This has become one of the most common forms of multiple risk models, whereby manifold risk variable exposure is operationalised in an additive manner, without accounting for statistical interactions between the risk variables. Whilst multiple risk encompasses any model with more than one risk variable, cumulative risk defines each variable categorically, with high levels of exposure to a variable necessary for it to be considered a risk variable (Evans et al., 2013). The basic premise of cumulative risk theory centres around the idea that children’s developmental outcomes are better predicted by combinations of risk variables rather than by single risk variables alone (Greenberg et al., 1999). Rutter argued that it was not any particular risk variable, but the accumulation of risk variables that led to psychiatric disorders, with higher cumulative risk leading to greater adjustment difficulties (Appleyard et al., 2005).
Although there are inconsistencies in the literature, cumulative risk theory typically has two main underlying assumptions. Firstly, the greater the number of risk variables, the greater the prevalence of clinical problems; those exposed to more risk variables are at an increased likelihood of problems compared to those exposed to fewer risks (Appleyard et al., 2005; Oldfield et al., 2015). Therefore, the addition of each extra risk variables brings about an increased detrimental effect on the child’s outcomes. Secondly, it is the accumulation of risk variables, rather than the presence or absence of particular risk variables or combinations of them, that impacts on developmental outcomes. In other words, it is the number of risk variables, rather than the type of risk variable, that is important. This is based on the principle of equifinality; that is, the same outcome can accrue from disparate sources, or there are multiple routes to the same outcome (Dodge & Pettit, 2003). This principle distinguishes cumulative risk from multiple risk theory; whilst multiple risk argues that it is the nature of the risks that is important, cumulative risk states it is the number of risks that is critical to children’s development. Cumulative risk theory is utilised in the current study, in order to account for the drawbacks of single risk research. Both assumptions of cumulative risk theory are tested independently, to further understanding of the power of cumulative risk when predicting pupils’ outcomes.

3.3 Theories of Cumulative Risk

Evans, Li and Whipple (2013) argued that one of the primary limitations of cumulative risk was the lack of theoretical explanation for its predictive power. Therefore, they posited three key rationales that could account for the superior predictive power of multiple versus singular risk on pupils’ outcomes. Whilst it is beyond the scope of the study to explicitly investigate these theories, each is acknowledged as a possible contributor to the negative effects of cumulative risk exposure.

Firstly, allostatic load is considered a possible explanation for the predictive power of cumulative risk. This is defined by Evans and colleagues (2007) as “a physiological
marker of cumulative wear and tear on the body caused by the mobilisation of multiple physiological systems in response to environmental demands” (p.341). In other words, it is thought that exposure to risk variables increases the activity of the body’s cardiovascular, immune, neuronal and endocrine systems in order to meet the heightened demands of the environment (Evans, 2003). Over time, the prolonged and repeated exposure to stressors alters the ability of these systems to respond efficiently to environmental demands (Evans et al., 2007, 2013).

A second explanation involves the underlying mediational mechanisms that can account for the negative impacts on children. It is thought that there may be commonality of the fundamental operating mechanisms; with results generally suggest that certain characteristics of the parents, child, the home environment, and broader socioeconomic environment account for some of the covariance between cumulative risk exposure and children’s outcomes. For example, commonly cited in the literature as mediating mechanisms are maternal factors such as education level (Gutman et al., 2003), mental health (Hooper et al., 1998) and parenting style (Gerard & Buehler, 1999).

Finally, Bronfenbrenner’s EST (1986) posits another explanation as to the superior predictive power of cumulative risk. It is argued that children exposed to multiple risk variables in multiple different contexts have a heightened vulnerability compared to those who only experience stress in a single context (Whipple et al., 2010). This is thought to be because a child generally has the capacity to handle a single disruption in one of their microsystems, providing they have the opportunity to cultivate energy from an alternative, uninterrupted source (Evans et al., 2013). However, exposure to multiple risk variables increases the likelihood of disrupting the proximal processes of development as they interfere with the complex exchanges and continuity of energy between the different ecological domains and the child (ibid). Children exposed to cumulative risks within multiple ecological domains are thought to be overwhelmed by the stressors and thus experience poorer outcomes (Whipple et al., 2010).
3.4 Functional Form

A handful of studies looking at the cumulative effect of risk variable exposure on children’s development have also explored its functional form. In other words, they have sought to identify whether the relationship between risk exposure and children’s outcomes is linear or non-linear (quadratic). In non-linear relationships, a threshold or saturation effect is possible. It was originally argued that the cumulative risk model is an additive one, in that it is assumed there are no statistical interactions between risk variables; this would suggest a linear relationship between risk variable exposure and outcomes. However, research has since suggested a different model, whereby risk variables potentiate each other, meaning that their combined effect is more detrimental than a summation of their separate effects (Appleyard et al., 2005; Evans et al., 2013).

Evans and colleagues’ review of the cumulative risk literature (2013) established that an equal number of studies found linear and non-linear relationships with children’s developmental outcomes, and so the form of the relationship is a disputed topic. As a result of this, they explained that outcome function warrants more attention in future research. Therefore, this study investigates the functional form of any cumulative risk relationship that is found, in an attempt to aid clarification in this area. Furthermore, as shown in the sections below, the majority of the studies investigating functional form focus on internalising and externalising difficulties, as opposed to academic outcomes, and rarely investigate both together. The current study thus provides a unique contribution to the field regarding the functional form of cumulative risk relationships.

3.4.1 Linear

The linear model suggests that as the number of risk variables increase, a similar steady increase in problematic outcomes occurs (Appleyard et al., 2005). One of the most frequently cited studies looking at cumulative risk, The Rochester Longitudinal Study, was conducted by Sameroff and colleagues in the 1980’s and 1990’s.
A set of risk variables were identified that were thought to explain the variance in children’s IQ scores over a nine year period. Sameroff found a clear downward linear relationship between risk variables and IQ; as the number of risk variables increased, the intellectual performance of children decreased at four years of age. This linear relationship was still evident at the follow-up study when the children were aged 13. Similarly, Appleyard and colleagues (2005) found evidence of a linear relationship between early risk variables at 12 months of age, and internalising and externalising problems at age 16, with more early risks relating to increased behaviour problems. The occurrence of an additional risk variable increased the odds of problem behaviour, but there did not appear to be a threshold at which outcomes became significantly worse.

3.4.2 Quadratic

A quadratic relationship is evident where there is a disproportionate increase in problem outcomes as the level of cumulative risk increases; this is also known as “mass accumulation”, which suggests that the combined effect that cumulative risk exerts on children’s outcomes is greater than the sum of its parts (Oldfield et al., 2015). Specifically, two models of quadratic relationships are identified in cumulative risk literature. Threshold effects indicate that after a certain number of risk variables there is a dramatic increase in problem outcomes (Appleyard et al., 2005). Alternatively, saturation or sensitisation effects suggest that once a certain number of risk variables is reached, there is a levelling off or plateauing of outcomes, whereby the addition of extra risk variables has no further detrimental effect (Gerard & Buehler, 2004b).

Various researchers have found evidence for the threshold effect in regards to cumulative risk and children’s outcomes. One of the most influential was the work done by Rutter (1979) in his Isle of Wight study who identified six variables significantly correlated with childhood psychiatric disorders. While he found that exposure to zero or one risk variable had little effect on children’s outcomes, the presence of two variables contributed to a four-fold increase in the likelihood of a
mental disorder, and four variables led to a ten-fold increase. These results suggested that after four risks a threshold effect occurred, leading to a dramatic increase in the likelihood of developing a psychiatric disorder. Rutter argued that this was due to the risk variables potentiating one another, causing the cumulated effect to be greater than the sum of each of the risk variables.

Several other studies have found results comparable to Rutter’s. For example, Biederman and colleagues’ research (1995) into risk variables for ADHD found that the odds of receiving a diagnosis were 9.5 times greater for children with two risk variables compared to no risk variables, but 24.6 times greater for children with three risk variables, demonstrating a clear threshold effect. Furthermore, Jones and colleagues (2002) discovered a “trigger point” for internalising problems between three and four risk variables, where the largest increase in problem outcomes occurred. Jones’ explanation for this threshold effect has links with the theory of allostatic load (see section 3.3). She suggested that this trigger point exists because children have adequate coping resources to deal with a couple of risk variables; however, by the fourth, these resources become overwhelmed, and the additional stressors cannot be tolerated.

Alternatively, Morales and Guerra (2006) identified a saturation effect in their study looking at the cumulative effect of exposure to risk variables on children’s adjustment (achievement, depression and aggression). They found that there was a clear linear decline in children’s academic scores until three risk variables were reached, followed by a levelling off between three and five risk variables. A clear linear effect was also evident with aggression until five risk variables were reached, after which the effects of extra risk variables had no additional impact on children’s aggression. A saturation effect is thought to occur as it is assumed that at a certain point the social environment presents too much of a challenge for the child. This means that their development is already compromised, and so the addition of extra stressors is likely to have little further impact (Gerard & Buehler, 2004b).
3.5 Measuring Cumulative Risk

Before cumulative risk can be evaluated, each possible variable first needs to be established as a risk variable. Generally, these are defined as risk variables if there is a significant association with the outcome variable. Each risk variable is then dichotomised by being coded as either “0” for absent or “1” for present. For binary variables, risk is coded as “1” if the risk variable is present and “0” if it is absent; for continuous variables, risk is coded as “1” if the scores fall at or above the 75th percentile (Gerard & Buehler, 2004b; Hebron et al., 2016), in other words, they are in the worst 25% of the sample, whilst everyone else in the sample is coded as “0” (Oldfield et al., 2015). These risk variables are then summed together for each individual to create a composite metric (Evans et al., 2013).

This approach means that no one risk variable is seen as more important than another, and so each is equally weighted; this is in line with cumulative risk theory which posits that it is not the type of risk variable, but the number, that is important. It also supports the notion of mass accumulation, suggesting that the total effect of the variables is greater than the sum of each (Flouri & Kallis, 2007). Furthermore, validity is increased as no single risk variable can adequately explain the variance in the outcomes (Evans et al., 2013; Flouri & Kallis, 2007). Aggregate variables of risk are also thought to be more stable than individual measures, and have an increased power to detect effects; this is because degrees of freedom are preserved, and measurement errors are reduced as the scores are summed (Burchinal et al., 2000).

However, it should be noted that cumulative risk indices do not provide an exhaustive list of risks, nor is there a set of risks identified as the “gold standard” (Lima et al., 2010; Oldfield et al., 2015). In fact, the number of risk indices previously identified in studies have ranged from two to 15 items or more (Lima et al., 2010), although the sample size of the study may affect the cumulative risk score that can be allocated to pupils. Generally, there will be fewer children in the very high cumulative risk groups, which can affect the power of the analysis;
therefore, previous studies have conflated the higher categories (e.g. Appleyard et al., 2005), with pupils with a cumulative risk score higher than five being grouped into one category. This approach is taken in the current study, although it is not without its drawbacks. Whilst this method preserves power, it may also hide any potentially important trends at the higher end of the cumulative risk scale.

It has been argued that a large amount of information is discarded when the risk variables are dichotomised, and the differing impact of the interactions between proximal and distal risk variables cannot be evaluated. This can make interpretation problematic when risk variables are found to be unexpectedly unrelated to distal outcomes. Also, cumulative risk cannot explain the mechanisms by which the risk variables account for the undesirable outcomes (Thompson, 2009). Although this is a potential weakness of the current study, by directly contrasting multiple and cumulative risk models, the researcher seeks to determine whether the cumulative effect of the risk variables accounts for additional variance above and beyond the summation of the individual risk variables, which may help to better explain children’s experiences of negative outcomes.

3.6 Chapter Summary

- Risk variables do not occur in isolation, and they can interact with each other; thus risk research should shift its focus to examine multiple risk variables when investigating pupils’ school functioning.
- There are two assumptions underlying cumulative risk theory, although they are not consistently measured in the literature.
- There is no clear explanation for the superior predictive power of cumulative risk, but three theories have been posited.
- The functional form of the risk-outcome relationship may be linear or quadratic, although further research is required in this area.
- Cumulative risk is distinguished from other multiple risk models as the risk variables are dichotomised, meaning that they are all equally weighted.
4. The Good Behaviour Game

4.1 Chapter Overview

This chapter aims to provide an overview of the GBG, a universal, school-based intervention. An outline of the intervention is provided, followed by a brief discussion of its history, and an examination of the theoretical underpinnings thought to explain the mechanisms through which the intervention triggers change. The logic model produced by the intervention developers is also provided to offer further details of the intervention characteristics. Finally, previous GBG research is explored and critiqued.

4.2 The Origins of the GBG

4.2.1 Overview of the GBG

The GBG is described by Tingstrom, Sterling-Turner and Wilczynski (2006) as an “interdependent group-oriented contingency management procedure” (p. 225). Generally, contingency management strategies reinforce individuals’ behaviour, although the rules apply to every person within the group (Grandy, Madsen, & Merssemen, 1973; Litoe & Pumroy, 1975). However, with interdependent systems, individuals’ reinforcement is dependent upon the performance of the whole group (Bandura, 1969). Interdependent strategies are thought to be easier to manage and more time effective than other types of contingencies; they have also been shown to increase prosocial behaviours such as sharing and co-operation (Skinner, Cashwell, & Dunn, 1996). Unlike other strategies, interdependent systems rely on peer influence; whilst this can have its drawbacks, such as scapegoating and sabotage, it is thought to encourage desired behaviours through social reinforcement (Maag, 1999).

Whilst playing the GBG, pupils in a class are divided into teams of up to seven children; these are typically balanced across genders and are heterogeneous in
behaviour and academic ability. Each team is also allocated a team leader. An individual child may also make up their own team if they have specific behavioural difficulties, or if other problems arise. Teams attempt to win the GBG in order to access certain rewards or privileges; all teams can win the game. To access these rewards, the team needs to have four or fewer infractions at the end of the game, which are recorded using “check-marks”. Whilst they are playing the game, the teacher records any infractions that occur as a result of a team member failing to follow one of the four rules: (1) we will work quietly, (2) we will be polite to others, (3) we will get out of our seats with permission, and (4) we will follow directions (Kellam et al., 2011). For rule 1, voice levels can be set, varying between zero (silence) and four (outside voice); this means that the game can be played during group work, as the voice level can be adjusted to suit the activity (Chan et al., 2012).

A draft logic model was developed for the GBG (see section 4.4; ibid), recommending that initially the game be played three times a week, for ten minutes each time; increasing over the year to every day for up to 30 minutes. The game should also be played at varying points throughout the day, during an assortment of lessons and activities (including both independent and group work). As the game is designed to be a behaviour management strategy, it is intended to be integrated into the existing curriculum without taking up any additional teaching time. Therefore, pupils complete tasks assigned by the teacher whilst playing the game. Over the course of the year, the types of rewards should also change; whilst at the beginning of the year there may be more tangible rewards such as stickers, towards the end of the year the rewards become intangible (e.g. free time). The period of time after the end of the game before the reward is given also extends for longer periods of time as the year progresses (Elswick & Casey, 2011).

Thus, the GBG is built around four core elements, each thought to contribute to pupils’ successful outcomes (AIR, 2014):

(1) Classroom rules – these provide clear expectations to students to encourage them to internalise helpful behaviours.
(2) Team membership – this helps students develop relationships and provides support and reinforcement.

(3) Monitoring behaviour – the teacher monitors infractions and students can practice monitoring each other’s behaviour.

(4) Positive reinforcement – this encourages and strengthens desired behaviours.

The use of the classroom as a social environment, and the influence of peers, are therefore central to the GBG; this aligns with two of the theoretical principles underpinning the GBG: social learning theory (Bandura, 1977), and social field theory (Kellam et al., 1975). Additionally, the use of rewards and punishments (check-marks) means that the GBG is based on a token economy, with roots in operant conditioning (Skinner, 1948) (see section 4.3).

Another fundamental aspect of the GBG is the coaching element. Teachers all receive two days of training for the GBG prior to implementation. During these sessions, teachers learn about the theoretical underpinnings of the game, and the importance of implementing the game with fidelity (see section 5.3). Additionally, teachers receive regular on-going support from a trained GBG coach, comprising up to 16 coaching visits over the course of a year. Coaches conduct observations before feeding back their findings to the teacher in order to increase fidelity. Data gathered from the UK pilot study of the GBG found that coaching was considered to be a necessary part of successful game implementation, with teachers valuing the support and feedback (Chan et al., 2012).

4.2.2 History of the GBG

The GBG was originally developed by Harriet Barrish in America when she noticed that one of her trainee teachers was using a game to help children behave (Embry,
2002); this inspired Barrish, along with her colleagues Saunders and Wolf, to go on to conduct research into the outcomes of the game, and publish their findings (1969). They conducted their study in a fourth-grade classroom of 24 students, recording each out-of-seat and talking-out response by a child. Although there were an additional seven rules in the original version of the game, they were generally extensions of the current four rules used today. Barrish found that the game had a beneficial effect, with out-of-seat and talking-out behaviours reducing by 71% and 77% respectively.

Following the demonstration of internal validity in the early studies of the GBG, in 1985 Kellam and colleagues extended GBG research in Baltimore, America to assess external validity. This longitudinal randomised controlled trial (RCT), along with its associated follow-up studies, is one of the landmark studies of the GBG, focusing on a wide range of outcomes, spanning 13 years and including around 1,000 students in 19 schools. This research was conducted whilst Kellam was affiliated with the American Institutes for Research (AIR). Although he has since retired, AIR have continued to study the outcomes of the game and scaled-up across America and Brazil; the results of Kellam’s research, and AIR’s further studies are discussed in more detail in section 4.5.

Most recently, AIR have expanded their research to the UK, conducting a pilot study in Oxfordshire, England. The results were promising, finding that the game was both feasible and acceptable to teachers. Although an average of five infractions occurred during the game fairly consistently over the course of the year, the number of infractions outside of the game decreased steadily from 20 to five infractions in each class. There were also significant reductions over time on all behaviour subscales (Coombes et al., 2016). The current study further contributes to the findings reported in the Oxford pilot, by providing more robust evidence in a UK setting.

In addition to AIR’s research, Embry and colleagues have attempted to improve the effectiveness of the original GBG, creating a variation known as PAX GBG. Whilst
similar to the original version in that it is a group-based token economy, PAX GBG has additional elements to promote a positive classroom environment (Becker et al., 2013). The pupils create rules and a vision for the class that will promote productivity, peace, health and happiness; this is known as “PAXIS” (Embry, 2002). Behaviours that prevent PAXIS are termed “spleems”; these are similar to the infractions received by pupils in the original version of the game, and so pupils work to prevent spleems occurring for their team. However, the evidence base for PAX GBG is extremely limited.

4.3 Theoretical Underpinnings of the GBG

There are three key theories that the developers of the GBG utilised to explain the mechanisms through which the game triggers change. Although these are not explored in detail in the current study, they are discussed here to aid understanding of how the GBG is thought to influence pupils’ outcomes.

4.3.1 Operant Conditioning

Operant conditioning (Skinner, 1948) involves the control of consequences to promote desirable behaviours and extinguish undesirable ones (McAllister et al., 1969; Richelle, 2003). Generally, reinforcers are used to strengthen desired behaviours, whilst punishments are used to decrease undesirable behaviours (Richelle, 2003). There are several different schedules that dictate when and how frequently reinforcers and punishments are provided. Whilst a continuous schedule sees every response reinforced, partial reinforcement schedules only reinforce behaviours some of the time. These can either be interval (following a certain length of time), or ratio schedules (following a number of correct responses). Interval and ratio schedules can be further broken down into fixed (specified time periods/number of responses), and variable schedules (varying amounts of time/responses (Boeree, 2006).
The GBG uses a variety of reinforcers and schedules during different aspects of the game. Whilst the check-marks for infractions are punishments following a continuous reinforcement schedule, the rewards for those who have won the game are reinforcers based on a fixed interval schedule. Furthermore, over the course of implementation, as the delay between the end of the game and the delivery of the reward extends, the schedule can remain as fixed interval, or may change to variable. Whilst continuous schedules help foster a solid understanding for pupils regarding the relationship between the behaviours and rewards, partial schedules prevent satiation as they only reinforce some of the correct responses. Therefore, they make a learned behaviour steadier and more resistant to extinction. Furthermore, extinction on interval schedules takes longer than on any other schedule, and so the behaviours learned during the game should remain for a longer period of time (Huitt & Hummel, 1997).

4.3.2 Social Learning Theory (SLT)

Whilst operant conditioning argues that behaviour is a result of antecedents and consequences, SLT (Bandura, 1977) focuses on the idea that humans are active information processors who are aware of the relationship between these two events. Bandura described learning as a result of “direct experiences [that] can occur on a vicarious basis through observation of other people’s behaviour and its consequences for them” (1971, p. 2). In other words, learning can occur entirely through observation and does not require direct reinforcement. Instead, individuals use “models”, or people in the environment that they observe and then imitate. The observed consequences of the behaviour affect how likely an individual is to imitate a behaviour (ibid).

Regarding the GBG, it is thought that the students learn appropriate behaviours through group reinforcement. They learn to observe their peers’ behaviour, and to encourage each other, so that the whole group can benefit. Teachers also use social reinforcement during the game by addressing rule-breaking behaviour, identifying the team, and praising the other teams for behaving correctly, thus letting the
pupils know which specific behaviours should be repeated (Ford et al., 2014). Pupils see the models in their class being praised for appropriate behaviour and imitate this, whilst avoiding the inappropriate behaviours that they have observed others being punished for with check-marks. However, there is an implicit assumption within the GBG that all pupils want to behave and to be praised; for certain pupils in the present sample, for example those with SENDs such as oppositional defiant disorder (ODD), who actively refuse to comply with requests (APA, 2013), there may be difficulties with this.

4.3.3 Life Course/Social Field Theory

Life course or social field theory was developed by Kellam and colleagues (1975), who also conducted the Baltimore trial of the GBG. This theory posits that at each stage in life there are several social fields, each with certain task demands. An individual’s ability to meet these demands is rated by other members of the field. The process of acknowledging social task demands and the behavioural responses to them is termed social adaptation, whilst the ratings given for these behaviours are known as social adaptational status (SAS). Successful adaptation is dependent upon an individual’s ability to meet the social task demands. The levels of success in a field can be defined by external dimensions (e.g. pupils’ SAS) or internal dimensions (e.g. psychological wellbeing (PWB)) (Kellam et al., 2008), and the reciprocal relationship between the two (Poduska, Kellam, Wang, Brown, et al., 2008).

In schools, task demands include being able to work well with others, paying attention and abiding by the rules; these are rated formally by pupils’ teachers and informally by peers (e.g. through rejection). All of these demands are encompassed by the GBG. It has previously been found that interventions that are directed at the social adaptational process of task demands in early life can increase SAS and PWB in other fields later in life (Kellam et al., 2008). Therefore, by improving the socialisation of pupils from an early age through the GBG, it is thought that they will learn the skills needed to be successful in the classroom field, which will then be
beneficial to them both externally in other fields in later life (e.g. the workplace), and internally by improving their mental health.

4.3.4 Objections to the Theoretical Underpinnings of the GBG

Although the three theories underpinning the GBG are widely accepted in the literature, they have also been met with some criticisms that should be acknowledged. In particular, it is argued that all three overlook biological processes and individual differences; for example, many individual responses to a stimulus come from the autonomic nervous system and thus are not learned, but partially inherited. Critics have also stated that in Bandura’s studies, the children were teased and manipulated into responding to an aggressive stimulus, deeming it unethical and morally wrong (Jeffery, 1990; Wortman, Lofus, & Weaver, 1998). Similarly, operant conditioning has been criticised due to the passive way it characterises individuals, thus simplifying complex learning processes, and its inability to explain more complex human behaviours such as language development (Seligman, 1970). However, Seligman (ibid) has offered an explanation for this, suggesting that in addition to behavioural conditioning, there is a third principle important in determining an individual’s behaviour: genetic preparedness to associate specific stimuli to responses. Individuals have different genetic tendencies, making conditioned and unconditioned stimuli more or less associable, depending on the preparedness of the individual. Seligman went on to argue that operant conditioning can only explain behaviours resulting from unprepared stimuli.

Finally, regarding social field theory, although the characteristics of the social field can encourage or inhibit adequate responses, each individual has a unique ability to respond appropriately and adapt in order to meet the demands of the field (Petras et al., 2008). This means that whilst the demands of the social field can encourage a desirable behaviour, there are other individual factors influencing the likelihood of these behaviours being adopted. Furthermore, social field theory was developed by Kellam and colleagues, who also developed the GBG, and is largely only referenced
in their studies; this calls into question the validity of the theory, and the theoretical robustness of the GBG.

4.4 GBG Logic Model

A draft logic model was developed by Chan and colleagues (Chan et al., 2012; Coombes et al., 2016) following the GBG pilot in Oxfordshire, England and is shown below. The logic model provides details regarding the key programme characteristics discussed previously (section 4.2.1), the expected outcomes for pupils, and the mechanisms through which these outcomes occur.
### Programme Characteristics

**Training**
Teacher attends initial 2-day training and subsequent 1-day booster session to understand programme theory, basics and application in order to implement with fidelity

**Coaching**
Coach visits teacher every 2-3 weeks for 90 minutes (60-minute lesson observation, 30-minute feedback session) to support quality of implementation over the initial year, and less frequently in second year

**Eligible Classrooms**
Primary (elementary) school classrooms with children in years 2, 3 or 4 (U.S. Grades 1, 2 or 3) and between 15 and 30 children per class. Children are split into equal sized teams balanced for gender, ability and behaviour

### Programme Processes

<table>
<thead>
<tr>
<th>Clear establishment and understanding of four simple class rules, against which children can self-monitor</th>
<th>Reduction of negative and increases in positive interactions with teachers</th>
<th><strong>Into late childhood and early adolescence:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>...along with...</strong></td>
<td><strong>Less disruptive and aggressive behaviour during lesson time</strong></td>
<td>Improve literacy and numeracy</td>
</tr>
<tr>
<td><strong>...leading to...</strong></td>
<td><strong>Consistent on-task behaviour, during the game and increasingly outside the game</strong></td>
<td>Improve curriculum progress</td>
</tr>
<tr>
<td><strong>Behaviour modification and intrinsic reinforcement so that modified behaviour is retained even after external reinforcement is removed (maintenance) and will be exhibited in all settings (generalization)</strong></td>
<td><strong>Increased use of positive reinforcement and monitoring of behaviours</strong></td>
<td>Decreased aggressive and disruptive behaviours</td>
</tr>
<tr>
<td><strong>Increased social awareness</strong></td>
<td><strong>Increased pro-social behaviours</strong></td>
<td>Reduce externalising behaviours</td>
</tr>
<tr>
<td><strong>Better definition and articulation of a appropriate behaviour inside and outside classroom</strong></td>
<td><strong>Increased functional skills such as following directions, independent work and concern for others</strong></td>
<td>Reduce bullying</td>
</tr>
<tr>
<td><strong>Reduce initiation into deviant and antisocial behaviour, and substance use</strong></td>
<td><strong>Reduce teacher stress</strong></td>
<td>Reduce social disengagement</td>
</tr>
<tr>
<td><strong>Provide a net benefit to Society</strong></td>
<td><strong>Lower demand for SEN support</strong></td>
<td>Lower rates of risky sexual behaviours</td>
</tr>
</tbody>
</table>

**Figure 4.1: GBG logic model (Chan et al., 2012)**
4.5 Research Outcomes

Research has been conducted into the effects of the GBG on a variety of pupils’ outcomes, in a whole host of countries worldwide. Outcomes examined in previous research include aggression, conduct problems, anti-social behaviour, suicide ideation, drug use, mental health difficulties and academic attainment (Ialongo et al., 1999; Kellam et al., 2008, 2011; Poduska, Kellam, Wang, Brown, et al., 2008; Reid et al., 1999; Wilcox et al., 2008). However, it is beyond the scope of the current study to investigate all outcomes; thus, only behavioural and academic outcomes are explored, in order to establish the impact of the game specifically on pupils’ school functioning. The data on academic outcomes may also bring clarity to the limited and ambiguous results of previous studies. Furthermore, research into the effects of the GBG in England is limited, with only one pilot study having been conducted (Chan et al., 2012; Coombes et al., 2016). It is not sufficient to rely on the studies from other countries as there are often issues with cultural transferability (see section 4.5.1), meaning that the effects of an exported intervention are reduced. The study’s findings therefore contribute to the evidence base regarding the cultural transferability of the GBG.

4.5.1 Behavioural Outcomes

The majority of the research surrounding the GBG has, unsurprisingly, focused on various behavioural outcomes, which have been broadly categorised into three main areas: low-level behaviours such as fidgeting, conduct problems including ADHD, and aggressive behaviours. Barrish’s study (see section 4.2.2) found that the use of the game significantly improved low-level classroom behaviours such as working quietly and remaining in seats, whilst a large longitudinal dataset from the Netherlands (see section 4.5.3) also produced promising results, suggesting that the game effectively reduced conduct problems for children with intermediate levels of ADHD and ODD.
A large portion of the research, looking specifically at aggression, has come from the early studies in Baltimore conducted by Kellam and colleagues. They found that the GBG significantly reduced aggressive behaviours, in both the short- and long-term, although effect sizes were generally small, and were only significant for certain groups of pupils, particularly males (Dolan et al., 1993; Kellam et al., 1994, 1998). Interestingly, being male is a risk variable for behavioural problems (see section 2.5.1.ii), suggesting that the GBG may be particularly beneficial for at-risk groups. However, in order to confirm this, the effects of the game on a wider variety of risk variables need to be investigated, hence why a cumulative risk approach is adopted in the present study. The differential effects of the GBG found in this early work, along with other studies, are discussed in more detail in section 4.5.3.

It is also noteworthy that in these early studies, a randomised block design was used, meaning that pupils were randomised at the classroom-level; thus even if some pupils were in the control group, they were attending a school where other pupils were receiving either the GBG, or another intervention focusing on aspects of school functioning. School-level variables such as overall behaviour have been shown to be associated with pupil-level outcomes (see section 2.5.3), and research has consistently demonstrated that students in one cluster (e.g. a school) tend to be more similar to each other in terms of an outcome variable than they are to students in another cluster (McCoach & Adelson, 2010). Therefore, it is possible that the pupils in the control groups were indirectly influenced by the interventions through school-level variables, which may explain why some groups of pupils failed to show a significant improvement in outcomes relative to controls.

Regarding conduct problems, only one study conducted in the Netherlands looked specifically at the effects of the GBG on ADHD and conduct problems (see section 4.5.3); although research as part of the Baltimore trial also investigated the effects of the GBG as part of a broader intervention for conduct problems (Ialongo et al., 2001). Pupils who participated in the initial Baltimore trial in first grade were followed up five years later to examine any lasting effects. It was found that pupils
who had received the intervention had significantly lower ratings for conduct problems relative to control pupils, although again effect sizes were small; they were also significantly less likely to meet the diagnostic criteria for CD. However, as these results come from the Baltimore dataset, the limitations described above still apply. Furthermore, the GBG was implemented as part of a “classroom-centred” (CC) intervention that consisted of curriculum enhancements and additional strategies for children not performing adequately, alongside the GBG. Thus, it is difficult to ascertain that the effects observed were due to the GBG specifically, and not one of the other aspects of the intervention, or a combination.

Reid and colleagues (1999) also investigated the effects of the GBG on pupils’ conduct problems, although again, the GBG was implemented as part of a broader ten-week intervention. They conducted a randomised trial of the “Linking the Interests of Families and Teachers” (LIFT) intervention, involving the GBG alongside other social, study, and problem-solving skills workshops, aimed at reducing ODD and CD. Data were collected in 12 elementary schools, involving 671 first and fifth graders. Results suggested that pupils’ behaviour improved following the intervention; effect sizes were generally small, although were larger for the fifth graders. However, the measures utilised were not direct measures of ODD and CD, but instead were thought to measure their antecedents. It therefore cannot be concluded that the intervention had an effect on pupils’ conduct problems, although it may have prevented later conduct problems from developing. In addition, as the GBG was part of a broader intervention, targeting behaviour in the family, classroom and playground, the effects of the GBG specifically could not be identified. The GBG was also modified as part of this intervention so immediate rewards were provided to pupils for positive behaviour, whilst negative behaviours were logged on a chart; pupils then worked towards both individual- and class-level rewards. This makes it difficult to compare the results from this study with other GBG studies, as fidelity to the intervention model would have differed greatly.

Flower and colleagues’ recent meta-analysis (2014) identified 22 peer-reviewed studies of the GBG, finding that it was primarily used for reducing aggression, and
disruptive (off-task, talking-out and out-of-seat) and challenging behaviours, with moderate to large effect sizes. Regarding disruptive behaviour, six of the eight studies investigating this outcome reported positive findings. Although early work with the GBG focused on disruptive behaviours (e.g. Barrish et al., 1969), much of the more recent longitudinal research has utilised aggression or conduct problems as the outcomes of interest (e.g. Kellam et al., 2008) (see logic model – section 4.4), and so current research into the effects of the GBG on disruptive behaviour is lacking. Furthermore, low-level disruptions were highlighted by Ofsted as an area of concern in UK schools (2014), and so disruptive behaviour is the measure utilised in the present study, to address this problem in UK schools specifically, and to provide more current findings regarding the GBG’s impact on lower-level behaviour problems.

Flower and colleagues’ results highlighted some inconsistencies with findings across studies regarding disruptive behaviour. For example, Leflot and colleagues’ (2010) study in Belgium found no significant impact of the GBG on out-of-seat behaviours. This conflicts with other studies that have found the GBG to be either mildly or very effective at reducing these behaviours (Barrish et al., 1969; Kleinman & Saigh, 2011; Saigh & Umar, 1983). However, these inconsistencies may be the result of adaptations that were made to the GBG in the Dutch version of the game (see section 4.5.3); indeed Coombes and colleagues (2016) emphasise the importance of adhering to the logic model specified by the programme developers (see section 4.4). Interestingly, when the effects of the Dutch version of the GBG were investigated in the USA (Lynne et al., 2017) positive results were found for disruptive behaviour, with large effect sizes. However, only one school was involved in the research, and no inferential statistical analyses were conducted; thus a causal relationship between the GBG and pupil behaviour cannot be concluded. Furthermore, a web-based system known as ClassDojo was utilised to record positive behaviours during the game; as this is a popular behaviour management tool with its own evidence base (Robacker, Rivera, & Warren, 2016), it is unclear to what extent this system confounded the results of the study.
These inconsistent findings surrounding the Dutch version of the GBG may also be due to cultural incompatibility of the intervention, as the vast majority of the research has been conducted in the USA. Cultural transferability of school-based interventions is a topic currently debated in the literature; whilst it is thought that some surface-level adaptations are necessary, deeper structural adaptations may compromise the critical components of the intervention, meaning that the intended mechanisms of change are not triggered. Whilst findings for exported interventions are often mixed, a recent meta-analysis of school-based interventions generally evidenced larger effect sizes for interventions implemented in their country of origin (Wigelsworth et al., 2016). This has implications not only for the interpretation of evidence from studies external to the USA, but it also poses a potential barrier to successful outcomes of the GBG in the present study.

However, a pilot study was conducted of the GBG in UK primary schools and found positive effects on pupils’ disruptive behaviour. A pre-post design was utilised, whereby six primary schools participated in the study for one year, producing data from 222 pupils in years 3 and 4. The revised version of the Teacher Observation of Classroom Adaptations (TOCA-R) was utilised as a measure of behaviour; this is the same measure utilised in the present study, although the Checklist version (TOCA-C) is used (see section 6.6.1.i). Results showed significant reductions on all TOCA-R subscales including emotional regulation, hyperactive/impulsive and attention/concentration. Furthermore, qualitative data analysis found decreased disruptive behaviour to be a common theme; in particular, this reduction was associated with observations of improved social skills, better self- and peer-management skills, and better teacher-pupil relationships. However, the lack of a comparison group meant that no firm conclusions could be drawn. The study was also conducted in a small geographical region, and thus may not be representative of the wider population. The current study builds on this pilot by evaluating the effectiveness of the GBG in a UK context utilising an experimental design, and by drawing from a much larger geographical area.
4.5.2 Academic Outcomes

Over the years, several studies have looked at the relationships between behaviour management strategies involving group contingencies and academic outcomes (Evans & Oswalt, 1968; Popkin & Skinner, 2003; Wilson & Williams, 1973); however, the influence of the GBG specifically on academic outcomes has not been so widely examined. Studies that have investigated the effect of the GBG on academic attainment have also utilised a wide variety of outcome measures including scores on reading and mathematics tests, as well as attention, school attendance, and degree receipt; thus making it difficult to identify trends across studies. Furthermore, results of these studies are generally ambiguous due to methodological limitations, discussed in more detail below. Therefore, the current study is investigating the effects of the GBG specifically on pupils’ reading outcomes, in an attempt to address these issues in the current literature base.

In a longitudinal randomised trial involving 658 high-school students (Bradshaw et al., 2009), the GBG was implemented as part of an enhanced academic curriculum. Results indicated that the intervention had a significant positive effect on reading performance and mathematics compared to pupils in the control group, although effect sizes were generally small. There was also a marginally significant effect on teacher-rated academic performance and the number of pupils graduating high school. However, this study utilised a relatively small sample, where the majority of the pupils were receiving FSM, and were of an ethnic minority group. This means that the sample may not necessarily be representative of the wider population. In addition, children in the GBG condition were those with the worst scores of academic readiness at baseline. As FSM eligibility, being part of an ethnic minority group, and previous academic attainment have previously been shown to be risk variables for poorer academic outcomes, it is unclear whether the GBG would benefit all pupils in the same way. Furthermore, as this intervention involved an academic curriculum that already had a proven positive effect on attainment, and the study did not explore the specific mechanisms mediating the change process, it is unclear what effect, if any, the game had on academic outcomes.
In Canada, an adaptation of the game was implemented in 30 schools for one year (Dion et al., 2011). Pupils were randomly allocated to either a peer-tutoring intervention (PTO), peer-tutoring plus the GBG (combined), or a control group. Pre- and post-test test assessments were conducted, finding that whilst PTO had no effect on attention, the combined intervention had a significant effect. Regarding reading, pupils in the combined group had positive and significant linear rates of reading progress, although this also had significant deceleration, suggesting that while progress was initially rapid, it slowed towards the end of the intervention. In addition, neither intervention significantly influenced word and non-word recognition or comprehension on end-of-year reading tests. This study has several limitations; firstly, the study was substantially under-powered to detect even medium effect sizes, and similarly to Bradshaw and colleagues, the game was not implemented on its own, thus making it is difficult to draw conclusions regarding the GBG specifically. The GBG was also adapted to focus explicitly on attention; teachers removed points from a team’s total when a disruption occurred during the game, and added points if the teams had been optimally attentive. This has implications regarding the logic model and theoretical underpinnings of the GBG (see sections 4.3 and 4.4); as key components of the GBG were not adhered to and different aspects of behavioural conditioning were utilised, critical mechanisms of change may not have been triggered. This may mean that similar results will not be replicated with other versions of the game.

Similarly to Dion and colleagues’ findings, Hemelt and colleagues’ results (2013) showed no significant impact of the game being played for two years in primary school on children’s post-secondary attendance and attainment. They also found no significant difference regarding college attendance or degree receipt; although they did find that the GBG appeared to benefit low-performing students more than higher performing pupils, with the intervention increasing the likelihood of them attending a two-year college course by seven percentage points. It also appeared to increase the likelihood of college attendance for girls, with them being 23% more likely to attend a two-year college than the controls. This is incongruent with the
general trend in other GBG studies investigating differential gains whereby boys seem to benefit significantly more. However, as women in the UK are already 35% more likely to attend university than men (Hillman & Robinson, 2016), Hemelt and colleagues’ (2013) findings may not be very surprising.

Hemelt and colleagues’ (ibid) results suggest that the GBG may have differential effects for varying groups of pupils regarding their academic outcomes, and thus does not benefit all pupils equally. However, they obtained data from the National Student Clearinghouse (NSC) to identify pupils that had gone to college. Not all schools participate in the NSC, and institutions covered by the NSC tend to favour private trade schools such as technology and nursing schools; thus the results may not accurately reflect overall college attendance, limiting the generalisability of the study’s results. Additionally, the dataset utilised was taken from Kellam and colleagues’ RCT of the GBG; while this makes attrition less of a problem, this dataset also has several limitations of its own (see section 4.5.1).

Following on from these findings, the current study is investigating the effects of the GBG on reading attainment in an attempt to aid clarity in this area. Whilst other studies have utilised more than one intervention simultaneously, here the GBG is assessed in isolation, thus contributing to the evidence base regarding academic outcomes. By also looking at the outcomes of pupils from multiple risk groups, it may be possible to establish if differential effects on attainment are occurring.

4.5.3 Differential Effects

Whilst studies of the GBG and similar interventions (Hutchings et al., 2013; Jones et al., 2011) show the positive gains on pupils’ outcomes, research into the GBG has also looked at the specific effects of the GBG on pupils exposed to the underlying risk variables associated with low school functioning. For instance, there is substantial research supporting the positive effects of the GBG on the behaviour of pupils from low SES backgrounds, and boys have consistently benefited from the
The initial research on the GBG (Dolan et al., 1993) showed that the game had a significant short-term impact on aggressive and shy behaviours, with boys with the highest aggression ratings at baseline demonstrating the greatest reduction. However, while the GBG did appear to benefit boys who had high initial levels of aggression, it did not seem to protect pupils who were not aggressive from becoming aggressive. In addition, none of the research conducted in Baltimore found any significant outcomes for girls regarding aggression, although in most studies girls already had lower levels prior to implementation.

During later follow up studies (Kellam & Anthony, 1998), boys who had played that game had significantly fewer problem outcomes including antisocial behaviour and violent crime. The GBG also appeared to have differential long-term gains for those with good behaviour at baseline, as they were significantly less likely to start smoking than the controls. Furthermore, the intervention significantly reduced the rates of antisocial personality disorder and violent and criminal behaviour amongst males with persistently high aggressive behaviours (Petras et al., 2008). These findings seem to suggest differential effects of the GBG for certain groups, and thus these are investigated further in the present study in order to better establish if there are any specific gains, either behaviourally or academically, for particular students. The study extends understanding of the differential effects of the GBG by taking an alternative approach to understanding these gains through the use of cumulative risk theory.

Positive effects have also been found for children with a range of SENDs (e.g. Johnson, Turner, & Konarski, 1978; Sy, Gratz, & Donaldson, 2016; Wiskow et al., 2018). For example, most of the research emerging from the Netherlands, utilising one longitudinal dataset has focussed primarily on ADHD (Huizink, van Lier, & Crijnen, 2009; Spilt, Koot, & van Lier, 2013; van Lier et al., 2004). Approximately 700 pupils in 13 schools were involved in this RCT, which lasted for around five years,
with the GBG being implemented for three academic years. It was found that playing the game at age six had a positive effect on the behaviour of pupils with intermediate ADHD and ODD, with a medium effect size. Results were also positive regarding the effects on externalising behaviours and peer relations including acceptance, mutual friends and proximity to others after two years; although the effect was only significant on boys’ behaviour, and the reductions found in externalising behaviours were mediated by peer relations (Witvliet et al., 2009). Whilst this study suggests differential effects for male pupils and pupils with ADHD, the GBG was adapted to be positively formulated for use in Dutch schools. This has similar implications to Dion and colleagues’ (2011) Canadian adaptation of the GBG discussed in section 4.5.2; as the logic model was not adhered to, it is unclear if these results will be replicated in the UK trial, where the GBG is implemented in its original format.

Much of the research examining pupils’ behavioural outcomes has often utilised aggression as a measure of behaviour, or focussed on pupils who have high levels of aggression at baseline, with these pupils deemed to be “at-risk”. Whilst current research largely suggests that at-risk pupils will benefit from the intervention, they generally use the term “at-risk” as a proxy for “highly aggressive”. This implies that risk is binary, when evidence suggests that it is relative, as implied by cumulative risk theory (e.g. Evans & Kim, 2007). However, the impact of the game on pupils at varying levels of risk has not been researched, and so the knowledge base is limited. As evidence exists to suggest that pupils at higher levels of risk experience significantly poorer outcomes (see section 3.2), it is thus important that this is explored in more detail. As Spilt and colleagues commented, “no research has yet examined the effectiveness of the GBG in relation to baseline risk profiles reflecting different constellations of risks across developmental domains” (2013, p. 480).

The current study addresses this issue by adopting a cumulative risk approach when examining the outcomes of the GBG. The results may enable a clearer understanding of who benefits most from the intervention.
4.6 Chapter Summary

- The GBG is a highly prescriptive, manualised intervention with roots in operant conditioning, SLT and life course/social field theory.
- The GBG has an evidence base spanning 40 years, and has recently been successfully introduced in UK schools.
- The GBG has been found to evidence improvements in pupils' disruptive and aggressive behaviours and conduct problems.
- The GBG has been found to improve academic outcomes, but only when implemented in conjunction with another curriculum-based intervention.
- There are findings to suggest differential gains for certain groups of pupils; however, more research is needed to determine the effects on pupils at varying levels of risk.
5. Implementation

5.1 Chapter Overview

This chapter aims to provide an overview of the field of implementation science, emphasising the importance of examining implementation when evaluating the effectiveness of an intervention. The different aspects of implementation are discussed, as are the factors affecting the implementation of an intervention. Various ways of measuring implementation exist, and a discussion of these is provided. Finally, previous research linking implementation and outcomes is explored and critiqued. A particular focus is placed on the GBG and other universal, school-based interventions.

5.2 The Importance of Implementation

Implementation is defined as “the process by which an intervention is put into place” (Lendrum & Humphrey, 2012, p. 635), or “what a program consists of when it is delivered in a particular setting” (Durlak & DuPre, 2008, p. 329). Derzon and colleagues (2005) argue that the implementation process will inevitably shape the quality of pupils’ experiences of an intervention, thus suggesting that the way an intervention is implemented can affect pupil outcomes. However, although the importance of implementation is emphasised in the literature, it is an extremely neglected area, with many prevention research studies failing to include aspects of implementation in their evaluation (Berkel et al., 2011; Domitrovich et al., 2010).

A main concern with the evaluation of interventions is the “black box” approach; this focuses on whether or not an intervention worked, without looking at what happened within the programme in order to establish how or why outcomes were affected (Harachi et al., 1999). Domitrovich and Greenberg (2000) supported this argument, suggesting that a key failing in many preventive intervention studies is that investigators assess programme outcomes while failing to examine most, or
any, aspects of implementation. In an attempt to rectify this issue, they went on to provide five rationales for conducting implementation research, outlined below.

Firstly, implementation data makes it possible to know what truly happened during an intervention, including what actually took place and any adaptations that were made (Dusenbury et al., 2003; Lendrum & Humphrey, 2012). This information can help to explain variations in outcomes and also means that critical programme components can be identified, thus increasing understanding of how they work and interact. Therefore, programme designers can be fully informed of the different components that contribute to the achievement of outcomes, and their relative significance (Lendrum & Humphrey, 2012). Secondly, collecting data linking implementation quality to programme outcomes means that the internal validity of an intervention is enhanced. This is essential for protecting against “Type III error”, that is “the inaccurate attribution of the cause of results” (Lendrum & Humphrey, 2012, p. 639) due to the intervention being delivered so poorly that the results of the analysis are invalidated (Domitrovich & Greenberg, 2000). Implementation data can thus determine whether an intervention failed due to poor programme design, or to poor implementation of a well-designed programme (Askell-Williams et al., 2012). Confirming that the key programme components and processes have been implemented means that links can be made between the achieved outcomes and the intervention (Lendrum & Humphrey, 2012).

Thirdly, implementation data aids understanding of the internal dynamics and operations of an intervention, otherwise known as a process evaluation (Greenberg et al., 2005). This type of evaluation explains how different aspects of an intervention fit together, and how implementers tackle and overcome obstacles. This allows for the examination of the factors affecting implementation, including both barriers to and facilitators of implementation. This knowledge can help to improve an intervention before it is more widely disseminated, hence improving the possibility of positive outcomes (Lendrum & Humphrey, 2012). Fourthly, implementation data can act as a source of ongoing feedback that can be useful for continuous quality improvement. If the implementation of an intervention is
monitored, problems will reveal themselves earlier, before outcomes are adversely affected, and appropriate adjustments can be made. Through the measurement of implementation in an evaluation, the amount of variance, or “noise”, unaccounted for in the design can be reduced (Greenberg et al., 2005).

Finally, studying implementation can help to advance knowledge on the best practices for replicating, maintaining, and diffusing programmes in more complex, “real world” systems. In the “real world”, interventions will be transferred and are frequently adapted across cultures and settings (Ringwalt et al., 2003). Therefore, documenting changes in the implementation process is necessary to interpret the significance of various elements of the intervention, and to understand the impact of any adaptations (Greenberg et al., 2005). The focus may also shift to look at the sustainability of an intervention, and how local and contextual factors affect implementation quality (Lendrum & Humphrey, 2012).

Based on the evidence discussed, it is clear that it is important that aspects of implementation are also examined when investigating the impact of an intervention. Therefore, in the present study, three aspects of implementation are assessed when evaluating the effectiveness of the GBG, in order to gain a clearer picture of how and why it influenced pupils’ outcomes. It also means that the cause of any non-significant findings can be attributed correctly, thus avoiding Type III error.

5.3 Aspects of Implementation

Durlak and DuPre’s review (2008) of approximately 600 interventions reported eight of the most commonly assessed aspects of implementation (see table 5.1). While some suggest that for a comprehensive evaluation to take place, all eight need to be examined in order to avoid Type III error (Dane & Schneider, 1998; Humphrey et al., 2016), Durlak and DuPre argue that although they are related, they can be separated for study.
Table 5.1: Aspects of implementation

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fidelity</td>
<td>The extent to which implementers adhere to the originally intended model/guidance – also commonly referred to as adherence, compliance and integrity</td>
</tr>
<tr>
<td>Dosage</td>
<td>How much of the original intended programme has been delivered and/or received</td>
</tr>
<tr>
<td>Quality</td>
<td>How well different components of the programme have been conducted/delivered</td>
</tr>
<tr>
<td>Participant Responsiveness</td>
<td>The degree to which the intervention stimulates and holds the interest of the participants, or the degree to which participants engage with the programme</td>
</tr>
<tr>
<td>Programme Differentiation</td>
<td>The extent to which the intervention’s practices or activities can be distinguished from other programmes, or from usual practice</td>
</tr>
<tr>
<td>Monitoring of Control/Comparison Conditions (in an RCT)</td>
<td>Determination of what is taking place in the usual practice group</td>
</tr>
<tr>
<td>Programme Reach</td>
<td>The rate and scope of participation</td>
</tr>
<tr>
<td>Adaptation</td>
<td>The extent and nature of changes made to the programme</td>
</tr>
</tbody>
</table>

Adapted from Durlak & DuPre (2008) and Humphrey et al (2016)

O’Donnell (2008) reported that the first five aspects listed were the most frequently utilised as measures of implementation, and could be divided into two groups: fidelity to structure (i.e. fidelity, dosage) and fidelity to process (i.e. quality, programme differentiation), whilst participant responsiveness took on characteristics of both. Durlak and DuPre (2008) agreed, stating that the latter three aspects of implementation would not be evaluated in their review, as they had not received as much research attention. Indeed, the measurement of implementation of school-based programmes has predominately focused on the
number of lessons delivered, the extent to which key aspects are delivered as intended, and the quality of the delivery (Hirschstein et al., 2007). However, whilst it is more commonly measured than other aspects of implementation, still only 10% of studies investigating implementation have measured quality (Durlak & DuPre, 2008). Domitrovich and Greenberg’s review (2000) found that fidelity and dosage were the two aspects of implementation monitored most often, with 59% of studies measuring fidelity, and 33% reporting dosage; although they suggested that “greater attention must be given to both the measurement of dosage and the quality and fidelity of intervention delivery, especially as empirically validated prevention programmes begin to ‘go to scale’” (p.194). For these reasons, each of these three aspects of implementation are investigated in the present study. Although ideally all eight aspects of implementation would be assessed to avoid Type III error, it is unfortunately beyond the scope of the study to do this. Due to the arguments discussed above, fidelity, quality and dosage have been chosen as the aspects that require the most attention.

5.3.1 Fidelity, Quality and Dosage

“Fidelity” is a term used ambiguously in the literature; whilst it is generally accepted that fidelity refers to the extent to which an intervention is delivered as intended by the developers, it is often operationalised differently, and thus two conceptualisations exist. Whilst some hold a “macro” view of fidelity, using fidelity as a term to describe all eight aspects of implementation, others view fidelity in the “micro” sense, whereby “implementation” is utilised as the overarching term, and “fidelity” refers to adherence to protocol. Thus, fidelity and adherence are terms often used interchangeably, depending on the researcher, although they can hold different definitions (Humphrey, Lendrum, et al., 2016). For the purposes of this study, the micro view of fidelity is adopted, referring to the extent that implementers adhere to intervention protocol. Specifically, fidelity refers to the extent to which teachers implement the GBG in accordance with the procedures outlined in the manual. Fidelity is arguably the aspect of implementation investigated most frequently in the literature (Durlak & DuPre, 2008), and
particularly in GBG research, possibly due to the highly prescriptive nature of the intervention. Therefore, it is important that fidelity is explored in the current study. Research regarding the fidelity of GBG implementation is discussed more detail in section 5.6.

Durlak and DuPre (2008) reported that whilst the majority of studies investigating implementation measured fidelity, only half measured dosage. Dosage refers to the amount of an intervention that has been implemented; with a classroom-based intervention such as the GBG, this would include the number of specific “units”, including the number of games per week and durations of each game. For reasons such as this, dosage is considered to be one of the easiest measures of implementation to quantify (Domitrovich et al., 2008) and may be why dosage is more commonly reported in GBG research, relative to other aspects of implementation, although it is still infrequent and is rarely linked to pupils’ outcomes. As the amount of dosage required is outlined in the GBG logic model (see section 4.4), it is important that dosage is measured when evaluating the effects of the game, as reduced dosage may mean that the assumed mechanisms of change are not triggered. However, in implementation literature dosage is often described relative to the amount of units outlined in the manual, and thus is reported as a proportion of the total number expected to be delivered (Humphrey, Lendrum, et al., 2016). This complicates the measurement of dosage in the GBG as the duration and frequency of games are expected to change over time. Nevertheless, as higher levels of implementation are often associated with better outcomes when fidelity and dosage are examined together (Durlak & DuPre, 2008), it is important that they are both included in the current study.

Whilst dosage and fidelity provide indications of what and how much was delivered as intended, quality refers to how well components of the intervention were implemented, meaning that the emphasis shifts from structure to process (Humphrey, Lendrum, et al., 2016). Quality therefore includes interactive, reciprocal techniques such as effective engagement of and responsiveness to pupils in order to guide students in gaining skills (Domitrovich et al., 2008; Dusenbury et
Dusenbury and colleagues (ibid) argue that these interventions require more than simply reading from a script; instead teachers are relied upon to act as facilitators. Therefore, quality is closely related to the professional expertise of the teacher, meaning that they should be enthusiastic, prepared and sensitive to the goals of the intervention (Askell-Williams et al., 2012). Thus, teacher attitudes regarding the perceived need for and benefit of the GBG, and the level of support afforded to them, may have an impact on the quality of GBG implementation. However, as it is extremely neglected in the GBG evidence base, by measuring levels of quality in the current study, a clearer picture can be built regarding its different dimensions and its importance to the outcomes of the GBG. Furthermore, if pupils are effectively engaged by teachers, other aspects of implementation such as participant responsiveness and programme reach may be affected, particularly if the teacher is invested in the intervention and its aims.

Knowing what was done (fidelity), how often (dosage) and how well (quality) is the foundation for any analysis of implementation. Berkel and colleagues (2011) deemed these three aspects to be the dimensions determined by programme implementers, and so they serve as potential sources of variability from the programme as designed; each have also individually been found to influence outcomes. Although it is unlikely that they will be unrelated, evaluations of interventions rarely examine more than one aspect of implementation, thus not allowing for possible relations between them (ibid). Therefore, fidelity, dosage and quality are investigated jointly in the current study to address for this gap in the literature.

5.4 Factors Affecting Implementation

A multitude of factors can interact with an intervention to bring about variability in implementation (Lendrum & Humphrey, 2012). Combinations of these factors can also interact, influencing the outcomes of an intervention, and its future viability (Domitrovich & Greenberg, 2000). Durlak and DuPre’s systematic review (2008) identified 23 factors in five categories that affected implementation, whilst both
Greenberg and colleagues (2005) and Humphrey and colleagues (2016) also reported five similar groups of factors. However, whilst there was some overlap with Durlak and DuPre’s categories, they differed in that they related to schools-based interventions specifically, and are thus outlined below.

Firstly, pre-planning and foundations in place prior to intervention delivery (e.g. awareness of need, capacity for change) can affect how well it is delivered (Humphrey, Lendrum, et al., 2016). Secondly, factors include the implementation support system, such as the provision of training, the quality of the programme materials (e.g. manuals), and on-going support (e.g. coaching) (ibid; Greenberg et al., 2005). Thirdly, the implementation environment can play an important role, including factors at the classroom-level (e.g. classroom climate) and school-level (e.g. school ethos), as well as the wider community and local authority levels (e.g. school location, current policy) (Mihalic, Fagan, & Argamaso, 2008; Humphrey, Lendrum, et al., 2016). Fourthly, implementer factors such as their professional and psychological characteristics, and perceptions of the intervention can have a powerful influence on intervention delivery (Humphrey et al., 2016). Finally, characteristics of an intervention such as form (e.g. universal, targeted, indicated), function (e.g. environmental, developmental), level and location (e.g. individual, school, community), complexity (e.g. single or multi-component), prescriptiveness (e.g. manualised), resources (required and provided) and procedures (e.g. what is done, how often) can all cause implementation to vary (Humphrey et al., 2016).

Regarding the GBG, and specifically implementation in the UK, *challenges to implementation* was identified as a key theme in the pilot study conducted in Oxfordshire (Chan et al., 2012; Coombes et al., 2016). Sub-themes included difficulties incorporating the GBG into teaching practices, time constraints, and clashes with other aspects of teaching. These factors reflected barriers to implementation such as teachers’ professional characteristics, and their knowledge and attitudes towards the need for, and benefits of, the game. While it is beyond the scope of the present study to explore all of these factors in detail, it is important that they are acknowledged here, and are considered when interpreting...
the quantitative results. These factors can influence not only whether the GBG is found to be successful in improving pupils’ outcomes, but are also important to take into account when investigating differential outcomes as a function of implementation variability.

5.5 Measuring Implementation

There is currently no widely accepted, standardised methodology for measuring aspects of implementation. Instead, most measures are tailored to the evaluation of a specific programme, as each intervention is unique in terms of the required processes (Dusenbury et al., 2003). However, it is argued that current measures of fidelity are weak, and thus there is a great need for valid instruments that operationalise and standardise the measurement of programme implementation; although the challenge of developing measures that can be used to assess interventions that differ greatly is considered to be one of the biggest obstacles to a standardised approach (Dusenbury et al., 2003; Ogden & Fixsen, 2014).

Methods utilised to collect implementation data vary from quantitative self-report surveys or questionnaires to qualitative interviews with the implementer, but can also include researcher observations and analysis of documents and artefacts associated with the intervention. Methods can also vary depending on the aspects of implementation being examined. For example, whilst fidelity is often assessed using highly structured questionnaires or observation schedules, the recording of adaptations can be more open-ended in nature to allow for unanticipated factors to emerge (O’Donnell, 2008). This may explain why there has been a 13% surge in the use of mixed methods for the evaluation of the implementation of interventions in recent years (Humphrey, Lendrum, et al., 2016). In addition, studies often use intervention-specific artefacts as a measure of student participation; such as Songer and Gotwal (2005), who used the percentage of students completing the worksheets of a curriculum-based science intervention as a measure of participant responsiveness. In an analysis of the GBG for instance, records of infractions could be used as an intervention-specific measure. However, it could be argued that this
is not truly measuring participant responsiveness, as it does not guarantee that pupils are truly engaged with the intervention.

Another important consideration when measuring implementation is the rater/respondent. In some cases it is the teacher or implementer who is relied upon to provide the implementation data. For example, Domitrovich and colleagues argued that “it is difficult to envision how anyone but the classroom teacher can provide an estimate of intervention dosage” (2010, p. 295); thus evaluations are often dependent on teachers’ weekly logs to establish how often the intervention was delivered. Fortunately with the present study, teachers utilised online scoreboards when playing the GBG, meaning that dosage data could be extracted without any additional input from the teacher. However, self-report can make collecting data on certain aspects of implementation more difficult, as it can often require them to collect data whilst attempting to implement the intervention, and teachers will not have a frame of reference. For instance, in an evaluation of the US HeadStart intervention, the use of self-report led teacher-rated participant responsiveness to be so near the ceiling of the rating scales that no variation between classrooms could be established (ibid). Social desirability and fear of a negative evaluation are frequently cited as issues with using teacher reports as measures of implementation, which is often why ratings by independent researchers are also utilised (ibid; Dusenbury et al., 2003; O’Donnell, 2008).

In fact, when self-reports are simultaneously used with researcher observations, it is often found that the independent observations are greater predictors of outcomes than teacher reports, increasing rigour and reliability (Domitrovich et al., 2010; Lillehoj, Griffin, & Spoth, 2004). Therefore, for the purposes of the current study, an independent researcher rated fidelity and quality. However, this method also has its own limitations; teachers who are aware that they are being observed may adapt their behaviour accordingly, and one observation only provides a single “snapshot” of implementation at a specific time on one day, thus one-off observations may not be representative of day-to-day implementation (Humphrey, Lendrum, et al., 2016).
5.6 Linking Implementation and Outcomes

As discussed in section 5.2, measuring implementation is vital when reporting the results of a school-based intervention, in order to avoid Type III error. However, many studies fail to do this; indeed, Dusenbury and colleagues (2003) found that only a third of studies in their review considered the impact of fidelity on outcomes, whilst Askell-Williams and colleagues (2012) suggested that identifying relationships between processes of implementation and outcomes needed an increased research focus. They argued that in order to be able to establish whether an intervention had an effect on student outcomes, both outcomes and processes need to be evaluated, as they are interrelated. Berman and McLaughlin’s earlier work (1976) supports this, demonstrating the long-standing nature of this issue, maintaining that the link between a promising idea and its impact on students is its implementation.

Information about implementation has been deemed critical to a robust evaluation, particularly for the interpretation of impact data in RCTs; however, most school-based interventions developed in the UK do not pass through the processes of linking implementation and outcomes (Lendrum & Humphrey, 2012). The present study addresses this issue by investigating levels of implementation needed to obtain desired programme outcomes, and by identifying which aspects of implementation are related to which outcomes, both of which were areas highlighted as a priority for future research by Durlak (2015). Furthermore, the study is amongst the first to attempt to identify differential effects of implementation on the outcomes for pupils at different levels of risk. This provides a vital contribution to the fields of implementation and prevention science.

Of the relatively few studies that have attempted to link implementation and student outcomes, they have found that implementation does play an important role. For example, O’Donnell’s review of the literature (2008) found that five of the 23 studies identified investigated the relationship between fidelity and outcomes, and they all consistently showed statistically significant higher outcomes when the
intervention was implemented with greater fidelity. Based on these findings, O’Donnell argued that the considerable variability often reported within the treatment condition raises concerns that this may be accounted for by differential implementation. However, as is typically the case in the field of implementation science, fidelity was the only aspect of implementation measured (Domitrovich & Greenberg, 2000).

However, an evaluation of the PATHS social and emotional learning curriculum in the USA (CPPRG, 1999) did investigate dosage and quality, finding that higher dosage was related to lower rating of aggression. Although the effects of the intervention on high-risk children were examined, a binary approach was taken to risk, and implementation was not linked to these pupils’ outcomes; thus it cannot be ascertained whether PATHS had differential effects for pupils at varying levels of risk. However, another evaluation in the UK found that only higher levels of quality were related to reductions in externalising problems; contrary to expectations, higher dosage was associated with reduced pro-social behaviour (Humphrey, Barlow, & Lendrum, 2017). Humphrey and colleagues (ibid) hypothesised that this may be because high levels of dosage were at the expense of quality, with teachers implementing quickly as opposed to well. This highlights the interrelated nature of the different aspects of implementation, and emphasises the need for multiple aspects to be investigated together. This may also be a cultural transferability issue, as the latter study was an efficacy trial of PATHS in the UK; thus it is important that the implementation of imported interventions is tested before they are scaled-up. Therefore, the present study is contributing to the evidence base in this field, by testing the cultural transferability of the GBG, exploring multiple aspects of implementation outside of its country of origin.

Another piece of research (Askell-Williams et al., 2012) involved creating an implementation index based on fidelity, quality and dosage to investigate differential effects of the KidsMatter mental health initiative. Results showed that social and emotional competencies only significantly improved in average- and high-implementing schools, although once again the differential effects of
implementation on specific groups of pupils were not investigated. However, one study that did investigate the relationship between implementation and outcomes for children with multiple risk variables was the Seattle Social Development Project (Abbott et al., 1998). It was found that the intervention had a greater association with pupils’ behavioural outcomes when implementation was high, suggesting that high implementation is important for high-risk children. However, academic outcomes were not significant, and the differential effects between low and high-risk children were not examined; thus it is unclear whether high-risk students experienced significantly different gains. Furthermore, “high implementation” was used as a proxy for “high fidelity” and so other aspects of implementation that could have influenced the results were neglected.

The results from research measuring implementation discussed above provide evidence that implementation variables account for differential outcomes, “although seldom are measures of implementation used to adjust for or interpret outcome measures” (O’Donnell, 2008, p. 34). Furthermore, there is extremely limited research available regarding the differential effects of implementation amongst pupils at-risk. The present study addresses this gap in the literature, by being amongst the first to examine differential gains of an intervention among pupils at varying levels of risk, as a function of implementation fidelity, quality and dosage.

5.6.1 Linking Implementation and the GBG

Whilst Flower’s meta-analysis (2014) identified eight GBG studies that reported fidelity, only three of these looked at the relationship between fidelity and student outcomes. The largest of these was conducted by Ialongo and colleagues (1999). There were significant main effects relating to teacher-rated problem behaviours for children in the high-fidelity class compared to controls, whilst the difference between low-fidelity and control classes was not significant. Furthermore, for boys, implementation level was associated with fewer nominations of aggressive behaviour. Thus, these results suggest that fidelity can act as a moderator for the
outcomes of the GBG, and can have differential effects for certain groups of pupils. However, as this study looked at the GBG as part of a CC strategy involving another universal intervention, it is unclear how much of a role the other intervention had in influencing pupils’ outcomes.

Lannie and McCurdy (2007) also investigated the effects of fidelity on disruptive behaviour, although, similarly to Ialongo and colleagues, they failed to explore other aspects of implementation. After three sessions of the GBG, an observer provided the teacher with feedback based on a fidelity checklist. During the games where fidelity was assessed and feedback provided, there was a small increase in on-task behaviour (3.08%). When feedback was removed, on-task behaviour continued to rise, whilst disruptive behaviour fell. It is unclear from these results whether high fidelity was maintained during the removal of feedback, and so it cannot be determined that high fidelity was the cause of the continued improvement in behaviour. It could instead be due to the fact that the pupils had been previously exposed to the game and so were more familiar with the rules, therefore finding it easier to adhere to them. The present study may aid clarification regarding the importance of fidelity, as the GBG is investigated in isolation, and implementation is monitored throughout the course of the trial. The need for this has been emphasised by Donaldson and colleagues (2011), as they recommended that “one area for future research could involve systematically evaluating the effects of changes in treatment integrity on the effectiveness of the GBG” (p.607).

Conversely, previous studies involving the GBG have been adapted several times for canteens at lunchtime (McCurdy, Lannie, & Barnabas, 2009), the library (Fishbein & Wasik, 1981), special needs classrooms (Johnson, Turner, & Konarski, 1978), to incorporate technology-based reward systems (Elswick et al., 2016; Lynne et al., 2017), and for Dutch classrooms where the game is positively formulated (Leflot et al., 2013). These interventions have all been considered successful in that behavioural problems were reduced, even though fidelity to the manual was low (e.g. different rules, removal of points following an infraction, addition of points...
following adherence to the rules, the removal of teams). This conflicts with the aforementioned studies that have found fidelity to be an important moderator for outcomes, and suggests that there may be more to effective implementation than high fidelity.

Thus, the type of adaptations made during implementation may have an influence on outcomes, with some contributing positively, and some negatively, to outcomes. Whereas minor surface level adaptations may help the intervention be a better fit for the target audience, deep, substantial modifications, such as the removal of core components, may mean that key mechanisms of change are not triggered (Humphrey, Lendrum, et al., 2016). There are considered to be three dimensions of adaptations (Moore, Bumbarger, & Cooper, 2013): fit (e.g. logical or philosophical reasons including accommodating constraints or appropriateness for the audience), timing (proactive or reactive, i.e. in advance of, or response to, problems), or valence (positive, neutral or negative, i.e. in keeping with, or deviating from, the goals of the intervention) (Humphrey, Lendrum, et al., 2016). However, the types of adaptations that are successful, and why this is the case, have not been previously explored in the GBG literature and so it is unclear which types of adaptations can still be used to produce successful outcomes, even in the context of low fidelity. Whilst it is beyond the scope of the present study to investigate adaptations in depth, more research needs to be conducted in this area in the future.

Regarding other aspects of implementation, Hagermoser Sanetti and Fallon (2011) identified 27 additional steps to increase the quality of implementation of the PAX GBG that are not required for adherence, but are thought to make for more comprehensive and effective implementation. Only six of these steps were implemented more than 80% of the time, whilst 19 components were implemented less than half the time. Hagermoser Sanetti and Fallon used these findings to argue that implementation is multi-faceted and has many dimensions that need to be reported to build a more comprehensive understanding; whilst evidence exists surrounding the relationship between high fidelity and student outcomes, other aspects of implementation are often neglected. Becker and colleagues (2013) agree,
arguing that research linking the quality of program fidelity and dosage with student outcomes demonstrates the importance of implementation quality for the success of preventive interventions; this suggests that various aspects of implementation act as important moderators in similar interventions. Indeed, studies of other universal interventions (Abbott et al., 1998; Kam et al., 2003) have found that high levels of implementation quality result in greater reductions in externalising behaviours, as well as an increase in pro-social student involvement, reinforcement, and student bonding to school. However, this type of evidence is lacking in GBG research; while a couple of other GBG studies mention measures of quality and dosage, and report their findings (Pas et al., 2015; Tanol, 2010), they do not go on to assess the effects of these factors on student outcomes.

For example, regarding dosage, GBG studies have previously found that students were only exposed to the recommended number of games a week 56% of the time, although 94% of teachers implemented the game for the correct length of time (Hagermoser Sanetti & Fallon, 2011). Furthermore, Pas and colleagues (2015) reported that teachers who had more contact with their coaches were more likely to implement a high number of games for a high number of minutes. However, benchmarks were not established regarding high and low dosage in either study, and the effects of the varying dosage on pupils’ outcomes were not investigated. As Becker and colleagues (2013) explained in their review of the GBG, “it is unknown whether a certain dosage or level of quality is necessary or sufficient to bring about student gains...as is typical of most interventions, these benchmarks have not been empirically validated” (p.221). By being amongst the first to explicitly establish implementation benchmarks, and to examine the differential gains of an intervention’s fidelity, quality and dosage, the present study advances current research in this field, as well as providing much needed clarity in this area.
5.7 Chapter Summary

- It is important to examine aspects of implementation when evaluating the effectiveness of an intervention, in order to avoid Type III error. However, many studies still fail to do this adequately.
- While eight different aspects of implementation exist, three are examined in the present study, and a justification for this is provided.
- There are many factors that can affect implementation, and these need to be considered when interpreting any results.
- Previous research suggests that implementation variability can influence pupils’ outcomes. However, only a small number of studies have examined this. Of those that have, it is typically only the relationship between fidelity and outcomes that is examined, although there is evidence to suggest that other aspects of implementation can also play a role.
- The vast majority of studies fail to examine associations between implementation variability and differential gains for pupils at varying levels of risk.
6. Methodology

6.1 Chapter Overview

This chapter aims to provide a detailed account of and justification for the methodology and methods employed for the current study. The chapter starts with an overview of the context of the study and its distinction from the wider GBG trial, before exploring the epistemological stance of the study. A rationale is given for the use of mixed methods, and information is provided pertaining to the study’s design, sample, materials, procedure and analytic strategy. A comprehensive review of the ethical considerations is also presented.

6.2 Context of the Study

The data utilised in this study was collected as part of a larger, national, independent evaluation of the GBG, carried out by a research team at the University of Manchester. This RCT was funded by the Education Endowment Foundation (EEF) to evaluate the effectiveness of the intervention regarding pupil and teacher outcomes. Mentor UK, an international charity working towards the prevention of harm caused by substance misuse, were also provided with funding to support the implementation of the GBG in UK schools. Two separate organisations were chosen to ensure an independent and unbiased evaluation.

6.2.1 The GBG Trial

Mentor UK and the University of Manchester received funding to implement and evaluate the GBG for two years, from 2015 to 2017. In total, 77 mainstream primary schools were recruited, with the children in year 3 in the first year of the trial being the target participants. Schools were situated in Greater Manchester, Yorkshire, and the Midlands. As this was an RCT, following recruitment and baseline data collection, schools were randomly allocated to either the intervention or comparison (usual practice; UP) arm of the trial. Mentor UK were responsible for
supporting the implementation and running of the GBG for the two year period, and so provided teachers with the appropriate resources and training. They also hired trained coaches to oversee the implementation in schools.

6.2.2 National Evaluation of the GBG

The evaluation of the GBG ran alongside its implementation. There were five research questions based on the hypotheses below, which were predominately quantitative in nature:

1. Children in primary schools implementing the GBG over a two-year period will demonstrate measurable improvements in reading (1a) and behaviour (1b) when compared to those children attending control schools
2. The effects outlined in H1 above will be amplified for boys exhibiting borderline/abnormal levels of conduct problems at baseline
3. The effects outlined in H1 above will be amplified for children eligible for FSM
4. The effects outlined in H1 above will be moderated by variation in implementation
5. Teachers implementing the GBG will demonstrate measurable improvements in efficacy in classroom management (5a) classroom stress (5b) and retention (5c) when compared to teachers in control schools

Data to address the quantitative hypotheses came from teacher surveys, structured observations and pupil reading tests. One cohort of children were followed for the two years of the trial, with data being collected at five time points (see table 6.1 below). Data regarding each child was also collected from the National Pupil Database (NPD).
Qualitative data were collected from six self-selecting case study schools at four time points throughout the trial. This was done in order to gain a deeper understanding of the processes involved in the implementation of the GBG. Data were collected from observations, as well as semi-structured interviews with teachers, head teachers, parents and coaches, and focus groups with pupils.

It should be noted that the author of this thesis was also employed as a research assistant on the GBG trial.

### 6.2.3 Independence of the Current Study

Whilst the data utilised for this thesis was collected as part of the GBG trial, it is important to note that key differences exist between the two studies. This study is independent of the wider GBG evaluation and so is an original piece of work that provides a unique contribution to the fields of prevention and implementation science, and risk research. Table 6.2 below presents a more detailed account of the independence of the current study.
### Table 6.2: Description of differences between the current study and the GBG trial

<table>
<thead>
<tr>
<th>Section</th>
<th>Doctoral Study</th>
<th>GBG Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outline</td>
<td>Independent doctoral study lasting three years.</td>
<td>Two year evaluation in conjunction with the EEF.</td>
</tr>
<tr>
<td>Staffing</td>
<td>Individual author of the thesis.</td>
<td>Large research team.</td>
</tr>
<tr>
<td>Aims (1)</td>
<td>To identify the variables that increase a pupil’s risk of behavioural and academic difficulties.</td>
<td>To evaluate the overall effectiveness of the GBG for all pupils and teachers relative to the comparison group.</td>
</tr>
<tr>
<td>Aims (2)</td>
<td>To evaluate the effectiveness of the GBG for pupils at different levels of risk exposure.</td>
<td>To evaluate the effectiveness of the GBG for boys at-risk of behavioural problems and pupils eligible for FSM specifically relative to the comparison group.</td>
</tr>
<tr>
<td>Aims (3)</td>
<td>To examine the differential effects of implementation fidelity, quality and dosage specifically on the outcomes of at-risk pupils.</td>
<td>To examine the effects of implementation fidelity, quality, dosage, reach and participant responsiveness on all pupils’ outcomes relative to the comparison group.</td>
</tr>
<tr>
<td>Temporality</td>
<td>Cross-sectional/longitudinal.</td>
<td>Longitudinal.</td>
</tr>
<tr>
<td>Design</td>
<td>Hybrid concurrent embedded mixed methods design (QUANqual).</td>
<td>Hybrid concurrent embedded mixed methods design (QUANqual).</td>
</tr>
<tr>
<td>Main Groups of Interest</td>
<td>At-risk pupils.</td>
<td>All pupils and teachers.</td>
</tr>
<tr>
<td>Participants – Quantitative</td>
<td>At-risk pupils receiving the GBG were the target participants. Pupils in the comparison group were</td>
<td>All pupils and their teachers in both the GBG and comparison arms of the trial were utilised for the trial.</td>
</tr>
</tbody>
</table>
utilised for RQ1 and RQ2. Teachers were only involved to complete the pupil behaviour surveys (TOCA-C).

<table>
<thead>
<tr>
<th>Participants – Qualitative</th>
<th>Pupils in the six case study schools were involved, as well as teachers and head teachers.</th>
<th>Pupils, teachers, head teachers, parents and coaches in the six case study schools were involved.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>Multi-level modelling (quantitative) and thematic analysis (qualitative).</td>
<td>Multi-level modelling (quantitative) and thematic analysis (qualitative).</td>
</tr>
<tr>
<td>Outcome Variables</td>
<td>Pupils’ disruptive behaviour and reading attainment.</td>
<td>Pupils’ concentration, disruptive and pro-social behaviour, and reading attainment. Teacher self-efficacy, classroom stress and teacher retention.</td>
</tr>
<tr>
<td>Target Audience</td>
<td>The academic community are the target audience.</td>
<td>The EEF and schools are the target audience.</td>
</tr>
</tbody>
</table>

### 6.3 Epistemology

At its broadest definition, a paradigm is considered a “worldview”, or an “all-encompassing way of experiencing and thinking about the world” (Morgan, 2007, p. 40). However, more specifically to the field of research, paradigms are defined as a “set of common beliefs and agreements shared between scientists about how problems should be understood and addressed” (Kuhn, 1962, p. 45). Within paradigms are ways of viewing reality (ontology), the tools we use to know about reality (methodology) and the meaning ascribed to the knowledge of that reality (epistemology) (see figure 6.1). Therefore, the paradigm chosen by the researcher will have great influence not only over the research carried out, but also the
methodological approach taken, and the interpreting and reporting of the findings (Darlaston-Jones, 2007).

Figure 6.1: Overview of a paradigm

Epistemological stances are considered to be distinct belief systems that influence the ways in which research questions are asked and answered (Morgan, 2007), of which there are four main standpoints: post-positivist, constructivist, transformative and pragmatic (Mertens, 2012). Although reviewing these in detail is beyond the scope of this research, at its purist form, constructivist epistemologies generally take a qualitative approach, while post-positivist epistemologies utilise quantitative methods.

6.3.1 Pragmatism

The dispute between purists regarding the best approach to use is longstanding, with qualitative and quantitative research viewed as dichotomous methods. This has led to what is known as the “incompatibility thesis”, which posits that quantitative and qualitative methods cannot and should not be mixed. However, some researchers are recognising the importance of using both methods in order to “draw from the strengths and minimize the weaknesses of both”; this has led to a third research paradigm known as pragmatism (Johnson & Onwuegbuzie, 2004, p. 15). Pragmatism emphasises creating knowledge and addressing problems through inquiry and appropriate lines of action (Morgan, 2014). Tashakkori and Teddlie
(1998) described the pragmatist as free to study what interests them in the ways deemed most appropriate, and to use the results to bring about positive consequences; thus pragmatic research is focused on finding workable approaches to problem-solving in order to answer specific questions (Morgan, 2014).

However, pragmatism has been subject to criticism from researchers for its “what works” attitude. Lincoln (2010) argued that pragmatists say nothing of their ontology or epistemology; it is suggested that they dismiss important metaphysical issues such as the nature of truth and reality (Morgan, 2014). Instead, researchers adopt this paradigm for ease, in order for them to be able to utilise the methods of their choosing, or to further their academic agendas (Mertens, 2015). However, pragmatists argue that in the new worldview, in which pragmatism is situated, knowledge consists of warranted assertions, resulting from taking action and experiencing the outcomes. They suggest that research has moved on, and pragmatism replaces an older, more abstract, philosophical way of thinking, with an emphasis on the pursuit of practical knowledge (Morgan, 2014).

Hence, the pragmatic paradigm supports the use of mixed methods in research (discussed further in the following section), arguing that there is not one set of methods that is appropriate; instead, the method should be chosen based on what fits with the research question in a specific study (Mertens, 2012). It is proposed that research methodologies should be mixed in ways that provide the best opportunities for answering important research questions. Furthermore, it is thought that taking a pragmatic stance will improve communication among researchers in order to advance knowledge (Johnson & Onwuegbuzie, 2004). Thus, pragmatism was the approach taken for the current research, as it allows for quantitative and qualitative methods to be utilised simultaneously, in order to provide the most comprehensive answers to the research questions.
6.3.2 Pragmatic Mixed Methods

Mixed methods research (MMR) has emerged as a dominant paradigm in recent years, particularly in educational research, as it is considered to combine the strengths of both quantitative and qualitative datasets (Moseholm & Fetters, 2017). However, inconsistencies still exist as to what constitutes mixed methods (Doyle, Brady, & Byrne, 2009). Although MMR was defined by Tashakkori and Creswell (2007) as “research in which the investigator collects and analyses data, integrates the findings and draws inferences using both qualitative and quantitative approaches or methods in a single study” (p.4), some definitions view MMR as simply the collection and analysis of quantitative and qualitative data, whilst others emphasise the importance of complete integration of the two approaches (Doyle et al., 2009).

Several researchers have sought to establish the benefits of MMR as a research methodology, the most well-known of which being Greene and colleagues (1989). Greene identified five purposes for conducting MMR, illustrated in table 6.3 below (adapted from Grey, 2018). More recently, Bryman (2006) conducted a review of 232 mixed methods papers and identified 16 reasons for conducting MMR; however, many of these reasons are similar to those identified by Greene and colleagues.

Table 6.3: Purposes for conducting MMR (Greene et al., 1989)

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Description</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triangulation</td>
<td>Convergence and corroboration of results from different methods</td>
<td>Increases the validity of constructs by allowing for biases present in one method</td>
</tr>
<tr>
<td>Complementarity</td>
<td>Elaborates on, enhances, illustrates or clarifies results of one method with results from the other method</td>
<td>Increases the meaningfulness and validity of constructs by capitalising on method strengths and minimising method</td>
</tr>
<tr>
<td></td>
<td>Weaknesses</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Development</strong></td>
<td>Utilises the results of one method to help or inform the other method</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increases the validity of constructs by capitalising on method strengths</td>
<td></td>
</tr>
<tr>
<td><strong>Initiation</strong></td>
<td>Uses contradictions in findings to inform or reform questions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increases the breadth and depth of results by analysing them from different perspectives</td>
<td></td>
</tr>
<tr>
<td><strong>Expansion</strong></td>
<td>Extends the breadth and range of research by using different methods</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increases the scope of research by selecting the methods most appropriate for multiple inquires</td>
<td></td>
</tr>
</tbody>
</table>

However, MMR has been met with much criticism from purist researchers over the years, with many arguments focusing around the incompatibility thesis, suggesting that quantitative and qualitative methods cannot be mixed in a single study. Furthermore, Sale and colleagues (2002) warned against researchers overlooking the underlying assumptions and differences between the two purist paradigms (Doyle et al., 2009). Practical issues have also been raised regarding the extra demand placed on time and resources, and the expectation for the researcher to have a good knowledge of quantitative and qualitative methods both independently, and how to mix them appropriately (ibid).

Nevertheless, other researchers have defended MMR, arguing that being able to combine different methods is beneficial; MMR does not follow a prescribed formula but instead is “an artful crafting of the kind of mix that will best fulfil the intended purposes for mixing within the practical resources and contexts at hand” (Greene, 2007, p. 129). In addition, Onwuegbuzie (2002) suggests that the dichotomies posited by purists are false; instead, positivist and non-positivist philosophies lie on an epistemological continuum with MMR occupying the middle ground. Finally, Tashakkori and Teddlie (2003) argue that pragmatists abide by the “dictatorship” of the research question; in other words, the research question is more important than the method underlying it. In the current study, pragmatic mixed methods were
chosen as the framework by which to answer the research questions, meaning that a deeper exploration could be conducted into not just “what works” for pupils, but also how and why it works under certain conditions and circumstances (Bonell et al., 2012). A purely quantitative or qualitative method would not have allowed for all of these questions to be answered so comprehensively.

6.4 Design

A two-year cluster-RCT was utilised, with schools as the unit of randomisation. Randomisation was conducted independently by the Manchester Academic Health Science Centre Clinical Trials Unit. Although in theory randomisation ensures that baseline characteristics are equally distributed and that sampling bias is reduced (Sullivan, 2011; Torgerson & Torgerson, 2001), this is not guaranteed; thus a minimisation algorithm was applied to the randomisation to ensure balance across the trial regarding school size and the proportion of pupils eligible for FSM. Although RCTs have been met with some controversy (Grossman & Mackenzie, 2005), they are generally considered to be the “gold standard” of research design when looking to identify the effects of a prescribed intervention on a specified set of outcomes (Maughan, 2013). However, Morrison (2001) argues that RCTs have a restricted view of causality and predictability, understate other data sources and factors such as context, and ignore or over-simplify the complex underlying processes taking place. Therefore, the present research also made use of qualitative techniques to provide context for the quantitative findings and explain the underlying mechanisms of change; thus answering the research questions more comprehensively.

In this case, the research design was a hybrid concurrent embedded (QUANqual) design, meaning that quantitative and qualitative data were collected and analysed separately, in order for the researcher to “best understand the research problem” (Creswell & Plano Clark, 2007, p. 64). Quantitative data was considered dominant,

---

1 Balance was required for school size to ensure an approximately equal number of pupils in each arm of the trial. Balance for pupils eligible for FSM was a requirement stipulated by the funders of the trial, The EEF.
and was utilised to determine pupils’ outcomes, whilst the qualitative data played a supportive, explanatory role. As embedded designs mean that one type of data are given less priority, the amount of extensive data collection that is required is limited; this has logistical benefits, particularly for postgraduate students, as the need for time and resources is reduced (ibid).

However, there are some limitations to this design; it is argued that this approach results in an unequal balance of evidence (Creswell, 2009), may introduce potential treatment bias (Creswell & Plano Clark, 2007), and requires the researcher to have a sufficient knowledge of both quantitative and qualitative methodologies (Doyle et al., 2009). Nevertheless, the basic premise of this design, that a single dataset is not sufficient, there are different types of questions that need to be answered, and that each requires a different type of data, is well-suited to the two different types of research questions in the current study; therefore allowing for an exploration into the effects of the GBG on at-risk pupils, and the ways in which these effects occurred.

6.4.1 Quantitative Approach

Whilst RQ2 and RQ3 were longitudinal in nature, looking for changes in pupils’ school functioning over a two-year period in response to an intervention, RQ1 only focused on data collected at baseline. The findings from RQ1 then informed the analysis for RQ2 and RQ3. Although longitudinal research has some drawbacks, such as a risk of attrition, it also makes it possible to infer causation, thus detecting changes in the target sample at both the individual- and the school-level (Ployhart & Vandenberg, 2010). A more detailed description of the quantitative approaches utilised to analyse these data are provided in section 6.8.

6.4.2 Qualitative Approach

A qualitative approach was utilised for RQ2a to explain the quantitative findings. This research question investigated the ways in which the GBG had an effect on
pupil outcomes; thus the qualitative data aids understanding of the underlying mechanisms affecting change. The use of qualitative techniques (outlined in more detail in section 6.9), and specifically a hybrid approach, enriched the quantitative findings by not only explaining them, but also by allowing for emergent themes to develop (Caracelli & Greene, 1993). This means that unanticipated findings could be explored in detail, further contributing to the evidence base in this field.

6.4.3 Integration of the Quantitative and Qualitative Strands

Creswell & Plano Clark (2007) cited one of the potential challenges of MMR to be the integration of the results of the two methods when they are used to answer different research questions. However, they also stated that the intention of an embedded design is not to converge two different sets of data to answer the same question; instead, the premise of this design is that different questions need to be answered, and that each type of question requires a different type of dataset. Therefore, the qualitative dataset can help to provide a greater insight into the quantitative findings when the datasets do converge. However, when the qualitative and quantitative data diverge, the qualitative data may draw attention to additional findings not highlighted by the quantitative results.

6.4.4 Variables in the Study

The outcome (dependent) variables in the study were teacher-reported disruptive behaviour and pupils’ reading test scores, at both baseline (for RQ1) and follow-up (for RQ2 and RQ3). The predictor (independent) variables were the 16 a priori risk variables established in the literature. Regarding RQ3, three aspects of implementation (fidelity, quality and dosage) were focused on in detail (as discussed in section 5.3.1).

Table 6.4 below provides a list of the variables identified in the previous research that have been found to increase the risk of poor academic and behavioural outcomes (see chapter 2). For the purposes of the multi-level analysis, these are
categorised into pupil-level and school-level variables. The sources that these data have been collected from are also listed.

Table 6.4: Potential risk variables investigated in the study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Risk Group</th>
<th>Source of Data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PUPIL-LEVEL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>NPD¹</td>
<td>Binary: Male or Female</td>
</tr>
<tr>
<td>Relative Age (season of birth)</td>
<td>Summer-born</td>
<td>NPD</td>
<td>Binary: <em>Summer Born</em> (June – August) or <em>Non-Summer Born</em> (September – May)</td>
</tr>
<tr>
<td>Looked-After Status</td>
<td>Looked-after child</td>
<td>NPD</td>
<td>Binary: Looked-After Child or Not Looked After</td>
</tr>
<tr>
<td>SEND Status</td>
<td>Identified as SEND</td>
<td>NPD</td>
<td>Binary: SEND or non-SEND</td>
</tr>
<tr>
<td>Familial Deprivation (FSM Eligibility)</td>
<td>FSM eligible (income &lt;60% of national median)</td>
<td>NPD</td>
<td>Binary: Eligible or Not Eligible</td>
</tr>
<tr>
<td>Ethnic Group</td>
<td>Minority ethnic group</td>
<td>NPD</td>
<td>Binary: White British or Other</td>
</tr>
<tr>
<td>First Language</td>
<td>EAL</td>
<td>NPD</td>
<td>Binary: EAL or English as First Language</td>
</tr>
<tr>
<td>Neighbourhood Deprivation (IDACI Score)</td>
<td>High IDACI score</td>
<td>LGA²</td>
<td>Continuous: the IDACI score for the child’s neighbourhood; higher scores indicate higher neighbourhood deprivation (0-1)</td>
</tr>
<tr>
<td><strong>SCHOOL-LEVEL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achievement</td>
<td>Low average achievement</td>
<td>SCPT³ (DfE)</td>
<td>Continuous: proportion of pupils achieving the NC expected</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>Source</td>
<td>Notes</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------------------------</td>
<td>-------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>SEND</td>
<td>High % of pupils with an SEND</td>
<td>SCPT (DfE)</td>
<td>Continuous: proportion of pupils identified as having an SEND; higher rates indicate higher numbers of pupils with an SEND (0-100%)</td>
</tr>
<tr>
<td>EAL</td>
<td>High % of EAL pupils</td>
<td>SCPT (DfE)</td>
<td>Continuous: proportion of pupils classified as EAL; higher rates indicate more EAL pupils (0-100%)</td>
</tr>
<tr>
<td>Attendance</td>
<td>High % average pupil absence</td>
<td>SCPT (DfE)</td>
<td>Continuous: the average pupil absence at the school; higher rates indicate more instances of absence (0-100%)</td>
</tr>
<tr>
<td>Behaviour</td>
<td>High % Behaviour Problems</td>
<td>Aggregate of baseline SDQ^4 conduct problems scores</td>
<td>Continuous: proportion of pupils scoring in the borderline/abnormal range for conduct problems at baseline; higher rates indicate higher levels of behaviour problems (0-100%)</td>
</tr>
<tr>
<td>Size</td>
<td>Large school size</td>
<td>EduBase (DfE)</td>
<td>Continuous: number of pupils enrolled at the school; higher numbers indicate larger schools</td>
</tr>
<tr>
<td>Urbanicity</td>
<td>Urban school</td>
<td>EduBase (DfE)</td>
<td>Binary: school is located in urban or rural area</td>
</tr>
<tr>
<td>Deprivation (FSM)</td>
<td>High % pupils eligible for FSM</td>
<td>EduBase (DfE)</td>
<td>Continuous: proportion of pupils receiving FSM; higher rates indicate greater numbers of pupils with a low SES (0-100%)</td>
</tr>
</tbody>
</table>
Overall, 16 risk variables for poor academic and behavioural outcomes that had been established to varying degrees in the literature (see section 2.5) were available in the GBG trial dataset, eight at the pupil-level and eight at the school-level.

### 6.5 Participants

Participants in this study were drawn from the sample utilised in the GBG trial; Mentor UK were responsible for the recruitment of schools to the trial (see section 6.2).

#### 6.5.1 Schools

Schools comprised state-funded mainstream primary schools in Greater Manchester (GM), West (WY) and South Yorkshire (SY), and the Midlands. The total number of schools was 77. As this was a cluster-RCT, with schools being the unit of randomisation, the schools were randomly allocated to either the intervention or comparison (UP) group. In total, 38 schools were allocated to the GBG and 39 schools were UP schools. In the GBG schools, there were 104 teachers from whom implementation data were sought. As can be seen in table 6.5, few substantive differences existed between the GBG and the UP schools. However, a medium-sized difference did exist regarding the proportion of pupils identified as having an SEND; this highlights that randomisation does not always ensure an equal balance across the different arms of a trial.

Regarding the overall sample, schools were generally above average on all socio-demographic characteristics, meaning that they were larger schools, with more pupils with SEND, eligible for FSM, speaking EAL, and those of an ethnic minority relative to the national average (NA). The exception to this was school-level
attainment; schools in the current sample had lower proportions of pupils achieving the KS2 targets. These discrepancies between the sample and overall population may be due to the fact that the majority of the schools were within a particular region, Greater Manchester, an ethnically diverse conurbation in the UK with some of the poorest areas in the country (CLES, 2012; Jivraj, 2013). Furthermore, these types of schools are potentially more interested in participating in a trial aiming to reduce behaviour problems and increase academic attainment, where the perceived need is typically greater.

Table 6.5: School-level socio-demographic data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall Sample (mean)</th>
<th>NA Sample (mean)</th>
<th>GBG Sample (mean)</th>
<th>UP Sample (mean)</th>
<th>Comparison of GBG and UP (Cohen’s d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size – number of pupils on roll¹</td>
<td>306.9</td>
<td>269</td>
<td>298.2</td>
<td>315.4</td>
<td>0.11</td>
</tr>
<tr>
<td>Attendance – overall % absence²</td>
<td>4.2</td>
<td>4.6</td>
<td>4.3</td>
<td>4.2</td>
<td>0.10</td>
</tr>
<tr>
<td>FSM – proportion of pupils eligible for FSM³</td>
<td>26</td>
<td>15.6</td>
<td>27.6</td>
<td>24.5</td>
<td>0.23</td>
</tr>
<tr>
<td>EAL – proportion of pupils speaking EAL³</td>
<td>22.6</td>
<td>19.4</td>
<td>22</td>
<td>23.2</td>
<td>0.04</td>
</tr>
<tr>
<td>Ethnic Minority – proportion of ethnic minority pupils³</td>
<td>32.9</td>
<td>30.4</td>
<td>32.4</td>
<td>33.3</td>
<td>0.03</td>
</tr>
<tr>
<td>SEND – proportion of</td>
<td>19.5</td>
<td>15.4</td>
<td>20.9</td>
<td>18.2</td>
<td>0.60</td>
</tr>
<tr>
<td>pupils with SEND(^4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attainment – proportion of pupils achieving Level 4+ at KS2(^5)</td>
<td>76.2</td>
<td>80</td>
<td>76.2</td>
<td>74.9</td>
<td>0.12</td>
</tr>
</tbody>
</table>


### 6.5.2 Pupils

The participants of this trial comprised pupils of the schools in year 3 (aged 7-8) at the beginning of the trial’s intervention phase (September 2015). These pupils were followed into year 4 for the second year of the trial. The total possible sample was 3084 children. Socio-demographic data were collected for each pupil from the NPD. As can be seen in table 6.6, only very small differences exist between pupils in the GBG and UP schools, suggesting that the arms of the trial were evenly balanced. Regarding the overall sample of pupils participating in the trial, they were generally above average on all socio-demographic characteristics, meaning that there were more pupils with SEND, eligible for FSM, speaking EAL, and those of an ethnic minority. The exception to this was attainment, pupils in the sample typically scored lower on their Key Stage One (KS1) reading tests, relative to the NA. This may be because schools with these demographics are more likely to have a perceived need for an intervention that increases attainment and reduces behavioural problems.
### Table 6.6: Pupil-level socio-demographic data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall Sample (mean)</th>
<th>GBG Sample (mean)</th>
<th>UP Sample (mean)</th>
<th>Comparison of GBG and UP (Cohen’s d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex – proportion of male students</td>
<td>52.6</td>
<td>50</td>
<td>50.4</td>
<td>54.9</td>
</tr>
<tr>
<td>FSM – proportion of pupils eligible for FSM¹</td>
<td>24.8</td>
<td>15.6</td>
<td>27.4</td>
<td>22.8</td>
</tr>
<tr>
<td>EAL – proportion of pupils speaking EAL¹</td>
<td>27.3</td>
<td>19.4</td>
<td>26.2</td>
<td>31</td>
</tr>
<tr>
<td>Ethnic Minority – proportion of ethnic minority pupils¹</td>
<td>33.5</td>
<td>30.4</td>
<td>32.8</td>
<td>34.2</td>
</tr>
<tr>
<td>SEND – proportion of pupils with SEND²</td>
<td>20.3</td>
<td>15.4</td>
<td>23.1</td>
<td>17.6</td>
</tr>
<tr>
<td>Attainment – average point score at KS1 reading³</td>
<td>15.2</td>
<td>16.6</td>
<td>15.0</td>
<td>15.4</td>
</tr>
<tr>
<td>SDQ Conduct Problems – proportion of pupils at risk</td>
<td>16.1</td>
<td>-</td>
<td>17.9</td>
<td>14.3</td>
</tr>
</tbody>
</table>

³ Phonics screening check and key stage 1 assessments in England, 2015. Nottingham: DfE
6.5.3 Quantitative Strand

The sample for RQ1 and RQ2 made use of the entire pupil sample recruited for the GBG trial. However, for RQ3, only data from pupils in the GBG arm were utilised, meaning that the total possible sample was 1560 pupils.

6.5.4 Qualitative Strand

There were six self-selecting case study schools in the qualitative strand of the study. All of these schools were in the GBG arm of the trial. It should be noted that on average the number of pupils with behaviour problems, eligible for FSM, and speaking EAL were higher than the NA. The demographic characteristics of each school are displayed in table 6.7 below. In order to anonymise the case study schools, these are displayed as either equal to (=) greater than (+) or less than (-), to indicate how close the schools were to the NA on each given characteristic. Data were collected from teachers, members of the senior leadership team, and pupils in the case study schools.

**Table 6.7: Case study school**

<table>
<thead>
<tr>
<th>School</th>
<th>Region</th>
<th>% FSM</th>
<th>% EAL</th>
<th>% Ethnic Minority</th>
<th>% SEND</th>
<th>% Absence</th>
<th>% NC Level 4</th>
<th>School Size</th>
<th>Location</th>
<th>No. Yr3 Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GM</td>
<td>+</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>-</td>
<td>Urban</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>GM</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>=</td>
<td>=</td>
<td>++</td>
<td>Urban</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>GM</td>
<td>++</td>
<td>++</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>+</td>
<td>Urban</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>WY</td>
<td>=</td>
<td>++</td>
<td>++</td>
<td>=</td>
<td>=</td>
<td>-</td>
<td>Urban</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>GM</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>=</td>
<td>--</td>
<td>++</td>
<td>Urban</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>SY</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>=</td>
<td>=</td>
<td>-</td>
<td>Urban</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

= within +/- 10% of NA
+ 10-20% higher than NA; ++ 20-30% higher than NA; +++ 30-40% higher than NA
- 10-20% lower than NA; -- 20-30% lower than NA; --- 30-40% lower than NA
6.5.5 Attrition

At baseline, 77 schools with 3084 pupils were initially included in the quantitative sample. Baseline (T1) disruptive behaviour scores were obtained for 2966 of those pupils, and 3014 baseline reading point scores. Whilst some classes ceased implementation of the GBG over the two years of the trial, all 77 schools provided both disruptive behaviour and reading data at the end of the trial (T2), meaning that the school-level attrition rate was 0%. However, there was a reduction in the number of disruptive behaviour and reading point scores obtained for pupils. 390 pupils (12.6%) left the school during the course of the trial, meaning that T2 data could not be collected for them. Teachers failed to provide T2 disruptive behaviour scores for a further 182 pupils (5.9%), meaning that disruptive behaviour scores were missing for 572 pupils (18.5%). A further 175 (5.7%) were absent on the day that the reading test was administered, meaning that T2 reading scores were missing for 565 pupils (18.3%).

Regarding risk grouping, the majority of pupils were allocated to a risk group for disruptive behaviour and reading attainment. However, some pupils (N = 232) could not be allocated to a reading risk group due to missing socio-demographic data; this was a result of them being included in the trial after the request for data had been submitted to the NPD. The number of pupils in each of the risk groups with complete data also fell over the course of the trial, as they were missing either disruptive behaviour or reading attainment scores. The number of pupils in each of the risk groups is outlined in table 6.8 (see section 7.6.1 for more detail).
Whilst all 38 GBG schools provided some outcome data at T2, some schools did cease implementation of the GBG over the course of the trial, meaning that FidQual and/or dosage data could not be collected. Six of the 60 classes ceased implementation in the first year of the trial, with a further eight classes ceasing implementation in the second year. Overall, 14 classes had ceased implementation by the end of the trial, resulting in an attrition rate of 23%. In the first year of the trial, 54 classes were observed playing the GBG (providing FidQual data), and 48 provided dosage data via the online scoreboard. In the second year of the trial, 45 of the classes were observed, and 39 provided scoreboard data. Missing data were addressing through the use of multiple imputation, discussed in more detail in section 7.2.

Regarding the qualitative dataset, six self-selecting schools were recruited at baseline. However, two of these schools ceased implementation of the GBG after the first year of the trial and so teacher interview (TI) and pupil focus group (PFG) data were not provided by these schools. There were further difficulties accessing one of the case study schools in the second year of the trial; although data were collected from this school at both time points, not all of the intended data were collected. Table 6.9 outlines visits to schools and data collected over the course of the trial.

<table>
<thead>
<tr>
<th>Risk Group (Behaviour)</th>
<th>N T1</th>
<th>N T2</th>
<th>Risk Group (Reading)</th>
<th>N T1</th>
<th>N T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-risk</td>
<td>1680</td>
<td>1638</td>
<td>No risk</td>
<td>406</td>
<td>406</td>
</tr>
<tr>
<td>Medium-risk</td>
<td>1228</td>
<td>1181</td>
<td>Low-risk</td>
<td>1710</td>
<td>1702</td>
</tr>
<tr>
<td>High-risk</td>
<td>129</td>
<td>119</td>
<td>Medium-risk</td>
<td>499</td>
<td>495</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>High-risk</td>
<td>237</td>
<td>232</td>
</tr>
<tr>
<td>Total</td>
<td>3037</td>
<td>2938</td>
<td>Total</td>
<td>2852</td>
<td>2835</td>
</tr>
</tbody>
</table>

*Table 6.8: Attrition by risk group*
Table 6.9: Case study data collection

<table>
<thead>
<tr>
<th>School</th>
<th>Visit 1</th>
<th>Visit 2</th>
<th>Visit 3</th>
<th>Visit 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gy985</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sa774</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tr248</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ci603</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hd887</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Tm517</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

6.6 Materials

6.6.1 Quantitative Strand

Data for the quantitative strand was collected from all schools in the GBG trial. Data were collected through various measures including teacher surveys, pupil tests, and researcher observation schedules. The measures are discussed in more detail below.

6.6.1.i Behavioural Outcomes – TOCA-C

Behaviour was assessed using the Teacher Observation of Children’s Adaptation Checklist (TOCA-C; Koth, Bradshaw, & Leaf, 2009) in all schools. The TOCA-C is a 21-item scale that assesses pupils’ concentration, and disruptive and pro-social behaviour (see appendix 1). Whilst the TOCA was originally a structured interview, it has since been adapted into a written checklist for time and financial purposes. Specifically for this study (as opposed to the wider GBG trial), only the 9-item disruptive behaviour subscale was utilised. In the checklist, statements are provided about the child (e.g. gets angry when provoked by other children), which teachers read and endorse on a 6-point scale (Never/Rarely/Sometimes/Often/Very
Often/Ashly Always). Children’s scores are then summed, with higher scores indicating more maladaptive behaviours (Kourkounasiou & Skordilis, 2014).

One issue with the use of the TOCA-C as a measure of disruptive behaviour in the current study is that different teachers were asked to complete them at the two time-points, which could affect the reliability of the findings. Koth and colleagues (2009) acknowledge that factors such as the time of year and pupils’ demographic characteristics can influence teachers’ responses. However, as this is the case across both arms of the trial, it affected both groups equally. Generally, the TOCA-C has previously been found to good psychometric properties. For instance, the Cronbach’s alphas reported during the initial development of the scale were above the widely accepted cut-off of 0.7 for all three of the subscales (α=.87-.97), indicating high internal consistency. A series of exploratory factor analyses were also conducted, identifying a three-factor solution; all items included in the final scale had pattern coefficients greater than 0.5 (P=.57-.96). Finally, confirmatory factor analysis conducted on two independent random samples found that all scores were near or above the cut-off levels (CFI=.92-.93; TLI=.91-.92; RMSEA=.06-.07; SRMR=.05), indicating adequate data fit. Further studies have also determined that the TOCA-C has high internal reliability (α=.89-.96; Bradshaw, Waasdorp, & Leaf, 2015).

In addition, studies utilising the TOCA have found it to have high test-retest reliability (r=.75-.94; Werthamer-Larsson, Kellam, & Wheeler, 1991) and predictive validity on subsequent violence in adolescence and adulthood (Petras et al., 2004); although the TOCA is different from the TOCA-C, the two measures include almost identical items and have been shown to be strongly correlated (r=.85-.91; Koth et al., 2009). An advantage of the TOCA-C is that it is considerably shorter than other scales, meaning less data burden for teachers, potentially reducing fatigue effects (Kourkounasiou & Skordilis, 2014). The TOCA is also the scale frequently used in previous research surrounding the GBG (e.g. Bradshaw et al., 2015; Chan et al., 2012; Kellam et al., 1994), although this may be due to the fact that it was developed by researchers conducting GBG trials at the time.
The Cronbach’s alpha for the TOCA-C at baseline in the current study was α=.94 for the disruptive behaviour subscale, indicating high internal consistency.

6.6.1.ii Academic Outcomes – Hodder Group Reading Test (HGRT)

As the baseline period for the trial coincided with end of KS1 NC reading tests, pupils’ baseline reading attainment scores obtained from the NPD were in the form of NC point scores. The HGRT (specifically, test sheet 2A) was chosen as the measure for post-test reading attainment, as it produces scores comparable to NC Levels. This means that the pre- (NC reading point scores) and post-test (HGRT raw scores) variables were comparable without having to conduct additional baseline reading tests, thus reducing data burden for schools.

The HGRT is designed to measure pupils’ reading comprehension at word, sentence, and continuous text level. The test can be administered in a whole class context and takes approximately 30 minutes to complete, minimising data burden for the schools. Furthermore, the HGRT is designed as a wide-range test, making it particularly useful for screening and monitoring mixed-ability groups. It has been standardised on over 13,000 children (α = 0.95; Devine et al., 2013) and is capable of reliably measuring reading ability over a broad chronological age range between seven and 16 years. The tests within the HGRT are made up of objectively-scored multiple-choice questions together with questions that assess pupils’ understanding of word meanings, culturally neutral sentence-completion questions, and higher-level “cloze” tasks that look at context and content simultaneously (Vincent & Crumpler, 2007).

6.6.1.iii Implementation (Fidelity and Quality) – Observation Schedule

To assess any moderating effects of implementation variability on the outcomes for at-risk pupils in GBG schools, a structured observation schedule was developed by the GBG research team (see appendix 2) based on the coaches’ fidelity checklist and
implementation manual developed by AIR (Ford et al., 2014). The GBG manual is highly prescriptive, detailing the ways in which AIR intend the GBG to be played, and thus the observation schedule was developed in accordance with these strict guidelines. Therefore, the schedule was designed to effectively become a “checklist” version of the manual, to indicate which of the requirements were followed, and to what extent. The observation schedule was further edited and developed based on the videos provided by AIR of the GBG being played; this ensured all aspects of the game were covered. Finally the observation schedule was edited to fit onto one A3 page for ease of use. The more detailed version was kept as a rubric that could be used as a point of reference during observations.

As three researchers conducted the observations, inter-rater reliability (IRR) needed to be established before observations began. IRR is used to demonstrate “consistency among observational ratings provided by multiple coders” (Hallgren, 2012, p. 23). IRR provides a way of quantifying the degree of agreement between coders making independent ratings and thus ensures that all researchers interpret the variables in the same way. Attaining a high IRR is thought to be essential in order to provide accurate representations of the variables being measured (McHugh, 2012). Therefore, in order to establish agreement between the three observers, previously unseen videos were utilised; raters independently watched and scored videos before their scores were analysed. Raters would then discuss their scores and, if necessary, the schedule was edited to account for any ambiguity. A new version of the schedule was then tested in the same way. This process was repeated until intra-class coefficients (ICC) (for ordinal items) and Cohen’s kappa scores (for nominal items) were all “almost perfect” (ICC: >.74; Kappa: >.8) or above (Hallgren, 2012), demonstrating high levels of IRR (see table 6.10).
Table 6.10: IRR scores for the implementation observation

<table>
<thead>
<tr>
<th></th>
<th>Fidelity (Cohen’s Kappa)</th>
<th>Overall Quality (ICC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observer 1</td>
<td>Observer 2</td>
</tr>
<tr>
<td>Observer 1</td>
<td></td>
<td>.93</td>
</tr>
<tr>
<td>Observer 3</td>
<td>1.00</td>
<td>.93</td>
</tr>
</tbody>
</table>

The schedule included a series of items relating to several of the aspects of implementation being examined in the wider GBG trial. For the purposes of this study however, only data from two sections of the schedule were utilised. Fidelity and quality were rated by trained observers on a fixed scale. Fidelity items were scored on a binary yes/no scale, indicating whether the manualised procedures were followed, whilst the observers rated quality on a scale of 0-2, with higher scores indicating higher quality of delivery.

Although teacher implementation surveys were also administered as part of the wider GBG trial, observations were chosen as the preferred measure for the current study as they are thought to increase rigour and validity; independent observations have been found to be greater predictors of outcomes than other forms of implementation measurement, including teacher self-report. Observation schedules are also thought to be less prone to bias and social desirability, and reduce data burden for teachers; although they are more costly than other measures, meaning that it is not always feasible to use them (Domitrovich et al., 2010; Lillehoj et al., 2004).

Prior to analysis of the data, a psychometric assessment of the GBG observation schedule was conducted by the research team, utilising scores from 100 teachers. Due to the small sample size, procedural fidelity items were summed to represent three items measuring pre-game, during game, and post-game fidelity. The five items measuring quality of delivery were also included, as well as the nine items measuring participant responsiveness. The structure of the schedule was assessed
using exploratory factor analysis (EFA) with Weighted Least Squares Mean and Variance adjusted (WLSMV), while accounting for clustering in the data. Only items with factor loadings above .32 were retained (Tabachnick & Fidell, 2015).

Parallel analysis indicated a two-factor structure for the observation schedule (see table 6.11). Subsequently, a two-factor EFA was conducted, with all items loading substantially onto one of two domains: fidelity/quality and participant responsiveness. This finding that fidelity and quality are in fact one factor not only indicates a clear demarcation of the observed behaviour of the implementer from the behaviour of the intervention recipients, but also suggests that fidelity and quality are not independent of each other. This is consistent with the GBG manual, whereby instructions to teachers regarding the implementation of the game also often include a quality component. For instance, teachers are not simply instructed to explain the task to pupils before the GBG begins, but to “thoroughly review the directions to the assignments, checking for understanding” (Ford et al., 2014, p. 15).

As a result of this, fidelity and quality were treated as one combined variable in analyses (FidQual), with teachers receiving a percentage score between zero and 100.

*Table 6.11: GBG observations 2-factor EFA*

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1</th>
<th>Factor 2b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pre-game fidelity</td>
<td>0.48 ***</td>
<td>-0.14</td>
</tr>
<tr>
<td>2. During game fidelity</td>
<td>0.48 ***</td>
<td>0.05</td>
</tr>
<tr>
<td>3. Post-game fidelity</td>
<td>0.37 ***</td>
<td>-0.12</td>
</tr>
<tr>
<td>4. How do pupils respond to the announcement of the game?</td>
<td>0.20</td>
<td>0.39 **</td>
</tr>
<tr>
<td>5. How attentive are pupils to the teacher’s instructions and examples regarding the game?</td>
<td>0.35 *</td>
<td>0.59 ***</td>
</tr>
<tr>
<td>6. How enthusiastic/willing to participate are pupils when discussing the game?</td>
<td>0.44 **</td>
<td>0.57 ***</td>
</tr>
</tbody>
</table>
7. Do rule breaking pupils correct behaviour following infraction?  
   \(-0.01 \quad 0.74 \三是\)

8. How well do pupils respond to members of their team getting a check?  
   \(-0.00 \quad 0.84 \三是\)

9. Teacher interest and enthusiasm  
   0.73 \三是\ 0.09

10. Teacher clarity of expression  
    0.84 \三是\ -0.12

11. Teacher preparedness  
    0.72 \三是\ -0.02

12. Teacher consistency of behaviour  
    0.51 \三是\ 0.01

13. Teacher engagement of pupils  
    0.81 \三是\ 0.12

* \( p < .05 \) ** \( p < .01 \) *** \( p < .001 \).

\( \chi^2 \) (53) = 116.106, \( p < .001 \); RMSEA = .109 (.082, .136), \( p < .05 \); CFI = .880; TLI = .824.

Correlation between factors: \( r = .45, \ p < .05 \)

a: \( \alpha = .65 \); Raykov \( \omega = .66 \)

b: \( \alpha = .70 \); Raykov \( \omega = .70 \)

6.6.1.iv Implementation (Dosage) – Online Scoreboard

An online scoreboard was developed for teachers to record details of games, including infractions, and length and frequency of play. The scoreboard was given to teachers to be used each time the GBG was played. The data regarding the frequency and duration of games was extracted in order to establish dosage. This method of data collection meant that teachers were not required to fill in extra surveys or provide additional dosage information. However, it relied on all the teachers using the scoreboards, and filling them in accurately and honestly.

As dosage is not \emph{consistently} measured with regards to evaluations of the GBG, or school-based interventions more generally, there is no set protocol or “gold standard” for collecting dosage data. Studies often rely on teacher or coach reports to establish dosage of the GBG, which can mean that data are collected sporadically or as an average, instead of being measured systematically. Warren, Fey and Yoder (2007) argue that often only one dimension of dosage is measured (e.g. the game was played three times a week) when evaluating intervention implementation, meaning that potentially vital pieces of information are missed. To overcome these issues, they propose calculating cumulative intervention intensity, which is the
product of dose (average duration of a game) and dose frequency (number of games), along with total intervention duration (total time period of the trial). This method allows for a more precise, multi-dimensional view of GBG dosage over the course of the trial. Whilst this would not be possible when utilising an annual teacher self-report survey as the method of data collection, the online scoreboard does allow for this more comprehensive approach. Therefore, for the analysis of the differential effects of implementation, dosage is measured in total minutes. This means that pupils’ total exposure to the GBG over the course of the trial is measured, allowing for varying lengths and numbers of games played.

6.6.2 Qualitative Strand

Data for the qualitative strand was taken from the case studies conducted for the GBG trial. It was collected through semi-structured interviews and focus groups with various stakeholders within each school, including parents, teachers, teaching assistants, pupils, coaches and members of the senior leadership team. For the purposes of this study, only data from teachers, members of the senior leadership team, and pupils were utilised, as they were the key stakeholders with the most exposure to the intervention.

6.6.2.i Semi-Structured Interviews and Focus Groups

The GBG research team developed semi-structured interview protocols covering a wide range of topics (see appendix 3); these were adapted accordingly for teachers and members of the senior leadership team. These interviews were carried out in each of the GBG case study schools by two members of the research team, one of whom is the author. Although other methods of data collection would have been viable options, interviews were chosen as they focus on the exploration of meaning and experience within a particular group of participants (King & Horrocks, 2010). Therefore, interviews in the current study allowed for the exploration of teachers’ personal experiences, attitudes and feelings towards the GBG. Whilst interviews are not a method recommended for establishing causality (ibid), they can focus on
individual perceptions of causality, thus helping to explain the ways in which teachers felt the GBG benefited their at-risk pupils.

Specifically, semi-structured interviews were utilised. Semi-structured interviews are thought to be sufficiently structured to address specific topics, whilst also leaving space for unanticipated themes to develop. The use of semi-structured interviews therefore suited the hybrid design of this study, as the interviews could be structured in sections and fully open-ended in others, thus allowing for both inductive and deductive themes (Galletta, 2013). The space for open-ended questions meant that teachers could provide details of any differential effects specific to their pupils, and the perceived reasons for these; this could help to explain any differences in outcomes for pupils at varying levels of risk. The interview schedules were designed to last for approximately 30 minutes.

Focus group schedules were also developed for use with pupils (see appendix 4) to triangulate data and to gain an understanding of how and why pupils believed the GBG had influenced their school functioning. It was decided that pupil data would be collected alongside teacher data for triangulation purposes, and to gain different perspectives of the GBG. Although focus groups are often not as in-depth as interviews, they were utilised for pupils as they allowed for more social interaction, meaning that discussions could develop. As well as being less formal than a one-to-one interview, which may be intimidating to children, they allowed for the exploration and development of themes, and for common trends to emerge (Krueger & Casey, 2015). Focus groups of no more than eight participants are recommended in the literature (Kitzinger, 1995), and so six pupils from each class were selected by their teacher to participate. The focus groups were designed to last approximately 20 minutes.

6.6.2.ii Content of the Interviews and Focus Groups

The interviews and focus groups utilised for the GBG trial aimed to develop an overall profile of each school, and to provide explanatory data for all five of the
research questions for the GBG trial. These schedules were developed jointly, and progressively refined, by the GBG research team. The focus of the questions in the schedule changed over the course of the year; while initial questions focused on foundations for the GBG in the school, including perceived need for the intervention, by the end of the year, questions aimed to establish how the GBG was implemented, facilitators and barriers to implementation, and feasibility of the GBG in UK schools. The schedule also aimed to identify any perceived benefits of the GBG, to both the pupils and the teachers, and to establish the reasons for these benefits.

Although each schedule focused on the same broad topics, a different schedule was designed for each type of informant. The senior leadership team were asked more questions regarding the foundations for the GBG, the overall profile of the school, and how the GBG fitted in; whilst teacher interviews focused on how the GBG was implemented, perceived barriers to implementation, and perceived benefits of the GBG. The pupil focus group schedules had considerably less questions, with the aim of facilitating discussion and allowing for unanticipated themes to emerge. The language used in the focus groups was also adjusted to be more age-appropriate. Pupils were asked to describe the GBG, to establish awareness and understanding; they were also asked how they felt about playing, including how they thought it helped them, to establish enjoyment, perceived benefits, and factors affecting implementation.

A subset of questions from the main interview schedules produced data relevant to RQ2a, that is, how and why the GBG may have produced differential gains for at-risk children. Specific questions were asked to answer RQ2a, including whether certain groups of pupils responded differently to the GBG, or experienced differential effects, and why the teachers thought this might be.
According to Gray (2018), for interviews to be credible, validity and reliability must be addressed. Arksey and Knight (1999) suggested that validity (that the instrument is measuring what it was intended to measure) can be strengthened in semi-structured interviews by: building rapport and trust with participants, prompting participants to expand on their initial responses, ensuring that the interviews are long enough to explore topics in depth, and by drawing on the relevant literature when designing the interview schedule. Conversely, to increase reliability (that the instrument consistently measures what it set out to measure) the “interviewer effect” must be avoided. Gray (2018) emphasised that interviewer bias can have a detrimental effect on the reliability of an interview, and can be reduced by the use of a protocol, standardising not only the interview schedule, but also the behaviour of the interviewer. This includes reading the questions as they are written, probing in a non-directive manner, and providing guidance and clarification without influencing the answer of the respondent.

In order to ensure validity and reliability in the current study, a preamble protocol was written to explain to the participants how the interview would be conducted, how long it would last, and some of the general topics that would be covered. It also ensured that participants were asked for permission to record the interview and gave them the opportunity to ask questions; additionally, they were guaranteed confidentiality and were told what the data would be used for. Gray (ibid) states the importance of including these points in a protocol to not only enhance reliability by reducing the researcher effect, but also to develop rapport with the participants to increase validity. Oppenheim (1992) described rapport as an elusive quality, but one that is necessary to ensure participants answer questions willingly; thus it was important that these points were included in a standardised way. Although the interview schedule itself followed a semi-structured format, questions were carefully phrased in the schedule so as to avoid leading the participant, and appropriate prompts and probes were also built into the schedule to reduce bias. Questions developed from previous literature were
grouped based on their similarity, with less probing and sensitive questions placed at the beginning of the schedule, to allow the participants to feel comfortable (Braun & Clarke, 2013). Finally, in terms of conducting the interviews, active and attentive listening is highlighted by Gray (2018) as an important interview skill; therefore care was taken by the researchers to listen more than they talked, identify incomplete responses, and probe where necessary.

A particularly important issue in the present study was the collection of qualitative data involving children. Cohen, Manion and Morrison (2007) emphasise that children differ from adults in their cognitive and linguistic development, attention span, recall ability, life experiences, what they consider to be important, status and power; all of which are significant factors to consider when conducting interviews and focus groups. Arksey and Knight (1999) emphasised several important factors, all of which were considered in the present study, to minimise these issues, including: establishing trust, putting the children at ease quickly, helping them to feel confident, making the interview non-threatening, using straight-forward and age appropriate language, and allowing the children time to think. In order to do this effectively, focus groups were utilised as they are considered to be less intimidating than one-to-one interviews, as pupils are with their classmates; they also allow pupils to talk to each other using their own language, thus helping to reduce potential power differentials. Eder and Fingerson (2003) argue that power and status dynamics are heavily implicated when interviewing children. However, it is thought that group interviewing in a familiar and natural setting are ways to ease these issues (Mayall, 1999); therefore, the researchers conducted the focus groups in a quiet room that the pupils were familiar with, were introduced to the pupils on a first name basis to help them seem approachable, and had a conversation with the pupils prior to the focus group beginning to make the process feel informal.

6.7 Procedure

Figure 6.2 provides an overview of the timeline specific to this study, as opposed to the wider GBG trial. Survey data and reading levels were collected at baseline, prior
to the implementation of the intervention, and then again at the end of the two-year period. One observation per GBG class occurred in the second term of both years of the trial in order to assess implementation. Two case study visits also occurred in both years of the trial, one before Christmas, and one after Easter.

Figure 6.2: Overview of the study procedure

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOCA-C distributed to yr2 teachers</td>
<td>Case study visit 1</td>
</tr>
<tr>
<td>Pupil data requested from NPD</td>
<td>Case study visit 2</td>
</tr>
<tr>
<td>Observation of GBG in yr3</td>
<td>Case study visit 3</td>
</tr>
<tr>
<td>Observation of GBG in yr4</td>
<td>Case study visit 4</td>
</tr>
</tbody>
</table>

6.7.1 Quantitative Strand

In May 2015, all schools participating in the GBG trial were sent information about the evaluation; this included consent letters that were to be sent to the parents of all pupils participating in the first year of the trial. Consent to participate was on an opt-out basis (see section 6.10). Data about the pupils, including names and gender, were collected from the NPD to create the baseline surveys. Year 2 teachers were asked to complete the baseline TOCA-C surveys regarding pupils’ behaviour in order to be eligible for randomisation. Prior to these survey windows opening, teachers were sent emails with instructions and unique passwords that they could use to log in to the secure online surveys. The survey window was open from May to July 2015. Additional pupil data such as CLA and FSM status were requested from the NPD in December 2015. The TOCA-C was distributed again in May 2017, this time to year 4 teachers. Upon completion, the information from the surveys was matched with
the information already collected about the pupils, and identifiable information was removed.

Between January and March 2016 and 2017, one of three members of the research team visited year 3 and 4 classes respectively in all GBG schools to carry out a structured observation assessing implementation (see section 6.6.1.iii). Teachers were asked to implement the GBG as normal while one of the trained observers sat at the back of the classroom and completed the observation schedule. Schools and teachers were given unique codes so that they would not be identifiable. It was decided that implementation data would be treated separately for both years of the trial, to account for bias caused by missing data, and to accurately reflect pupils’ experiences of the GBG over the course of the trial. Over the two year period pupils typically had two different teachers; as implementers can vary by 20 to 40% (Durlak & DuPre, 2008), it was considered important to investigate the effects of individual teachers’ implementation on outcomes. Utilising data from only one year may not have accurately reflected the pupils’ overall experiences of the GBG, whilst aggregating years may have caused high and low implementation to negate each other.

6.7.2 Qualitative Strand

The qualitative strand ran alongside the quantitative one. In September 2015, all teachers allocated to the GBG attended a two-day training event, at which time they were asked to volunteer to be case study schools. A member of the research team visited each of the case study schools twice a year. Each visit involved an observation, and interviews and focus groups with various stakeholders (see table 6.12). Stakeholders were provided with information sheets and opt-in consent was sought from teachers, pupils, and parents of the pupils prior to each interview (see section 6.10).

Semi-structured interviews and focus groups were utilised and recorded before being transcribed by a third-party company. All identifying information was
removed during this process and pseudonyms were allocated. For the pupil focus groups, teachers were encouraged to select a range of participants based on variables such as SEND and FSM status, gender and ability.

*Table 6.12: Details of data collection at each case study visit*

<table>
<thead>
<tr>
<th>Visit</th>
<th>Interviews/Focus Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. December 2015</td>
<td>Year 3 teacher(s)</td>
</tr>
<tr>
<td></td>
<td>Member of senior leadership team/GBG lead</td>
</tr>
<tr>
<td>2. April 2016</td>
<td>Year 3 teacher(s)</td>
</tr>
<tr>
<td></td>
<td>6 pupils (per class)</td>
</tr>
<tr>
<td>3. December 2016</td>
<td>Year 4 teacher(s)</td>
</tr>
<tr>
<td></td>
<td>Member of senior leadership team/GBG lead</td>
</tr>
<tr>
<td>4. March 2017</td>
<td>Year 4 teacher(s)</td>
</tr>
<tr>
<td></td>
<td>Member of senior leadership team/GBG lead</td>
</tr>
<tr>
<td></td>
<td>6 pupils (per class)</td>
</tr>
</tbody>
</table>

**6.8 Quantitative Analytic Strategy**

**6.8.1 Multi-Level Modelling (MLM)**

MLM was first developed for educational research, when researchers realised that the performances of students in the same class were not independent of each other. As the pupil-level is nested in the school-level, pupils attending the same school are more likely to have similar outcomes than those pupils attending different schools, meaning their scores may be correlated (Twisk, 2006). “Standard” statistical techniques, such as multiple regression, assume independent observations; this is the assumption that participants are not affected by an outside influence common to several other participants, and that the participants are not influenced by each other (Papay, 2009). This assumption is thus violated when conducting educational research in schools.
Therefore, MLM was developed as an extension of multiple regression, also designed to assess how much of the variance in a dependent variable (e.g. reading scores) can be accounted for by a range of independent variables (e.g. risk variables). However, unlike standard multiple regression, MLM takes into account the dependency of observations, accounting for clustered and hierarchical data. MLM therefore allows for pupil-level variables (e.g. gender, ethnicity) to be modelled alongside school-level variables (e.g. school size, urbanicity), whilst also recognising the correlated nature of the data at these levels (Twisk, 2006).

It is important to use statistical techniques that allow for clustering, even if the outcome of interest is at the pupil-level, as ignoring clustering can mean that the standard errors and confidence intervals obtained may be spurious. This can lead to the conclusion of real effects, when in reality all that occurred was random variation (Buxton, 2008). MLM was particularly valuable in the current study, as not only did it allow for the correlations of academic and behaviour scores within schools, it also allowed for the clustered nature of the implementation data within the classes. As pupils in the same class all experienced the same levels of fidelity, quality and dosage, which may have differed to pupils’ experiences in other classes and schools, it was important to take this into account when modelling the data. Therefore, MLM was the statistical technique utilised for all three of the research questions.

6.8.2 Key Concepts in MLM

6.8.2.i Empty and Full Models

There are two distinct steps when building a MLM. Firstly, the null, or empty, model is developed, whereby no predictors are added. This empty model partitions the unexplained variance in the outcome variable into within- and between-group components; therefore indicating how much of the variance lies between the schools, and how much is attributable to the pupil-level. The proportion of the variance at each level is displayed as a percentage, known as the ICC (see section
6.8.2.ii below) (Heck, Thomas, & Tabata, 2010). Secondly, the full model is then created with the addition of the predictor variables at the appropriate levels. The full model indicates which of the variables are significant predictors of the outcome variable. Additionally, the empty and full models can be compared using the -2*log likelihood statistic (see section below).

6.8.2.ii Intraclass Correlation Coefficient (ICC)

The ICC is a measure of between-group variability, used to explain the amount of variance between clusters, as opposed to the variance between individuals within clusters. An ICC can be generated for each level of a model; thus in the present study, this means that the ICC was used to explain the proportion of unexplained variance at both the pupil- and school-level. Typically, the ICC is calculated from the empty model, by dividing the variance between clusters by the total variance, where total variance is defined by the summation of the variance between clusters and the variance within clusters. The greater the ICC, the larger the between-group variance (i.e. between schools), and the smaller the within-group variance (i.e. within schools) (Twisk, 2006).

\[
\text{ICC} = \frac{\text{variance(between)}}{\text{(variance(between) + variance(within))}}
\]

6.8.2.iii The -2*log likelihood statistic

This statistic is a measure of overall model fit; indicating how much unexplained variance is left after the model has been fitted. Therefore, the larger the value, the more unexplained variance there is, and thus the poorer the model fit. The log likelihood statistic is based on summing the probabilities associated with the predicted and observed outcomes; hence larger values indicate larger deviations from the expected outcomes (Field, 2009). This statistic was used to compare both of the cumulative models in RQ1 (as well as empty and full models) in the current study, in order to establish the best fitting model; a reduction in the log likelihood
statistic between models would indicate that the additional predictor variables were accounting for significant variance in the outcome variables.

6.8.2.iv The β Coefficients

The β coefficient indicates the amount of variance explained in the outcome variable by each of the predictor variables. This is done in one of two ways. For continuous variables, the coefficient indicates the resultant change in the outcome variable following a one unit change in the predictor variable. For example, a β coefficient value of 0.1 would demonstrate that for every one unit increase in the predictor variable (e.g. a risk variable), scores on the outcome variable (e.g. disruptive behaviour scale) increase by 0.1. Conversely, categorical variables are entered into the model with a reference category, and thus the coefficient indicates the change in the outcome variable as a result of being exposed to the predictor variable, relative to the reference category. For example, a β coefficient value of 0.1 would demonstrate that exposure to the predictor variable (e.g. being male) results in a 0.1 increase in the outcome variable (e.g. disruptive behaviour scores), relative to the reference category (e.g. being female). However, it is worth noting that unless the outcome variable is standardised prior to analysis, the β coefficients must be interpreted as raw scores, and thus direct comparisons between predictor variables and other outcome variables cannot be drawn.

6.8.2.v The intercept $\beta_{0ij}$

This reflects the average score for the average pupil within the average school, whereby the subscript $i$ represents the pupil-level, and $j$ represents the school-level. Standard deviations are also given alongside this in brackets.

6.8.2.vi p Significance

This is commonly used across all areas of psychology and education research and is a measure of statistical significance. The widely accepted cut-off point for statistical
significance in social science research is \( p < .05 \), meaning that the probability of a result occurring by chance (assuming that the null hypothesis is true) is less than 5%. Any predictor variables with a \( p \) statistic of less than .05 are therefore considered statistically significant, and their relationship with the outcome variable is thought to not be due to chance (Field, 2005). \( P \) values are calculated by dividing the coefficient by its standard error, resulting in a \( t \) statistic. This \( t \) statistic is then calculated alongside degrees of freedom to generate the \( p \) value.

6.8.2.vii Standard Error

This refers to how well a particular sample represents the population, and is used when multiple different samples are available in a dataset (e.g. multiple schools). The standard error can be thought of as a measure of the standard deviation of sample means, indicating how well a particular sample (school), represents the other schools in the dataset. The standard error can be calculated by dividing the sample standard deviation (\( s \)) by the square root of the sample size (\( n \)) (Field, 2009).

\[
\sigma_k = \frac{s}{\sqrt{n}}
\]

6.8.3 Screening of the Data

6.8.3.i Power at Point of Randomisation

Prior to the implementation of the GBG, at the point of randomisation, a priori power calculations were conducted. These power calculations were based on the assumption of a two-year trial using disruptive behaviour and reading attainment as the primary outcomes, through which a single year group of pupils would be tracked, and intention-to-treat (ITT) analyses for all pupils conducted. Power calculations were conducted to establish the minimum detectable effect sizes (MDES) for each of the RQs. The MDES is the smallest effect that, if true, has an 80% chance of producing an impact estimate that is statistically significant (Bloom, 1995).
With an ICC of 0.04, an average pre-post correlation of 0.3 (both based on PATHS SDQ data; Humphrey et al., 2016), a cluster size of 40, 77 schools, and alpha and power set to 0.05 and 0.8 respectively (Barker Bausell & Li, 2002), the MDES for ITT analysis was 0.15 (N=3084) for disruptive behaviour. Relative to mean effect sizes (EFs) of between 0.3 and 0.75 reported in the literature (Ialongo et al., 1999; Spilt et al., 2013), this means that the study was more than adequately powered to detect likely treatment effects of the GBG on disruptive behaviour.

Regarding reading outcomes, with an ICC of 0.06 (ICC at baseline – KS1 reading scores), an average pre-post correlation of 0.7 (based on EEF estimates; EEF, 2013), a cluster size of 40, 77 schools and alpha and power values of 0.05 and 0.8 respectively, the MDES for ITT analysis was 0.13 (N=3084). Relative to mean EFs of between 0.25 and 0.51 reported in the literature (Ialongo et al., 1999), this means that the study was more than adequately powered to detect likely treatment effects of the GBG on academic attainment.

6.8.3.ii Power at Point of Analysis

Post-test analyses of MDES for behaviour scores and reading attainment were conducted for all pupils, investigating MDESs for complete case analyses and multiply imputed models (accounting for missing data; more detail can be found in section 7.2). MDESs were also calculated for pupils in allocated risk groups. Risk groups were developed from the descriptive statistics explored in RQ1; this is discussed further in section 7.6.1.

Regarding behaviour, the sample for the imputed analysis (77 schools, N=3084, average of 40 pupils per cluster) provided an MDES of 0.15 in an ITT analysis, utilising an ICC of 0.08 for the TOCA-C, a pre-post correlation of 0.63 and power and alpha at 0.8 and 0.05 respectively. The complete case analysis (77 schools, N=2512, average of 30 pupils per cluster) provided an MDES of 0.16, utilising the same assumptions.
Regarding reading, the sample for the imputed analysis (77 schools, N=3084, average of 40 pupils per cluster) provided an MDES of 0.14 in an ITT analysis, utilising an ICC of 0.08 for the HGRT, a pre-post correlation of 0.75 and power and alpha at 0.8 and 0.05 respectively. The complete case analysis (77 schools, N=2519, average of 33 pupils per cluster) provided an MDES of 0.14, utilising the same assumptions.

Subgroup analyses were conducted based on risk status for disruptive behaviour and reading attainment, utilising the same basic assumptions reported above. The MDESs for the risk categories are reported in table 6.13 for behaviour and 6.14 for reading below.

Table 6.13: MDESs by behaviour risk group

<table>
<thead>
<tr>
<th>Behaviour Risk Group</th>
<th>N</th>
<th>% of Sample</th>
<th>MDES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-risk</td>
<td>1638</td>
<td>56</td>
<td>0.17</td>
</tr>
<tr>
<td>Medium-risk</td>
<td>1181</td>
<td>40</td>
<td>0.20</td>
</tr>
<tr>
<td>High-risk</td>
<td>119</td>
<td>4</td>
<td>0.45</td>
</tr>
<tr>
<td>Total</td>
<td>2938</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.14: MDESs by reading attainment risk group

<table>
<thead>
<tr>
<th>Reading Risk Group</th>
<th>N</th>
<th>% of Sample</th>
<th>MDES</th>
</tr>
</thead>
<tbody>
<tr>
<td>No risk</td>
<td>406</td>
<td>14</td>
<td>0.30</td>
</tr>
<tr>
<td>Low-risk</td>
<td>1702</td>
<td>60</td>
<td>0.15</td>
</tr>
<tr>
<td>Medium-risk</td>
<td>495</td>
<td>18</td>
<td>0.27</td>
</tr>
<tr>
<td>High-risk</td>
<td>232</td>
<td>8</td>
<td>0.37</td>
</tr>
<tr>
<td>Total</td>
<td>2835</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Regarding RQ3, analysis also involved two levels; however, whilst pupils remained at the first level, they were clustered by class instead of by school, as implementation varies between teachers, regardless of school-level clustering. In
multi-level analysis, the main issue is typically attaining an appropriate sample size at the group-level; although this is always smaller than the total sample, it is considered to be more important (Maas & Hox, 2005). Therefore, calculations need to be conducted to ascertain whether an appropriate group-level sample size has been achieved in the present study, to ensure that overfitting of the models does not occur (Austin & Steyerberg, 2015). One such way of establishing this is to calculate the ratio of subjects per variable (SPV). In the present study, the SPV ratio for RQ3 was 13.5 (54 classes, 4 variables) for year 3 implementation, and 11.5 (46 classes, 4 variables) for year 4 implementation. The minimum SPV required for accurate estimation of a regression model is contested in the literature, but the range generally lies between 10 and 20 (ibid); therefore, the present study is thought to have an adequate sample size for the number of predictor variables in the models for RQ3.

Given the exploratory nature of RQ3, calculations were conducted to confirm that there were a satisfactory number of predictor variables at the second level of the model. A recent study, of a similar design and sample, conducted to evaluate the PATHS intervention utilised a comparable two-level model, with implementation variables at the second level and pupils’ academic attainment at the first (Humphrey et al., 2017). The $f^2$ statistic, a measure of effect size suitable for regression models (Cohen, 1992), was found to be 0.087 for this model. This suggests that the implementation factors at the second level explained a significant proportion of the class-level variance; therefore, it could be concluded that this is an appropriate number of predictor variables to be included at the second level.

6.9 Qualitative Analytic Strategy

Thematic analysis was deemed to be the most suitable approach for the analysis of the qualitative data, in order to capture important aspects and identify patterned responses and meaning within the dataset. This form of analysis is thought to summarise key features of a large dataset and provide rich descriptions of it; it can also highlight similarities across the dataset and generate unanticipated
explanations for the data (Braun & Clarke, 2006). Although thematic analysis as a research method in its own right is contested in the literature (ibid), it was beneficial for this study as it allowed for the identification of common perceptions regarding the GBG, and any perceived impact on pupils’ school functioning. This includes how and why the GBG influenced at-risk children, and the reasons considered to underpin any perceived impact. Common justifications for any differential gains could also be explored.

As well as fitting well with a pragmatic mixed methods approach, the flexibility of thematic analysis allowed for a hybrid approach to be taken, meaning that data could be analysed both inductively and deductively. Deductive analysis was utilised with regards to pupils’ risk variables, as they were based on a priori variables that had been previously identified in the quantitative analysis. Analysis of the ways in which the GBG affected at-risk pupils’ outcomes was conducted inductively to allow for unexpected and emergent themes to develop (Nowell et al., 2017).

Braun and Clarke (2006) argued that thematic analysis “should be considered a method in its own right” (p.78) and thus developed a six-phase guide to performing thematic analysis. This guide was utilised during the analysis of the data and is outlined in table 6.15 below.

Table 6.15: Six-phase guide to thematic analysis (Braun & Clarke, 2006)

<table>
<thead>
<tr>
<th>Phase</th>
<th>Description of the Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data familiarisation</td>
<td>As interviews were transcribed externally, this stage involved reading and re-reading the data and noting possible codes.</td>
</tr>
<tr>
<td>Generating initial codes</td>
<td>As the analysis was both inductive and deductive, codes regarding risk variables were taken from the preceding quantitative analysis, whilst codes regarding mechanisms of change were searched for in the dataset. These data were coded in a</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Searching for themes</td>
<td>Collated codes were gathered into potential themes using broad headings such as “male pupils”. An “other” theme was created for codes that did not fit with a priori themes, thus allowing for further emergent themes to develop.</td>
</tr>
<tr>
<td>Reviewing themes</td>
<td>Emergent themes were reviewed, and data were re-coded where necessary.</td>
</tr>
<tr>
<td>Defining and naming themes</td>
<td>Codes in emergent themes were examined to ensure that they were representative of the themes; final names were then decided for themes.</td>
</tr>
<tr>
<td>Producing the report</td>
<td>The findings were written up in the results chapters, including interesting extracts linking back to the research questions and the previous literature.</td>
</tr>
</tbody>
</table>

### 6.10 Ethical Considerations

The University Ethics Committee granted ethical approval for the GBG trial, including this doctoral thesis, in May 2015 (ref: 15126).

#### 6.10.1 Informed Consent

Every parent with a child in a year 3 class received an information sheet prior to the collection of baseline data (see appendix 5). The letter outlined the aims of the research, who would be involved, and how and what data were being collected. Parents received these letters again before the second round of data collection. Consent was established through an opt-out form, thus anybody not completing this form consented to their child being involved in the research. In total, 108 pupils were opted-out of the study by their parents (3% of the sample).
For case study schools, teachers were provided with a similar information sheet explaining what they would be asked to do prior to the interviews beginning (see appendix 6). Parents of children selected for focus groups also received these information sheets (see appendix 7). Consent was established through an opt-in form, and so parents and teachers were required to sign the form in order to consent to being part of the qualitative aspect of the study.

For all pupils, assent was utilised, meaning that it was assumed that the children were happy to take part in the trial unless they expressed otherwise. For pupils in the case study focus groups, opt-in consent forms were provided (see appendix 8), meaning that they needed to sign the form in order to consent to be a part of the qualitative research.

6.10.2 Right to Withdraw

An opt-out consent form was sent to parents along with the information sheet (see appendix 5). This form required parents to sign and return it to the university (via freepost, email or telephone) if they did not wish for their child to be part of the study. This included having a teacher completing a survey on their child’s behalf. However, the opt-out form only removed them from the research project, these children would still be playing the GBG in their classroom (if in the intervention arm). The information sheet made clear that parents could withdraw their child from the study at any time, and parents were assured that any data already collected about their child would be destroyed.

For the case study schools, teachers who were selected for interviews were told that they could change their mind and withdraw at any time, without needing to give a reason, by contacting the researchers directly. Parents of pupils participating in the focus groups also received opt-in consent sheets (see appendix 7) explaining that they were free to withdraw their child from either the case study focus group, or the whole study, at any time, and that all data would be destroyed. Pupils in the focus groups were given an opt-in form, whereby they could consent to the
research by ticking the “smiley face”. They were also briefed by the researcher as to how the data would be used. They were given the opportunity to say that they did not want to participate in the focus group, at which point they could go back to their classroom.

Prior to the interviews and focus groups beginning, participants were asked again if they were still happy to take part, and for the interview to be recorded.

6.10.3 Anonymity and Confidentiality

Prior to the trial beginning, it was necessary for schools to provide some identifying information about the pupils, such as their name and date of birth, in order to match data correctly. Data were also collected from the NPD regarding information such as pupils’ prior attainment, SEND status and ethnicity. Having this information meant that pupils’ data could be matched across the different time points in the trial, and the number of risk variables each child was exposed to could be ascertained. Once the dataset was completed at the end of the two-year period, identifying information was destroyed so that children were anonymised.

For case study schools, teachers and pupils were given pseudonyms during transcription of the interviews, and recordings were deleted. Schools and pupils were given unique codes so that data could be matched across time points and participants could be matched to their schools.

Information sheets assured participants that any data collected from pupils or teachers would be treated as confidential. The website housing the surveys was secure and password protected. All survey data were stored on a secured network and was password protected. The network was only accessible to principal members of the research team. Schools and parents were also informed that the data from the study may be used to write reports for funders, academic journal articles, books, and doctoral theses, although these would not include any identifying information.
6.10.4 Incentives for Participation

No incentive was given in order for pupils, parents or teachers to agree to participate in the trial. However, in order to be eligible for randomisation for the GBG trial, schools were required to complete at least 90% of their baseline surveys, and to sign a Memorandum of Agreement with the EEF. As part of this, schools were asked to pay an application fee relevant to the size of their school (£750 per form entry), which contributed towards the cost of the GBG, training and resources (see appendix 9). Following randomisation, schools that were allocated to the UP arm of the trial received their payment back in stages in accordance with data collection (two-thirds at the beginning of the trial), as well as an additional £750 per form entry. Therefore, although there was no incentive given for schools to continue in the trial, withdrawing may have had financial implications, which could affect their decision. However, some GBG schools did cease implementation (see section 6.5.5), despite the cost. Case study schools, including the parents, pupils and individual teachers, did not receive any additional incentives for their participation.

6.10.5 Protection of Participants

Although every effort was made to minimise the risk of harm to participants, it is possible that teachers felt under pressure to complete their surveys as participation had cost implications for the school, or felt that they were being judged during the fieldwork visits. In order to address this, participants were reminded that they could withdraw at any time by contacting the researcher confidentially. Members of the research team also delivered a presentation at each of the GBG training events to explain the purposes of the fieldwork visits, and to reassure teachers that they were collecting data on the GBG, and not on the teachers’ abilities.

For case study schools, teachers and pupils were also given an additional chance to withdraw prior to the interviews beginning. At the end of the interviews,
participants were thanked and asked if they had any further questions about the trial or how their data would be used.

The GBG research team had all undergone an Enhanced Disclosure and Barring Service checks for the purposes of working within schools prior to fieldwork visits.

6.11 Chapter Summary

- The current study utilised data collected as part of a wider RCT of the GBG, but is independent from the wider trial and provides several distinct contributions to knowledge.
- The study utilised a hybrid concurrent embedded mixed methods design, to establish not just “what works”, but what works for whom and why, under certain conditions and circumstances.
- A total of 3084 pupils were involved in the study, although some attrition occurred over the course of the trial. Six schools were case studies for the qualitative strand.
- The TOCA-C was utilised to collect data on pupils’ disruptive behaviour, whilst the HGRT was utilised to obtain reading scores. Structured observation schedules and an online scoreboard were used to collect implementation data. Semi-structured interview and focus group schedules were developed for the qualitative strand of the study.
- Data were analysed using MLM and thematic analysis, and an overview of these is provided.
- A detailed account of the ethical considerations is outlined. Consent was sought from teachers, pupils and parents.
7. Quantitative Results

7.1 Chapter Overview

The following chapter presents the findings pertaining to the quantitative research questions of the present study. An analysis of missing data and the data assumptions for MLM are presented first, before moving on to the specific research questions. The risk variables associated with pupils’ school functioning are presented, followed by the calculations of pupils’ cumulative risk scores and an examination of the predictive power of cumulative risk. Finally, the effects of the GBG on outcomes for pupils at varying levels of risk and any differential associations with implementation variability are investigated.

7.2 Missing Data

One of the main problems in longitudinal studies is missing data (Twisk, 2006), particularly in the context of educational research, due to the constraints associated with conducting research in schools (Pampaka, Hutcheson, & Williams, 2016). Standard statistical procedures were developed for complete datasets (Peugh & Enders, 2004) and thus failure to account for missing data can spoil inferences (Heitjan & Basu, 1996). However, even though issues surrounding missing data are widely acknowledged, it is common practice to ignore missing data or utilise techniques that delete all cases with some missing data; King and colleagues (2001) commented that the use of listwise deletion to eliminate entire observations results in “a loss of valuable information at best and severe selection bias at worst” (p.49). This issue is particularly salient when data are not missing completely at random, as is often the case in educational research (Pampaka et al., 2016). Rubin (1976) outlined three “mechanisms” that can be used to explain why data are missing: missing completely at random (MCAR), whereby the missingness of an observed variable is unrelated to both other variables in the dataset and the underlying values of the observed variable itself; missing at random (MAR), whereby missingness may depend on other variables in the dataset, but not the
observed variable itself; or missing not at random (MNAR), whereby missingness of an observed variable depends on both the other variables in the dataset, and values of the observed variable itself (Heitjan & Basu, 1996).

Analysis of disruptive behaviour and reading attainment outcomes in the current study revealed that data were collected from all 77 schools in the trial, indicating 0% attrition at the school-level. However, at the pupil-level, 19.1% of cases were partially observed (either baseline or follow-up data available) for disruptive behaviour scores (80.1% complete cases; both baseline and follow-up data available) and 18.8% of cases partially observed for reading point scores (81.2% complete cases), with outcome data missing at either baseline or T2. Therefore, missing value analysis was conducted on the two pupil-level outcome variables to establish the mechanisms through which data were missing in the present study. Binary logistic regression identified five variables that predicted partially observed disruptive behaviour data (school size, school-level absence, school-level behaviour, pupil SEND status, and pupil looked-after status) and three variables that predicted partially observed reading attainment data (school-level SEND, pupil FSM status, and pupil SEND status). Therefore, missing data in the present study was considered to be MAR. While experts have not reached a consensus regarding the percentage at which missing data becomes problematic (Schlomer, Bauman, & Card, 2010), Bennett (2001) suggests that when more than 10% is missing, analyses are likely to be biased. Thus, in the present study, appropriate steps were taken to account for the missing data.

Multiple imputation (MI) is increasingly becoming established as the leading approach for analysing partially observed datasets when working under the assumption that data are MAR (Carpenter, Goldstein, & Kenward, 2011), as it has a strong theoretical framework and substantial evidence base (Peugh & Enders, 2004). Indeed, Pampaka and colleagues (2016) found that MI was a valuable technique, producing results comparable to complete datasets, even when over 60% of the data are missing. Utilising MI means that bias due to missingness is reduced, and that the sample size is preserved, hence maintaining statistical power. MI involves
replacing missing values utilising suitable estimates so that standard complete-data analysis methods can be applied to the dataset (ibid). Unlike other methods, MI creates a number of imputed datasets containing different plausible estimates for the missing values; the analysis is then performed on each of these datasets, before averaging the parameter estimates to produce a single set of results (Peugh & Enders, 2004).

Therefore, MI was chosen as the technique to address the missing data in the present study. REALCOM-Impute was the software chosen for conducting MI as, unlike some other packages, it allows for mixed response types (i.e. discrete and continuous data), as well as the multi-level structure of the models (Carpenter et al., 2011). Guidance produced by Carpenter and colleagues (ibid) was followed, and is discussed in more detail in section 7.6.3.

7.3 Data Assumptions for MLM

In order to be able to draw conclusions about the wider population from a sample, several data assumptions regarding MLMs must be met. Therefore, prior to data analysis, it is important that these assumptions are checked, and that any violations are highlighted (Field, 2005). As MLM is an extension of linear regression, all of the assumptions for standard linear regression also apply to multi-level analysis, with the addition of a further assumption relating to the random coefficients (Field, 2009; Twisk, 2006). Therefore, nine data assumptions were tested for the two models, disruptive behaviour and reading attainment. The dataset for RQ2 was utilised, as this research question, exploring the impact of the GBG on pupils’ at different levels of risk exposure, was considered to be the primary analysis being conducted for this study.

Table 7.1 below displays the data assumptions for MLM. As shown, the majority of the assumptions were met for at least one of the two models, with the exception of homoscedasticity. However, in large samples, the test for homoscedasticity, Levene’s test, can be significant even when group variances are not very different,
and thus this statistic should be interpreted with caution in the current study (Field, 2009). This means that the reading attainment model satisfied all assumptions; as two additional assumptions were not met for the disruptive behaviour model, this model was less robust than the reading attainment model.

**Table 7.1: Summary of data assumptions**

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Requirement*</th>
<th>Behaviour Model</th>
<th>Reading Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable Types</strong></td>
<td>All of the predictor variables are either quantitative or categorical, and the outcome variables are quantitative, continuous and unbounded.</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Non-Zero Variance</strong></td>
<td>The predictor variables have some variation in value, i.e. they do not have variances of zero (appendix 10a).</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Multicollinearity</strong></td>
<td>The Variance Inflation Factor (VIF) value must lie between 0 and 10.</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Homoscedasticity</strong></td>
<td>Levene’s test should not be statistically significant (appendix 10b).</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td><strong>Independent Errors</strong></td>
<td>The Durbin-Watson statistic should be between 1 and 3.</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Normally Distributed Error</strong></td>
<td>A histogram of standardised residuals should be normally distributed. Values on P-P plots of regression standardised residuals should fit along a</td>
<td>×</td>
<td>✓</td>
</tr>
</tbody>
</table>
All of the values of the outcome variables should be independent; in other words, they all come from a separate entity.

Standardised residuals should be plotted alongside the standardised predicted values; the overall shape of the scatterplot should be rectangular as opposed to curved (appendix 10d).

The Cook’s distance statistic should be less than one.

Histograms of the random intercepts should be normally distributed (appendix 10e).

*(Field, 2009)*

### 7.4 Risk Variables

The aim of this section is to answer research question 1 and one of the sub-questions within it. The research questions are:

1. What are the risk variables associated with lower school functioning?
   a. For those variables identified, what is the magnitude of their association with behavioural (disruptive behaviours) and/or academic (reading) outcomes?
7.4.1 Descriptive Statistics

Table 7.2: Means and standard deviations for outcome variables at baseline

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Mean Score</th>
<th>Standard Deviation</th>
<th>Minimum Score</th>
<th>Maximum Score</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disruptive Behaviour (TOCA-C)</td>
<td>1.66</td>
<td>0.81</td>
<td>1.00</td>
<td>5.78</td>
<td>1.799</td>
<td>3.593</td>
</tr>
<tr>
<td>Reading (KS1 Point Score)</td>
<td>15.19</td>
<td>3.80</td>
<td>3.00</td>
<td>24.00</td>
<td>-0.682</td>
<td>0.746</td>
</tr>
</tbody>
</table>

The minimum and maximum scores for disruptive behaviour on the TOCA-C are 0 and 6. Table 7.2 therefore shows that the mean score in the sample for disruptive behaviour was low relative to other studies; for example Koth and colleagues’ study (2009) of the TOCA-C found an average mean score of 2.25. However, the standard deviation does suggest some spread in the data. Indeed, whilst the value for skewness of the TOCA-C was within the acceptable range (+/-2; Field, 2009), the kurtosis value indicated that the distribution of the scores was leptokurtic. Further analysis of histograms (not reported here) supported this, demonstrating a peak in scores at the lower end of the scale.

Regarding reading attainment, point scores allocated to pupils following KS1 assessments can range between 3 and 27. The mean score of 15.19 in the present sample represented a reading ability in line with the NC expected standard. However, the standard deviation suggested some spread in the data, although assessment of the skewness and kurtosis values indicated that the data are normally distributed.
### Table 7.3: Correlations between predictor variables and outcome variables at baseline

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Scale of Measurement</th>
<th>Disruptive Behaviour</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SCHOOL-LEVEL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School Size</td>
<td>Number of pupils on roll</td>
<td>-.08**</td>
<td>-.08**</td>
</tr>
<tr>
<td>School Urbanicity</td>
<td>urban vs rural</td>
<td>.02</td>
<td>-.04</td>
</tr>
<tr>
<td>School FSM Eligibility</td>
<td>Proportion of pupils eligible for FSM</td>
<td>.01</td>
<td>-.11**</td>
</tr>
<tr>
<td>School EAL</td>
<td>Proportion of EAL pupils</td>
<td>.02</td>
<td>-.1**</td>
</tr>
<tr>
<td>School Absence</td>
<td>Proportion of pupils absent</td>
<td>-.033</td>
<td>-.08**</td>
</tr>
<tr>
<td>School SEND</td>
<td>Proportion of pupils registered as SEND</td>
<td>.08**</td>
<td>-.16**</td>
</tr>
<tr>
<td>School Achievement</td>
<td>Proportion of pupils achieving Level 4 English and Maths</td>
<td>.04*</td>
<td>.02</td>
</tr>
<tr>
<td>School Behaviour</td>
<td>Proportion of pupils in the borderline/abnormal range for conduct problems</td>
<td>.32**</td>
<td>-.05**</td>
</tr>
<tr>
<td><strong>PUPIL-LEVEL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Male vs female</td>
<td>.28**</td>
<td>-.17**</td>
</tr>
<tr>
<td>Relative Age</td>
<td>Summer born vs other</td>
<td>-.02</td>
<td>-.15**</td>
</tr>
<tr>
<td>FSM</td>
<td>Yes vs no</td>
<td>.12**</td>
<td>-.15**</td>
</tr>
<tr>
<td>Ethnic Minority</td>
<td>White vs other</td>
<td>-.07**</td>
<td>.01</td>
</tr>
<tr>
<td>EAL</td>
<td>Yes vs no</td>
<td>-.08**</td>
<td>-.06**</td>
</tr>
</tbody>
</table>
Correlations were conducted to support the inclusion of the predictor variables in the current study, along with evidence available in the extant literature (see section 2.5). To be included in the study, predictor variables needed to be correlated with at least one of the outcome variables, or have a clear theoretical rationale for inclusion. Table 7.3 shows that all of the predictor variables were significantly associated with at least one of the outcome variables, apart from school urbanicity, which failed to reach significance for either variable. As the vast majority of the variables demonstrated a relationship with either one or both of the outcome variables, they were all included in the study. Although school urbanicity was not a significant correlate, it was still investigated as a potential risk variable due to the limited and inconsistent findings regarding the effects of school location on pupil outcomes in the existing evidence base, alongside a clear theoretical rationale for its effects (see section 2.5.3.viii). Including it as a risk variable in the present study may help to aid clarity in this area.

Although all 15 of the variables were significant correlates, mostly at the $p < .01$ level, the strength of the relationships with the outcome variables varied. Most variables were only weakly correlated ($r = < +/-.30$; Cohen, 1992) with either disruptive behaviour or reading attainment, with the exception of SEND status which had a medium correlation with reading ($r = .52$; ibid). School-level behaviour and gender also had weak relationships with behaviour. These mostly very weak, but statistically significant, correlations with the outcome variables align with the rationale postulated in cumulative risk theory (see section 3.2), which suggests that

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Type of Relationship</th>
<th>Correlation (p &lt; .05)</th>
<th>Correlation (p &lt; .01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEND</td>
<td>Yes vs no</td>
<td>.25**</td>
<td>-.52**</td>
</tr>
<tr>
<td>Looked-After Child</td>
<td>Yes vs no</td>
<td>.09**</td>
<td>-.06**</td>
</tr>
<tr>
<td>IDACI Score</td>
<td>IDACI score for the child’s neighbourhood</td>
<td>.04*</td>
<td>-.14**</td>
</tr>
</tbody>
</table>

* Correlation significant at $p < .05$
** Correlation significant at $p < .01$

Pearson’s $r$ coefficients are displayed for the continuous predictor variables, whilst the categorical predictor variable coefficients reflect point biserial correlations.
the presence of no single risk variable is enough to cause serious detriment to a child’s outcomes. Instead, problems only begin to arise when the number of risk variables present in a child’s life begins to overwhelm their coping mechanisms. Furthermore, the large sample size in the present study means that there is a heightened sensitivity to detect effects (see section 6.5; ibid).

Further correlations (not reported here) were also run between each of the predictor variables. The results showed that ethnic minority status and EAL were highly correlated \( (r = .66, p < .01) \). Although this is just within the acceptable limit \( (r = .7; \text{Shieh & Fouladi, 2003}) \) for bivariate correlations regarding multicollinearity, it was decided that as it was a borderline case, and due to the high degree of overlap between the two variables (68% of ethnic minority pupils were also classified as EAL), the two would be combined to create one categorical variable in order to avoid violating an assumption of MLM. Thus, pupils were re-grouped to be categorised as either White & non-EAL, White EAL, ethnic minority & non-EAL, or ethnic minority & EAL.

### 7.4.2 Research Question 1 and 1a

1. **What are the risk variables associated with lower school functioning?**
   
a. **For those variables identified, what is the magnitude of their association with behavioural (disruptive behaviours) and/or academic (reading) outcomes?**

The dataset for this research question consisted of all pupils in the trial, in both the GBG and UP conditions, with pupils grouped by school. Two separate MLMs were utilised to answer this RQ, one for disruptive behaviour, and one for reading attainment. When utilising MLMs for statistical analysis, empty models are calculated first, before the addition of predictor variables, to establish the amount of unexplained variance in the outcome variable accounted for at each of the two levels. The amount of unexplained variance is measured using the ICC statistic, which provides the percentage of the variance in behaviour and reading outcomes.
explained at each level. In the empty models, school- and pupil-level data were added as predictor variables, with either disruptive behaviour scores or reading point scores added as outcome variables. The outcome variables were standardised by converting them to z scores prior to being entered into the model. Z scores represent the number of standard deviations above or the below the sample mean a particular score is, allowing for meaningful comparisons within and between models.

Empty models, including the ICC for each level, are presented below in table 7.4 for disruptive behaviour, and table 7.5 for reading attainment.

Table 7.4: Empty MLM for disruptive behaviour outcome variable

<table>
<thead>
<tr>
<th>Empty Model: Disruptive Behaviour</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>β0ij = 0.028(0.047)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2*log likelihood = 8148.686</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School-Level</td>
<td>0.146</td>
<td>0.027</td>
<td>&lt;.001**</td>
<td>14.4%</td>
</tr>
<tr>
<td>Pupil-Level</td>
<td>0.868</td>
<td>0.023</td>
<td>&lt;.001**</td>
<td>85.6%</td>
</tr>
</tbody>
</table>

Table 7.4\(^2\) shows that there was a significant association between disruptive behaviour scores and the school-level (μ0j = 0.146, p < .001) and the pupil-level (μ0j = 0.868, p < .001). The ICCs demonstrate that school-level variance accounted for 14.4% of the total variance in disruptive behaviour scores, whilst pupil-level variance accounted for the vast majority (85.6%) of the variance in pupils’ behavioural outcomes. The proportion of the variance in behavioural outcomes explained at the school-level is higher than in previous research in the field. For instance, Gutman and Feinstein (2008) found that the school-level accounted for 1-3% of the variance in outcomes; however, this difference may be accounted for by the use of pupil self-report in Gutman and Feinstein’s study.

\(^2\)N.B. While descriptive statistics were rounded to two decimal places, inferential statistics were rounded to three, to account for the smaller discrepancies between values.
Table 7.5: Empty MLM for reading attainment outcome variable

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School-Level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.056</td>
<td>0.013</td>
<td>&lt;.001**</td>
<td>5.6%</td>
</tr>
<tr>
<td><strong>Pupil-Level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.940</td>
<td>0.025</td>
<td>&lt;.001**</td>
<td>94.4%</td>
</tr>
</tbody>
</table>

Table 7.5 shows that there was a significant association between reading point scores and the school-level (μ0j = 0.056, p < .001) and the pupil-level (μ0j = 0.940, p < .001). The ICCs demonstrate that school-level variance accounted for 5.6% of the total variance in reading scores, whilst pupil-level variance accounted for the vast majority (94.4%) of the variance in pupils’ reading outcomes. The proportion of the variance in reading outcomes explained at the school-level is similar to previous research in field. For instance, Gutman and Feinstein (2008) found that the school-level accounted for 7-10% of the variance in outcomes.

The next stage of the analysis involved running two full models, whereby all of the predictor variables were added into the models to establish which of the variables were significant predictors of disruptive behaviour and/or reading attainment. The identification of significant predictors at either the school- or the pupil-level would thus demonstrate the risk variables for poor school functioning. Regarding behaviour problems, predictor variables that are significantly positively associated with the disruptive behaviour outcome variable are considered to be risk variables, as higher scores on the TOCA-C indicate more behavioural difficulties. However, for reading attainment, variables must be significantly negatively related to the reading point score to be considered risk variables, as higher point scores are indicative of better reading outcomes.
Similarly to running the empty models, standardised disruptive behaviour and reading point scores respectively were first added to each of the full models; predictor variables at the school- and pupil-level were then fitted. Continuous and binary variables were easily added to the models in their original state, whereas categorical variables with more than two categories needed to be allocated a reference category. Typically, the reference category is the group with the largest sample size (Steele, 2013), and so this was the approach taken with the current analysis. Only one of the predictor variables had more than two categories, ethnicity and EAL status, and so white & non-EAL was utilised as the reference category. The full models for behavioural and academic difficulties are displayed in tables 7.6 and 7.7 respectively.

Table 7.6: Full MLM for disruptive behaviour outcome variable

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Model: Disruptive Behaviour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{ij} = -0.257(0.326)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-2\times$log likelihood = 7142.054</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School-Level (ICC =4.2% )</td>
<td>0.034</td>
<td>0.009</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>School Size</td>
<td>-0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>School Urbanicity (if urban)</td>
<td>-0.047</td>
<td>0.159</td>
<td>.383</td>
</tr>
<tr>
<td>Proportion FSM Eligibility</td>
<td>0.001</td>
<td>0.003</td>
<td>.371</td>
</tr>
<tr>
<td>Proportion EAL</td>
<td>0.004</td>
<td>0.001</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Proportion Absence</td>
<td>-0.076</td>
<td>0.040</td>
<td>.031*</td>
</tr>
<tr>
<td>Proportion SEND</td>
<td>-0.008</td>
<td>0.005</td>
<td>.057</td>
</tr>
<tr>
<td>Proportion Level 4 English and Maths</td>
<td>-0.001</td>
<td>0.003</td>
<td>.371</td>
</tr>
<tr>
<td>Proportion Borderline/Abnormal Conduct Problems</td>
<td>0.029</td>
<td>0.003</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Pupil-Level</td>
<td>0.767</td>
<td>0.021</td>
<td>&lt;.001**</td>
</tr>
</tbody>
</table>
As is shown in table 7.6, several variables were predictors of poorer behaviour scores, as demonstrated by a significant, positive coefficient. At the school-level, the proportion of EAL pupils ($\beta_{0j} = 0.004, p < .001$), and the proportion of pupils in the school at-risk for conduct problems ($\beta_{0j} = 0.029, p < .001$) were significantly associated with higher disruptive behaviour scores at the pupil-level. The positive coefficients show that for every 1% increase in the proportion of EAL pupils and those with conduct problems, there were associated 0.004 and 0.029 standard deviation increases in the pupils’ disruptive behaviour scores respectively. This means that pupils demonstrated 0.003 and 0.02 higher scores on the TOCA-C for every 1% increase in the proportion of EAL pupils and those with conduct problems respectively.

At the pupil-level, there were four significant predictors of high disruptive behaviour scores, all of which were binary variables. Being male ($\beta_{0j} = 0.394, p < .001$) was associated with 0.394 standard deviation increase in disruptive behaviour scores, whilst being eligible for FSM ($\beta_{0j} = 0.216, p < .001$) was associated with a 0.216 increase. Furthermore, being registered as having any form of SEND ($\beta_{0j} = 0.505, p < .001$) emerged as a significant predictor, indicating a 0.505

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (if male)</td>
<td>0.394</td>
<td>0.034</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Relative Age (if summer born)</td>
<td>-0.102</td>
<td>0.038</td>
<td>.005**</td>
</tr>
<tr>
<td>FSM Eligibility (if eligible)</td>
<td>0.216</td>
<td>0.041</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Ethnicity &amp; EAL Category:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White &amp; non-EAL</td>
<td>◊</td>
<td>◊</td>
<td>◊</td>
</tr>
<tr>
<td>White EAL</td>
<td>-0.020</td>
<td>0.088</td>
<td>.409</td>
</tr>
<tr>
<td>Ethnic Minority &amp; non-EAL</td>
<td>-0.093</td>
<td>0.060</td>
<td>.063</td>
</tr>
<tr>
<td>Ethnic Minority &amp; EAL</td>
<td>-0.344</td>
<td>0.062</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>SEND (if yes)</td>
<td>0.505</td>
<td>0.045</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Looked-After Child (if yes)</td>
<td>0.732</td>
<td>0.205</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>IDACI Score</td>
<td>0.158</td>
<td>0.152</td>
<td>.298</td>
</tr>
</tbody>
</table>

◊ = reference category
standard deviation increase in disruptive behaviour. Finally, being registered as a looked-after child was also associated with higher levels of disruptive behaviour ($\beta_{0j} = 0.732, p < .001$); specifically, being looked-after was associated with a 0.732 standard deviation increase.

The coefficients for each of the variables discussed previously can be interpreted as standardised effect sizes; thus demonstrating each of the predictor variables relative association with pupils’ disruptive behaviour scores. Most of the variables only had small effects ($d < +/- 0.2$; Cohen, 1992), with the exception of SEND status and looked-after status, meaning that a large increase would need to be observed in the risk variables to evidence a small change in the disruptive behaviour score. Generally pupil-level risk variables have much larger effects than those variables at the school-level. This is further supported by a comparison of the school- and pupil-level ICCs, which show that only 4.2% of the total variance was accounted for by school-level variables.

Several of the predictor variables, namely school-level absence ($\beta_{0j} = -0.076, p = .031$), being summer born ($\beta_{0j} = -0.102, p < .005$), and being of both ethnic minority and EAL status ($\beta_{0j} = -0.344, p < .001$) were significantly negatively associated with the outcome variable, meaning that they were associated with lower disruptive behaviour scores. This contradicts previous research, suggesting that they were in fact promotive factors. However, as previously discussed (see section 2.4), it is beyond the scope of the present study to investigate both risk and promotive factors; and so these variables were not included in the cumulative risk analysis.
Table 7.7: Full MLM for reading attainment outcome variable

<table>
<thead>
<tr>
<th>Full Model: Reading Attainment</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>School-Level (ICC = 3%)</td>
<td>0.020</td>
<td>0.006</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>School Size</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td>School Urbanicity (if urban)</td>
<td>0.186</td>
<td>0.134</td>
<td>.084</td>
</tr>
<tr>
<td>Proportion FSM Eligibility</td>
<td>-0.000</td>
<td>0.002</td>
<td>.5</td>
</tr>
<tr>
<td>Proportion EAL</td>
<td>-0.002</td>
<td>0.001</td>
<td>.025*</td>
</tr>
<tr>
<td>Proportion Absence</td>
<td>-0.037</td>
<td>0.033</td>
<td>.133</td>
</tr>
<tr>
<td>Proportion SEND</td>
<td>-0.003</td>
<td>0.004</td>
<td>.4228</td>
</tr>
<tr>
<td>Proportion Level 4 English and Maths</td>
<td>-0.002</td>
<td>0.002</td>
<td>.16</td>
</tr>
<tr>
<td>Proportion Borderline/Abnormal Conduct Problems</td>
<td>0.003</td>
<td>0.002</td>
<td>.069</td>
</tr>
<tr>
<td>Pupil-Level (ICC = 97%)</td>
<td>0.640</td>
<td>0.017</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Gender (if male)</td>
<td>-0.204</td>
<td>0.031</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Relative Age (if summer born)</td>
<td>-0.265</td>
<td>0.033</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>FSM Eligibility (if eligible)</td>
<td>-0.178</td>
<td>0.037</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Ethnicity &amp; EAL Category:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White &amp; non-EAL</td>
<td>◯</td>
<td>◯</td>
<td>◯</td>
</tr>
<tr>
<td>White EAL</td>
<td>-0.456</td>
<td>0.079</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Ethnic Minority &amp; non-EAL</td>
<td>0.100</td>
<td>0.054</td>
<td>.034*</td>
</tr>
<tr>
<td>Ethnic Minority &amp; EAL</td>
<td>0.033</td>
<td>0.056</td>
<td>.279</td>
</tr>
<tr>
<td>SEND (if yes)</td>
<td>-1.178</td>
<td>0.041</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Looked-After Child (if yes)</td>
<td>-0.004</td>
<td>0.174</td>
<td>.492</td>
</tr>
<tr>
<td>IDACI Score</td>
<td>-0.469</td>
<td>0.136</td>
<td>.0005**</td>
</tr>
</tbody>
</table>
As is shown in table 7.7, several variables were predictors of poorer reading scores, as demonstrated by a significant negative coefficient. At the school-level, only one variable was a significant predictor of pupils’ poorer reading scores, the proportion of EAL pupils \( (\beta_{0j} = -0.002, p < .05) \). The negative coefficient shows that for every 1% increase in the proportion of EAL pupils, there was a 0.002 standard deviation decrease in pupils’ reading point scores. This means that pupils demonstrated 0.008 lower reading point scores for every 1% increase in the proportion of EAL pupils in the school.

At the pupil-level, there were six significant predictors of lower reading scores, four of which were binary, one continuous, and one categorical. Being male \( (\beta_{0j} = -0.204, p < .001) \) was associated with a 0.204 standard deviation decrease in pupils’ reading scores, whilst being born in the summer months \( (\beta_{0j} = -0.265, p < .001) \) equated to a 0.265 decrease. Being identified as having any form of SEND \( (\beta_{0j} = -1.178, p < .001) \) also emerged as a significant predictor, and was associated with a 1.178 standard deviation decrease in reading attainment. Furthermore, neighbourhood-level (IDACI score) and familial-level poverty (FSM eligibility) were significant predictors of reading attainment, meaning that they both explained unique variance. Whilst a 1% increase in IDACI score \( (\beta_{0j} = -0.469, p < .001) \) was associated with 0.469 standard deviation decrease in reading point scores, being eligible for FSM \( (\beta_{0j} = -0.178, p < .001) \) was associated with a 0.178 decrease.

Finally, one categorical variable was a significant predictor of reading scores, ethnicity and EAL status, although only one group acted as a risk variable, being white and speaking EAL \( (\beta_{0j} = -0.456, p < .001) \). Being a member of this risk group was associated with a 0.456 standard deviation decrease in reading point scores. However, being a member of an ethnic minority group and speaking English as a first language \( (\beta_{0j} = 0.100, p < .05) \) was associated with a 0.1 increase in reading point scores, making it a promotive factor. Being of both ethnic minority and EAL status was not a significant predictor.
Although two of the school-level variables, school urbanicity and school-level behaviour, were not significant (p = .084 and .069 respectively), in some studies, particularly those that are exploratory in nature, p values of <.1 would be considered to suggest a trend towards significance (Schumm et al., 2013), and thus school urbanicity and school-level behaviour would be deemed risk variables for reading attainment. However, the number of participants in the current sample makes the study robust and so the analysis would be sensitive enough to detect significance at the p<.05 level. Therefore, school urbanicity and school-level behaviour were not included as risk variables for reading attainment in the current study.

As mentioned previously, the coefficients for each of these variables can be interpreted in the same way as standardised effect sizes; thus demonstrating each of the predictor variables relative association with pupils’ reading point scores. Most of the variables only have small effects (d < +/− 0.2; Cohen, 1992), with the exception of IDACI score, meaning that a large increase would need to be observed in the risk variables to evidence a small change in reading attainment. The pupil-level risk variables had much larger effects than the school-level variable, proportion of EAL pupils, which evidenced an extremely small effect (-0.002). This was further supported by a comparison of the school and pupil-level ICCs, which showed that only 3% of the total variance was accounted for by school-level variables.

There were therefore seven risk variables associated with a decrease in reading attainment, and six risk variables associated with an increase in disruptive behaviour scores. These are summarised below in figure 7.1.
Figure 7.1: Significant risk variables for school functioning

- School-level behaviour
- Looked-after child

Both Outcomes
- School-level EAL
  - Male
  - FSM
  - SEND

Reading
- Summer Born
- White EAL
- Higher IDACI Score
7.5 Cumulative Risk

The aim of this section is to answer the remaining two sub-questions of research question 1. These questions are:

b. Is there evidence of a cumulative effect of risk exposure on these outcomes?

c. What is the functional form of the risk-outcome relationship?

d. Is the number of risk variables a better predictor of outcomes than their relative independent strength?

7.5.1 Cumulative Risk Score

A cumulative risk score was created for each pupil for behaviour and academic difficulties, based on the risk variables identified in RQ1 and RQ1a. In keeping with the extant literature (Appleyard et al., 2005; Gerard & Buehler, 2004a; Oldfield et al., 2015), firstly each of the previously identified risk variables were recoded dichotomously to either 1 or 0 to represent the presence or absence of the variable for each pupil. For binary and categorical variables, the group denoting risk was coded as 1, whilst all other categories were recoded to 0. For continuous variables, the criterion for risk status is a score that falls on or above the 75th percentile; thus the lowest, or poorest-performing, 25% of the sample were coded as 1, and the remaining 75% recoded as 0.

Next, the risk variables were summed to create two cumulative risk scores for each pupil in the sample, one for disruptive behaviour and one for reading attainment. Scores could range from 0 to 6 for behaviour, and 0 to 7 for reading, with higher scores indicating higher cumulative risk. More details regarding the cumulative risk scores are presented in the following section.
7.5.2 Descriptive Statistics

Table 7.8: Number and percentage of participants per risk level

<table>
<thead>
<tr>
<th>Risk Level</th>
<th>Behaviour</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Sample at Risk Level</td>
<td>Percentage of Sample at Risk Level</td>
</tr>
<tr>
<td>0</td>
<td>599</td>
<td>19.4%</td>
</tr>
<tr>
<td>1</td>
<td>1081</td>
<td>35.1%</td>
</tr>
<tr>
<td>2</td>
<td>818</td>
<td>26.5%</td>
</tr>
<tr>
<td>3</td>
<td>410</td>
<td>13.3%</td>
</tr>
<tr>
<td>4</td>
<td>112</td>
<td>3.6%</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>0.6%</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Missing</td>
<td>47</td>
<td>1.5%</td>
</tr>
<tr>
<td>Total</td>
<td>3084</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 7.8 displays the total number of participants at each risk level, for both disruptive behaviour and reading attainment. As can be seen in the table, the majority of the pupils had a risk score of 1 or 2, for both behaviour and attainment. As the number of risks increases, the number of participants at that risk level falls and hence the proportion of pupils at the highest levels is extremely small. Indeed, only 0.6% of the sample were exposed to five or more risk variables for disruptive behaviour, and 1.5% for reading attainment. This small sample size means that there is a possibility that any results could be skewed or spurious, and so the higher risk levels were combined prior to further analyses. As is consistent with previous literature (Oldfield et al., 2015), pupils with four or more risk variables were collapsed into a new category, “4+ risks”, to increase the sample size. The final numbers of pupils at each of the new risk levels are displayed in table 7.9.
Regarding the reading risk grouping, 232 pupils (7.5% of the sample) were missing, predominately due to missing dates of birth from the NPD. While there is currently no consensus in the literature regarding the point at which missing data becomes problematic (Schlomer et al., 2010), it has been suggested that analyses are only likely to be biased when over 10% of the data are missing (Bennett, 2001). Indeed, some even argue that missing data are only a problem over 20% (Peng et al., 2006). Thus, for the purposes of the current study (see section 7.2) a 10% cut-off was utilised, in line with Bennett’s recommendations (2001). Hence, missing data were not considered to be an issue for RQ1, and so MI was not employed for this research question.

<table>
<thead>
<tr>
<th>Risk Level</th>
<th>Behaviour</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proportion of Sample at Risk Level</td>
<td>Percentage of Sample at Risk Level</td>
</tr>
<tr>
<td>0</td>
<td>599</td>
<td>19.4%</td>
</tr>
<tr>
<td>1</td>
<td>1081</td>
<td>35.1%</td>
</tr>
<tr>
<td>2</td>
<td>818</td>
<td>26.5%</td>
</tr>
<tr>
<td>3</td>
<td>410</td>
<td>13.3%</td>
</tr>
<tr>
<td>4+</td>
<td>129</td>
<td>4.2%</td>
</tr>
<tr>
<td>Missing</td>
<td>47</td>
<td>1.5%</td>
</tr>
<tr>
<td>Total</td>
<td>3084</td>
<td>100%</td>
</tr>
</tbody>
</table>

As can be seen in table 7.10 below, following the combining of the risk categories, each pupil was now in a category of between 0 and 4 for both disruptive behaviour and reading attainment scores. The mean cumulative risk score for reading attainment at 1.75 was slightly higher than it was for disruptive behaviour, which was 1.47.
<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disruptive Behaviour</td>
<td>1.47</td>
<td>1.08</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Reading Attainment</td>
<td>1.75</td>
<td>1.15</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

7.5.3 Research Question 1b

b. Is there evidence of a cumulative effect of risk exposure on these outcomes?

In order to establish if there was evidence of a cumulative risk effect in the data, MLMs were utilised. Two models were created, one with disruptive behaviour scores as the outcome variable, and one with reading point scores. The cumulative risk score was then fitted to the models, as a continuous variable. Tables 7.11 and 7.12 show the results for disruptive behaviour and reading attainment.

Table 7.11: Cumulative risk MLM for disruptive behaviour outcome variable

<table>
<thead>
<tr>
<th>Cumulative Risk Model: Disruptive Behaviour</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{0ij} = -0.521(0.049)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2*log likelihood = 7701.906</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School-Level (ICC = 12.2%)</td>
<td>0.107</td>
<td>0.021</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Pupil-Level (ICC = 87.8%)</td>
<td>0.768</td>
<td>0.020</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>CUMULATIVE RISK SCORE</td>
<td>0.384</td>
<td>0.019</td>
<td>&lt;.001**</td>
</tr>
</tbody>
</table>

Table 7.11 above indicates that the cumulative risk score alone ($\beta_{0j} = 0.384$, $p < .001$) was a significant predictor of pupils’ disruptive behaviour; the positive coefficient shows that each additional risk was associated with an increase in pupils’ disruptive behaviour scores by 0.384 of a standard deviation. In other words, for each additional risk variable present, pupils’ disruptive behaviour scores on the
TOCA-C were 0.3 points higher. This means that the number of risk variables a child was exposed to was a predictor of disruptive behaviour. Results therefore support the first assumption of cumulative risk theory: the greater the number of risk variables, the greater the prevalence of problems.

Table 7.12: Cumulative risk MLM for reading attainment outcome variable

<table>
<thead>
<tr>
<th>Cumulative Risk Model: Reading Attainment</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{0j}$ = 0.697(0.039)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2*log likelihood = 7315.919</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| School-Level (ICC = 4.7%)                | 0.037        | 0.010          | <.001** |
| Pupil-Level (ICC = 95.3%)                | 0.752        | 0.020          | <.001** |
| CUMULATIVE RISK SCORE                    | -0.397       | 0.016          | <.001** |

The model displayed in table 7.12 indicates that the cumulative risk score alone ($\beta_{0j} = -0.397, p < .001$) was a significant predictor of pupils’ reading attainment; the negative coefficient shows that each additional risk was associated with a decrease in pupils’ reading point scores by 0.397 of a standard deviation. In other words, for each additional risk variable present, pupils’ reading point scores were 1.5 points lower. This means that the number of risk variables a child was exposed to was a predictor of reading attainment. Results therefore support the first assumption of cumulative risk theory noted above.
7.5.4 Research Question 1c

c. What is the functional form of the risk-outcome relationship?

Table 7.13: Means (standard errors) of outcome variables at baseline for each risk level

<table>
<thead>
<tr>
<th>Risk Level</th>
<th>Mean Baseline Behaviour Score (SE)</th>
<th>Mean Baseline Reading Score (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.31 (.018)</td>
<td>17.56 (.134)</td>
</tr>
<tr>
<td>1</td>
<td>1.50 (.020)</td>
<td>16.31 (.107)</td>
</tr>
<tr>
<td>2</td>
<td>1.80 (.031)</td>
<td>15.11 (.118)</td>
</tr>
<tr>
<td>3</td>
<td>2.15 (.054)</td>
<td>13.38 (.180)</td>
</tr>
<tr>
<td>4+</td>
<td>2.34 (.089)</td>
<td>11.78 (.244)</td>
</tr>
</tbody>
</table>

As is evident in table 7.13 above, an increase in the number of risk variables that a child is exposed to was associated with an increase in behavioural and academic difficulties. However, what is unclear is the form of this relationship. Thus to answer the research question, the MLMs from RQ1b were extended, in an attempt to establish whether the cumulative risk score had a linear or quadratic relationship with the two outcome variables.

One way that this has previously been done in the literature is by adding a squared term of the cumulative risk score (i.e. the quadratic term) into the model alongside the original cumulative risk score (i.e. the linear term) to establish which has the best overall model fit. If this squared term accounts for additional variance beyond the linear cumulative risk score, then a disproportionate relationship is present. However, prior to this analysis, it is necessary to mean centre the cumulative risk score before it is squared to avoid issues surrounding multicollinearity (Oldfield et al., 2015).

After mean centring and squaring the cumulative risk scores, they were added to the existing models from RQ1b as additional predictors to produce quadratic
models. The results for both disruptive behaviour and reading attainment outcomes are displayed in tables 7.14 and 7.15 below.

Table 7.14: Quadratic cumulative risk MLM for disruptive behaviour outcome variable

<table>
<thead>
<tr>
<th>Quadratic Cumulative Risk Model: Disruptive Behaviour</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>β0ij = -0.539(0.050)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2*log likelihood = 7697.134</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School-Level (ICC = 12.4%)</td>
<td>0.109</td>
<td>0.021</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Pupil-Level (ICC = 87.6%)</td>
<td>0.767</td>
<td>0.020</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>CUMULATIVE RISK SCORE</td>
<td>0.373</td>
<td>0.019</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>CUMULATIVE RISK SCORE SQUARED</td>
<td>0.028</td>
<td>0.013</td>
<td>.015*</td>
</tr>
</tbody>
</table>

The results displayed in the quadratic model in table 7.14 regarding disruptive behaviour scores show that the squared cumulative risk score was a significant predictor of disruptive behaviour \( (\beta_0j = 0.028, p < .05) \), explaining additional variance after accounting for the linear cumulative risk score \( (\beta_0j = 0.2373, p < .05) \). This suggests that the relationship between the risk score and disruptive behaviour score was non-linear. To confirm this, a likelihood-ratio test was conducted to establish which model had the best overall fit. A significant reduction in the log likelihood value from the linear model to the quadratic model was observed \( (x^2(1, n = 3084) = 4.772, p < .05) \), meaning that the quadratic relationship had the best model fit.
Table 7.15: Quadratic cumulative risk MLM for reading attainment outcome variable

<table>
<thead>
<tr>
<th>Quadratic Cumulative Risk Model: Reading Attainment</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>β0ij = 0.714(0.040)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2*log likelihood = 7312.580</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School-Level (ICC = 4.7%)</td>
<td>0.037</td>
<td>0.009</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Pupil-Level (ICC = 95.3%)</td>
<td>0.752</td>
<td>0.020</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>CUMULATIVE RISK SCORE</td>
<td>-0.391</td>
<td>0.016</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>CUMULATIVE RISK SCORE SQUARED</td>
<td>-0.021</td>
<td>0.011</td>
<td>.03*</td>
</tr>
</tbody>
</table>

The results displayed in the quadratic model in table 7.15 regarding reading point scores show that the squared cumulative risk score coefficient was a significant predictor of reading attainment (β0j = -0.021, p < .05), explaining additional variance after accounting for the linear cumulative risk score (β0j = -0.391, p < .05). This suggests that the relationship between the risk score and reading scores was also non-linear. To confirm this, a likelihood-ratio test was conducted. Whilst the log likelihood value did reduce from the linear to the quadratic model, this difference was not statistically significant (χ²(1, n = 3084) = 3.339, p = .068). However, it was extremely close to significance, suggesting that the functional form of the relationship between cumulative risk and reading attainment was borderline linear/non-linear.

The graphs displayed in figures 7.2 and 7.3 provide a further visual representation of the relationships between cumulative risk score and the two outcome variables. As can be seen in the graphs, both lines are indicative of a mass accumulation effect, as demonstrated by their deviation from the straight trend line. In particular, there are two distinct elbow points on the disruptive behaviour graph, at exposure to one and three risk variables respectively. Although the line on the reading attainment
graph is closer to the straight trend line, there is still evidence of one elbow point at exposure to two risk variables. This confirms that while indeed the functional form of this risk-outcome relationship for reading attainment was borderline, there is evidence to suggest the relationship was tending towards quadratic, and so is treated as such in further analyses.

*Figure 7.2: Functional form of risk-outcome relationship for disruptive behaviour*
7.5.5 Research Question 1d

\textit{d. Is the number of risk variables a better predictor of outcomes than their relative independent strength?}

In order to establish if the cumulative risk model had superior power in predicting pupils’ outcomes, relative to the predictive power of the individual risk variables, multi-level models were utilised. Two models were created, one with disruptive behaviour scores as the outcome variable, and one with reading point scores. The dichotomised versions of the predictor variables that were found to be significant risk variables in the models from RQ1 and 1a were fitted to the new model, along with the squared cumulative risk score evidenced as a significant predictor of outcomes for both disruptive behaviour and reading attainment in RQ1c. Table 7.16 and 7.17 show the results for disruptive behaviour and reading attainment.
The model displayed in table 7.16 indicates that the squared cumulative risk score was a significant predictor of pupils’ disruptive behaviour scores ($\beta_{0j} = 0.024$, $p < .05$), even after accounting for the variance explained by the individual risk variables. Results therefore support the second assumption of cumulative risk theory: the accumulation of risk variables is more important than the presence or absence of particular risk variables.
Table 7.17: Cumulative risk and risk variables model for reading outcome variable

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cumulative Risk Full Model: Reading Attainment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_{0ij} = 0.537(0.036))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-2*\text{log likelihood} = 6850.891)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>School-Level</strong></td>
<td>0.025</td>
<td>0.007</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>(ICC = 3.8%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High Proportion EAL</strong></td>
<td>-0.186</td>
<td>0.062</td>
<td>.002**</td>
</tr>
<tr>
<td><strong>Pupil-Level</strong></td>
<td>0.641</td>
<td>0.017</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>(ICC = 96.2%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong> (if male)</td>
<td>-0.196</td>
<td>0.031</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>Relative Age</strong> (if summer born)</td>
<td>-0.262</td>
<td>0.034</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>FSM Eligibility</strong> (if eligible)</td>
<td>-0.214</td>
<td>0.037</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>White EAL</strong> (vs other)</td>
<td>-0.505</td>
<td>0.078</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>SEND</strong> (if yes)</td>
<td>-1.201</td>
<td>0.041</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>High IDACI Score</strong></td>
<td>-0.073</td>
<td>0.040</td>
<td>.036*</td>
</tr>
<tr>
<td><strong>CUMULATIVE RISK SCORE SQUARED</strong></td>
<td>0.020</td>
<td>0.011</td>
<td>.038*</td>
</tr>
</tbody>
</table>

The model displayed in table 7.17 indicates that the squared cumulative risk score was a significant predictor of pupils’ reading point scores \(\beta_{0j} = 0.020, p <.05\), even after accounting for the variance explained by the individual risk variables. Results therefore support the second assumption of cumulative risk theory noted above.

7.5.6 Summary of RQ1

- Regarding disruptive behaviour, a total of six risk variables were found to be significantly associated with increased disruptive behaviour scores, two at the school-level, and four at the pupil-level.
• Regarding reading attainment, a total of seven variables were found to be significantly associated with lower reading point scores, one at the school-level, and six at the pupil-level.

• Pupils’ cumulative risk scores were a predictor of their disruptive behaviour and reading attainment, with higher levels of risk associated with higher disruptive behaviour scores and lower reading point scores.

• Cumulative risk scores also accounted for additional variance over and above the variance explained by the individual risk variables.

• The squared term of the cumulative risk scores indicated that the relationships between cumulative risk scores and disruptive behaviour and reading point scores were quadratic.
7.6 Good Behaviour Game

The aim of this section is to answer research question 2. The research question is:

2. Are there differential intervention gains in behavioural and/or academic outcomes among children at different levels of cumulative risk exposure?

7.6.1 Risk Grouping

Prior to analysis, each pupil was categorised into a risk group for disruptive behaviour and reading attainment. To establish these risk groups, calculations of effect size were conducted to determine if there were any notable differences in mean disruptive behaviour and reading point scores between risk levels (see table 7.18).

Similar differences were found between each of the risk levels for disruptive behaviour, with effect sizes ranging from .35 to .39, except between risk levels 3 and 4+, where the change in scores was smaller (d=.19). This is consistent with the cumulative risk graph (see figure 7.4), where an elbow point is visible after exposure to three risk variables. As the increases in disruptive behaviour scores between each of the risk levels were typically of similar proportions, categorisation was done in accordance with the elbows present on the cumulative risk graph, accounting for the exponential nature of the relationship. This means that pupils were split where a visible change in disruptive behaviour scores occurred. Pupils were therefore categorised into one of three risk groups (see table 7.19): the low-risk group, consisting of pupils exposed to either zero or one risk variables, the medium-risk group, consisting of pupils with either two or three risk variables, or the high-risk group, consisting of pupils exposed to four or more risk variables.
Table 7.18: Change in disruptive behaviour scores between risk levels

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Effect Size (Cohen’s d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>0.35</td>
</tr>
<tr>
<td>1-2</td>
<td>0.39</td>
</tr>
<tr>
<td>2-3</td>
<td>0.36</td>
</tr>
<tr>
<td>3-4+</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Figure 7.4: Cumulative risk graph for disruptive behaviour, with elbow points

Table 7.19: Number of pupils per risk group for disruptive behaviour, by trial group

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>N Risk</th>
<th>Total N</th>
<th>% of the Sample</th>
<th>N GBG</th>
<th>N UP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-Risk</td>
<td>0+1</td>
<td>1638</td>
<td>56</td>
<td>762</td>
<td>876</td>
</tr>
<tr>
<td>Medium-Risk</td>
<td>2+3</td>
<td>1181</td>
<td>40</td>
<td>644</td>
<td>537</td>
</tr>
<tr>
<td>High-Risk</td>
<td>4+</td>
<td>119</td>
<td>4</td>
<td>82</td>
<td>37</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2938</td>
<td>100</td>
<td>1488</td>
<td>1450</td>
</tr>
</tbody>
</table>
Categorising the pupils in this way increased the sample sizes of each group, meaning that the analysis was suitably statistically powered. In order to ensure this, calculations to determine the MDES were conducted. Both the low and medium-risk group had over 1,000 pupils in total (see table 7.19), and had MDESs of .17 and .20 respectively (see section 6.8.3.ii). The high-risk group was small in comparison, with only 129 pupils in this group, and so the MDES was .45. In spite of this, the groups were categorised in this way, in order to be consistent with the smaller effect size difference between risk levels 3 and 4+, and the visible relationship displayed on the cumulative risk graph. Furthermore, as the GBG is expected to have the largest effect on pupils in the medium-risk group (Muthén et al., 2002), the MDES for the high-risk group was not considered to be as important.

Regarding the creation of reading attainment risk groups (see table 7.20), whilst there were similar, small differences in reading point scores between the majority of the risk levels, larger differences were evident between risk levels 0 and 1 (d=.43), and 2 and 3 (d=.56). This is consistent with the cumulative risk graph (figure 7.5), where an elbow point is visible after exposure to two risk variables. Pupils were therefore categorised into one of four groups, in line with the findings resulting from the effect size calculations (see table 7.21): the no risk group, consisting of pupils exposed to zero risk variables, the low-risk group, consisting of pupils with one or two risk variables, the medium-risk group, consisting of pupils exposed to three risk variables, or the high-risk group, consisting of pupils with four or more risk variables. Similarly to above, sample sizes were increased (see table 7.21), and calculations to determine the MDESs were conducted (see section 6.8.3). All subgroups were powered to detect small effect sizes.
Table 7.20: Change in reading point scores between risk levels

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Effect Size (Cohen’s d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>0.43</td>
</tr>
<tr>
<td>1-2</td>
<td>0.37</td>
</tr>
<tr>
<td>2-3</td>
<td>0.56</td>
</tr>
<tr>
<td>3-4+</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Figure 7.5: Cumulative risk graph for reading attainment, with elbow points

Table 7.21: Number of pupils per risk group for reading attainment, by trial group

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>N Risk Variables</th>
<th>Total N Pupils</th>
<th>% of the Sample</th>
<th>N GBG Pupils</th>
<th>N UP Pupils</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Risk</td>
<td>0</td>
<td>406</td>
<td>14.3</td>
<td>215</td>
<td>191</td>
</tr>
<tr>
<td>Low-Risk</td>
<td>1+2</td>
<td>1702</td>
<td>60.0</td>
<td>803</td>
<td>899</td>
</tr>
<tr>
<td>Medium-Risk</td>
<td>3</td>
<td>495</td>
<td>17.5</td>
<td>270</td>
<td>225</td>
</tr>
<tr>
<td>High-Risk</td>
<td>4+</td>
<td>232</td>
<td>8.2</td>
<td>125</td>
<td>107</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2835</td>
<td>100</td>
<td>1413</td>
<td>1422</td>
</tr>
</tbody>
</table>
7.6.2 Descriptive Statistics

Table 7.22: Mean scores for behaviour outcome variable, by trial group allocation

<table>
<thead>
<tr>
<th></th>
<th>GBG Sample</th>
<th>UP Sample</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Schools</td>
<td>38</td>
<td>39</td>
<td>77</td>
</tr>
<tr>
<td>N Pupils T1</td>
<td>1497</td>
<td>1469</td>
<td>2966</td>
</tr>
<tr>
<td>N Pupils T2</td>
<td>1202</td>
<td>1310</td>
<td>2512</td>
</tr>
<tr>
<td>Mean (SE)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disruptive Behaviour Score (T1)</td>
<td>1.71(0.21)</td>
<td>1.61(0.21)</td>
<td>1.66(0.81)</td>
</tr>
<tr>
<td>Disruptive Behaviour Score (T2)</td>
<td>1.74(0.02)</td>
<td>1.65(0.02)</td>
<td>1.69(0.17)</td>
</tr>
</tbody>
</table>

Table 7.22 above demonstrates the pre- and post-test mean scores for disruptive behaviour for pupils in each of the trial conditions. The table indicates the number of pupils who provided data at the two time points; as can be seen, although disruptive behaviour data were collected from all schools, the number of pupils providing data fell between T1 and T2. Table 7.22 also indicates that mean disruptive behaviour scores increased in both trial conditions, suggesting that pupils’ behaviour got marginally worse over the course of the trial. However, mean disruptive behaviour scores were lower in the UP condition at both time points.

Table 7.23: Mean scores for behaviour variable, by risk and trial group allocation

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Mean Disruptive Behaviour Scores (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GBG T1</td>
</tr>
<tr>
<td>Low-Risk</td>
<td>1.46(0.21)</td>
</tr>
<tr>
<td>Medium-Risk</td>
<td>1.91(0.37)</td>
</tr>
<tr>
<td>High-Risk</td>
<td>2.37(0.10)</td>
</tr>
</tbody>
</table>

Table 7.23 above demonstrates the pre- and post-test mean scores for disruptive behaviour for pupils in each of the three risk groups, in each of the trial conditions.
The table indicates that disruptive behaviour scores increased as risk group did, both at baseline and T2. Pupils in the UP condition also typically had lower disruptive behaviour scores at baseline, compared to pupils in the GBG condition. All mean disruptive behaviour scores increased over the course of the trial, except for the high-risk pupils who were exposed to the GBG, where the mean score fell by 0.36.

**Table 7.24: Mean scores for reading outcome variable, by trial group allocation**

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>GBG Sample</th>
<th>UP Sample</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Schools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBG T1</td>
<td>1533</td>
<td>1481</td>
<td>3014</td>
</tr>
<tr>
<td>GBG T2</td>
<td>1264</td>
<td>1255</td>
<td>2519</td>
</tr>
<tr>
<td>Mean (SE)</td>
<td>15.31(0.10)</td>
<td>15.72(0.10)</td>
<td>15.19(3.80)</td>
</tr>
</tbody>
</table>

Table 7.24 above demonstrates the pre- and post-test mean scores for reading attainment for pupils in each of the trial conditions. The table indicates the number of pupils who provided data at the two time points; as can be seen, although reading attainment data were collected from all schools, the number of pupils providing data fell between T1 and T2. Table 7.25 also indicates that mean reading scores were higher in the UP condition at both time points.

**Table 7.25: Mean scores for reading variable, by risk and trial group allocation**

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>GBG T1 - KS1</th>
<th>GBG T2 - HGRT</th>
<th>UP T1 - KS1</th>
<th>UP T2 - HGRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Risk</td>
<td>17.51(0.20)</td>
<td>37.17(0.66)</td>
<td>17.55(0.21)</td>
<td>38.44(0.65)</td>
</tr>
<tr>
<td>Low-Risk</td>
<td>15.80(0.12)</td>
<td>34.20(0.37)</td>
<td>16.06(0.12)</td>
<td>33.75(0.36)</td>
</tr>
<tr>
<td>Medium-Risk</td>
<td>13.65(0.26)</td>
<td>28.65(0.70)</td>
<td>14.25(0.28)</td>
<td>28.94(0.81)</td>
</tr>
<tr>
<td>High-Risk</td>
<td>12.10(0.36)</td>
<td>24.57(1.00)</td>
<td>12.42(0.43)</td>
<td>25.35(1.08)</td>
</tr>
</tbody>
</table>
Table 7.25 demonstrates the pre- and post-test mean scores for reading attainment for pupils in each of the four risk groups, in each of the trial conditions. The table indicates that reading point scores decreased as risk group increased, both at baseline and T2. Typically, pupils in the UP condition had higher reading point scores at both time points, except for pupils in the low-risk group, where pupils exposed to the GBG had slightly higher reading point scores at T2.

7.6.3 Research Question 2

2. Are there differential intervention gains in behavioural and/or academic outcomes among children at different levels of cumulative risk exposure?

The dataset for this research question consisted of all pupils in the trial, in both the GBG and UP conditions. Pupils were grouped by school, with their T2 disruptive behaviour and reading point scores added to the dataset. In order to establish if there was evidence of differential effects of the GBG on pupils at different levels of risk exposure, MLMs were utilised. Two models were created, one with disruptive behaviour scores as the outcome variable, and one with reading point scores. Prior to creating the MLMs, the outcome variables were standardised by converting them to z scores. Similarly to RQ1, empty models were calculated first and are presented below in table 7.26 for disruptive behaviour, and table 7.27 for reading attainment.

Table 7.26: Empty MLM for disruptive behaviour outcome variable at T2

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>School-Level</td>
<td>0.084</td>
<td>0.019</td>
<td>&lt;.001**</td>
<td>8.4%</td>
</tr>
<tr>
<td>Pupil-Level</td>
<td>0.918</td>
<td>0.026</td>
<td>&lt;.001**</td>
<td>91.6%</td>
</tr>
</tbody>
</table>
Table 7.26 shows that there was a significant school-level ($\mu_0j = 0.084, p < .001$) and pupil-level ($\mu_0j = 0.918, p < .001$) effect on disruptive behaviour scores at T2. The ICCs demonstrate that school-level variance accounted for 8.4% of the total variance in disruptive behaviour scores, whilst pupil-level variance accounted for the vast majority (91.6%) of the variance in pupils’ behavioural outcomes.

Table 7.27: Empty MLM for reading attainment outcome variable at T2

<table>
<thead>
<tr>
<th>Empty Model: Reading Attainment</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School-Level</strong></td>
<td>0.084</td>
<td>0.019</td>
<td>&lt;.001**</td>
<td>8.4%</td>
</tr>
<tr>
<td><strong>Pupil-Level</strong></td>
<td>0.912</td>
<td>0.026</td>
<td>&lt;.001**</td>
<td>91.6%</td>
</tr>
</tbody>
</table>

Table 7.27 shows that there was a significant school-level ($\mu_0j = 0.084, p < .001$) and pupil-level ($\mu_0j = 0.912, p < .001$) effect on reading attainment at T2. The ICCs demonstrate that school-level variance accounted for 8.4% of the total variance in reading point scores, whilst pupil-level variance accounted for the vast majority (91.6%) of the variance in pupils’ reading outcomes.

The next stage of the analysis involved running two full models, one with disruptive behaviour scores as the outcome variable, and one with reading point scores, to establish any effects of the GBG on pupils’ at different levels of risk exposure. Similarly to running the empty models, standardised disruptive behaviour and reading point scores respectively were first added to each of the full models, before adding either baseline disruptive behaviour and reading point scores, to control for pupils’ prior behaviour and attainment. Trial group allocation and risk group were also then added to determine any effects of the GBG on pupils’ outcomes. Finally, to identify any differential effects of the GBG on pupils at different levels of risk exposure, interaction terms between the trial group and risk group were then added.
Regarding behaviour problems, interaction terms that were significantly negatively associated with the disruptive behaviour outcome variable would suggest that exposure to the GBG had resulted in a reduction in disruptive behaviour, as lower scores on the TOCA-C indicate fewer behavioural difficulties. For reading attainment, interaction terms that were significantly positively associated with the reading point score outcome variable would suggest that exposure to the GBG had resulted in an increase in reading attainment, as higher point scores are indicative of better reading outcomes.

Due to the amount of missing data, MI procedures were conducted (see section 7.2) to maintain the sample size, thus reducing the bias associated with attrition and allowing for the use of techniques designed for complete data (Pampaka et al., 2016). MI was conducted in REALCOM-Impute. Pupil gender, trial group allocation and the constant were added as auxiliary variables, whilst school and pupil demographics utilised in RQ1 as risk variables were added as outcome variables. Pupils’ T1 and T2 TOCA-C scores (all subscales) and reading point scores were also added as response variables. REALCOM-Impute default settings of 1000 iterations, a burn-in of 100 and a refresh of 10 were utilised, in accordance with guidance produced by Carpenter and colleagues (2011) for multi-level imputation with mixed response types (Humphrey et al., 2017).

The multiply imputed full models are shown in table 7.28 for disruptive behaviour and table 7.29 for reading attainment below. Whilst complete case analyses were also performed (see appendix 11), only the imputed models are shown here, as there were no substantive differences in findings.
As shown in table 7.28 above, the multi-level analysis revealed that the GBG did not impact upon pupils’ disruptive behaviour ($\beta_{0j} = 0.024, p = .390$). There were also no statistically significant differential effects of the GBG for pupils in any of the risk groups (medium-risk: $\beta_{0j} = 0.073, p = .146$). Although descriptive statistics indicated lower mean disruptive behaviour scores for pupils in the high-risk group, this reduction was not found to be significant ($\beta_{0j} = -0.191, p = .142$).
### Table 7.29: Imputed MLM for differential effects on pupils’ reading attainment

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Model: Reading Attainment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β0ij = -3.262(0.093)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>School-Level (ICC = 14.3%)</strong></td>
<td>0.062</td>
<td>0.012</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>Trial Group (if GBG)</strong></td>
<td>-0.056</td>
<td>0.089</td>
<td>.265</td>
</tr>
<tr>
<td><strong>Pupil-Level (ICC = 85.7%)</strong></td>
<td>0.373</td>
<td>0.011</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>T1 KS1 Reading Point Score</strong></td>
<td>0.211</td>
<td>0.004</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>Risk Group:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Risk</td>
<td>◯</td>
<td>◯</td>
<td>◯</td>
</tr>
<tr>
<td>Low-Risk</td>
<td>-0.029</td>
<td>0.054</td>
<td>.296</td>
</tr>
<tr>
<td>Medium-Risk</td>
<td>-0.025</td>
<td>0.072</td>
<td>.365</td>
</tr>
<tr>
<td>High-Risk</td>
<td>0.006</td>
<td>0.093</td>
<td>.474</td>
</tr>
<tr>
<td>GBG*Low-Risk</td>
<td>0.115</td>
<td>0.075</td>
<td>.070</td>
</tr>
<tr>
<td>GBG*Medium-Risk</td>
<td>0.039</td>
<td>0.097</td>
<td>.345</td>
</tr>
<tr>
<td>GBG*High-Risk</td>
<td>-0.055</td>
<td>0.129</td>
<td>.322</td>
</tr>
</tbody>
</table>

As shown in table 7.29 above, the multi-level analysis revealed that the GBG did not impact upon pupils’ reading attainment ($β0j = -0.056, p = .265$). There were also no statistically significant differential effects of the GBG for pupils in any of the risk groups (medium-risk: $β0j = 0.039, p = .345$; high-risk: $β0j = -0.055, p = .474$). However, there was a marginal trend indicative of higher reading point scores for low-risk pupils in the GBG condition ($β0j = 0.115, p = .070$). This is in line with the descriptive statistics outlined in table 7.25 previously, which indicated that pupils in the low-risk group exposed to the GBG had slightly higher mean reading scores (34.20) at T2 than pupils in the UP condition (33.75).
7.6.4 Summary of RQ2

- Pupils were categorised into one of three risk groups for disruptive behaviour, and into one of four risk groups for reading attainment.
- There were no statistically significant subgroup effects of the GBG on pupils’ disruptive behaviour or reading attainment.
  - However, there was a trend towards significance indicative of higher reading point scores for low-risk pupils exposed to the GBG.
7.7 Implementation

The aim of this section is to answer research question 3:

3. Are there any differential associations between implementation of the GBG (specifically fidelity, quality, dosage) and behavioural and/or academic outcomes for at-risk children?

7.7.1 Descriptive Statistics

As discussed in section 6.6.1.iii, fidelity and quality were treated as one combined variable for the analyses (FidQual), with teachers receiving a percentage score between zero and 100. Regarding dosage, in line with Warren, Fey and Yoder’s method (2007), pupils’ total exposure to the game was measured in minutes, calculated utilising data collected from teachers’ online scoreboards.

Due to the number of teachers ceasing implementation across the two years of the trial (see section 6.5.5), analyses was conducted separately for each year, to minimise bias resulting from missing data. Table 7.30 below shows the number of classes implementing the GBG in each of the two years of the trial, and the mean scores for FidQual, and dosage. As can be seen in table 7.30, 54 of the 60 year 3 classes were observed playing the GBG during the first year of the trial, whilst 45 of the 60 year 4 classes were observed in the second year of the trial. In total, 100 classes were observed from 35 schools over the course of the trial. FidQual scores were consistent across the trial, with mean scores at approximately 70% (year 3: 69.79%; year 4: 70.11%).

Online scoreboard data were collected from 48 of the 60 year 3 classes in the first year of the trial, and 39 of the 60 year 4 classes in the second year. In total, dosage data were collected from 87 classes from the 35 schools over the course of the trial. Dosage saw an increase from the first year to the second year of the trial, from
approximately nine hours over the year (539.16 minutes) to approximately 11 hours over the year (657.3 minutes).

**Table 7.30: Number of classes and mean (SD) scores for implementation**

<table>
<thead>
<tr>
<th>Implementation Variable</th>
<th>Year 3 Classes</th>
<th>Year 4 Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Classes</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>FidQual (%)</td>
<td>54</td>
<td>69.79 (12.35)</td>
</tr>
<tr>
<td>Dosage (total number of minutes)</td>
<td>48</td>
<td>539.16 (352.24)</td>
</tr>
</tbody>
</table>

Prior to analysis, FidQual scores and dosage (number of minutes) were both categorised, producing high/low FidQual scores and high/low dosage scores for each teacher, for each year of the trial. Consistent with previous implementation fidelity literature, FidQual scores were categorised using an external cut-off of 80% (Savignac & Dunbar, 2014), with teachers scoring above 80% categorised as high FidQual, and those below 80% categorised as low FidQual. Regarding dosage, no external standardised cut-off exists, as required dosage levels are unique to individual interventions; dosage in the GBG can also vary by both length and frequency of games, and is expected to evolve over time. Therefore, teachers’ dosage was categorised using the 75th percentile as a cut-off; teachers were ranked by total number of minutes played in each year group, with those ranking in the top 25% being categorised as high dosage and the rest categorised as low dosage.

Sample sizes and mean scores for each category are displayed in table 7.31 below. As shown, similarly to the overall means, FidQual scores were fairly consistent across both years of the trial for the low and high groups. However, whilst dosage increased overall, and also increased in the high dosage category, the mean number of minutes for the low dosage group fell during the second year of the trial from
approximately eight hours (489.63 minutes) to approximately seven hours (433.39 minutes).

Table 7.31: Sample size (classes) and mean scores (SE) for implementation

<table>
<thead>
<tr>
<th></th>
<th>Year 3 Classes</th>
<th></th>
<th>Year 4 Classes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Mean (SE)</td>
<td>N</td>
<td>Mean (SE)</td>
</tr>
<tr>
<td>FidQual (%)</td>
<td>43</td>
<td>65.88 (0.32)</td>
<td>11</td>
<td>84.95 (0.18)</td>
</tr>
<tr>
<td>Dosage (total number of minutes)</td>
<td>35</td>
<td>489.63 (14.86)</td>
<td>13</td>
<td>1154.0 (3.89)</td>
</tr>
</tbody>
</table>

As the analysis for RQ3 relates to the interaction between implementation category and risk group, sample sizes for each of the risk groups, for both behaviour and reading, are also reported in the tables below.

Table 7.32: Sample sizes for implementation groups, by risk group for behaviour

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Year 3</th>
<th></th>
<th>Year 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FidQual</td>
<td>Dosage</td>
<td>FidQual</td>
<td>Dosage</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Low-Risk</td>
<td>548</td>
<td>148</td>
<td>443</td>
<td>150</td>
</tr>
<tr>
<td>Medium-Risk</td>
<td>429</td>
<td>151</td>
<td>407</td>
<td>151</td>
</tr>
<tr>
<td>High-Risk</td>
<td>71</td>
<td>9</td>
<td>52</td>
<td>29</td>
</tr>
<tr>
<td>Total</td>
<td>1048</td>
<td>308</td>
<td>902</td>
<td>330</td>
</tr>
</tbody>
</table>

Table 7.32 above outlines the sample sizes of each implementation category for the disruptive behaviour outcome variable. Typically, the low-risk groups had the largest sample size for both implementation variables, and the high-risk groups had
the smallest. Due to missing data in the second year of the trial (see section 6.5.5), the sample sizes were generally smaller for all risk groups in year 4.

Table 7.33: Mean T2 behaviour scores (SE) by risk group for implementation groups

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>FidQual Year 3</th>
<th>Dosage Year 3</th>
<th>FidQual Year 4</th>
<th>Dosage Year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-Risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1.62 (0.04)</td>
<td>1.49 (0.05)</td>
<td>1.54 (0.03)</td>
<td>1.82 (0.10)</td>
</tr>
<tr>
<td>High</td>
<td>1.59 (0.04)</td>
<td>1.43 (0.06)</td>
<td>1.59 (0.04)</td>
<td>1.42 (0.06)</td>
</tr>
<tr>
<td>Medium-Risk</td>
<td>1.84 (0.05)</td>
<td>1.98 (0.08)</td>
<td>1.89 (0.05)</td>
<td>1.91 (0.05)</td>
</tr>
<tr>
<td>Low</td>
<td>1.81 (0.08)</td>
<td>1.91 (0.05)</td>
<td>1.78 (0.09)</td>
<td>1.91 (0.05)</td>
</tr>
<tr>
<td>High</td>
<td>1.78 (0.09)</td>
<td>1.78 (0.09)</td>
<td>1.78 (0.09)</td>
<td>1.78 (0.09)</td>
</tr>
<tr>
<td>High-Risk</td>
<td>1.89 (0.11)</td>
<td>2.69 (0.36)</td>
<td>2.07 (0.14)</td>
<td>2.18 (0.16)</td>
</tr>
<tr>
<td>Low</td>
<td>1.86 (0.19)</td>
<td>2.18 (0.16)</td>
<td>2.09 (0.40)</td>
<td>2.18 (0.16)</td>
</tr>
<tr>
<td>High</td>
<td>2.18 (0.16)</td>
<td>2.09 (0.40)</td>
<td>2.09 (0.40)</td>
<td>2.09 (0.40)</td>
</tr>
<tr>
<td>Overall</td>
<td>1.73 (0.03)</td>
<td>1.76 (0.05)</td>
<td>1.72 (0.03)</td>
<td>1.81 (0.06)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.33 above outlines mean T2 TOCA-C disruptive behaviour scores for each of the risk groups, for both FidQual and dosage categories. Regarding FidQual, mean disruptive behaviour scores were typically higher in the low FidQual classrooms in both years for pupils in all risk groups, with the exception of the high-risk group in year 3, where mean disruptive behaviour scores were higher in the high FidQual group (2.69) relative to the low FidQual group (1.89). A similar pattern was evident for dosage, with mean disruptive behaviour scores typically higher in the low dosage classrooms in both years for pupils in all risk groups, with the exception of the low-risk group in year 3, where mean disruptive behaviour scores were higher in the high dosage group (1.82) relative to the low dosage group (1.54).
Table 7.34: Sample sizes for implementation groups, by risk group for reading

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Year 3</th>
<th></th>
<th>Year 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FidQual</td>
<td>Dosage</td>
<td>FidQual</td>
<td>Dosage</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>No Risk</td>
<td>141</td>
<td>49</td>
<td>125</td>
<td>34</td>
</tr>
<tr>
<td>Low-Risk</td>
<td>526</td>
<td>184</td>
<td>444</td>
<td>190</td>
</tr>
<tr>
<td>Medium-Risk</td>
<td>199</td>
<td>50</td>
<td>167</td>
<td>71</td>
</tr>
<tr>
<td>High-Risk</td>
<td>95</td>
<td>22</td>
<td>76</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>961</td>
<td>305</td>
<td>812</td>
<td>330</td>
</tr>
</tbody>
</table>

Table 7.34 above outlines the sample sizes of each implementation category for the reading attainment outcome variable. Typically, the low-risk groups had the largest sample size for both implementation variables, and the high-risk groups had the smallest. Due to missing data in the second year of the trial (see section 6.5.5), the sample sizes were generally smaller for all risk groups in year 4.

Table 7.35: Mean T2 reading point scores (SE) by risk group for implementation groups

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Year 3</th>
<th></th>
<th>Year 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FidQual</td>
<td>Dosage</td>
<td>FidQual</td>
<td>Dosage</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>No Risk</td>
<td>35.79</td>
<td>(0.81)</td>
<td>36.70</td>
<td>(0.85)</td>
</tr>
<tr>
<td>Low-Risk</td>
<td>33.63</td>
<td>(0.45)</td>
<td>33.74</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Medium-Risk</td>
<td>27.74</td>
<td>(0.80)</td>
<td>28.32</td>
<td>(0.90)</td>
</tr>
<tr>
<td>High-Risk</td>
<td>24.63</td>
<td>(1.22)</td>
<td>24.68</td>
<td>(1.20)</td>
</tr>
<tr>
<td>Overall</td>
<td>31.66</td>
<td>(0.35)</td>
<td>34.97</td>
<td>(0.59)</td>
</tr>
</tbody>
</table>
Table 7.35 outlines mean T2 HGRT reading scores for each of the risk groups, for both FidQual and dosage categories. Regarding FidQual, reading scores were typically higher in the high FidQual classrooms in both years for pupils in all risk groups, with the exception of pupils with no or low-risk in year 4, who had lower mean reading scores in the high FidQual group (no risk = 28.18; low-risk = 33.83) relative to the low FidQual group (no risk = 36.12; low-risk = 34.33). The mean reading scores for dosage were somewhat more inconsistent; while pupils in the low (35.31) and medium (29.71) risk groups in year 3 had higher reading scores in high dosage classrooms relative to low dosage classrooms (low-risk = 33.74; medium-risk = 28.32), pupils in the no risk (35.97) and high-risk (22.73) groups had lower reading point scores in high dosage classrooms relative to low dosage classrooms (no risk = 36.70; high-risk = 24.68). In year 4, while reading scores were lower in high dosage classrooms for pupils in the medium (27.16) and high (22.04) risk groups relative to low dosage classrooms (medium-risk = 28.92; high-risk = 25.38), pupils in the no risk (39.11) and low risk (34.52) groups had higher reading scores in the high dosage classrooms relative to the low dosage classrooms (no risk = 34.94; low-risk = 33.91).

7.7.2 Research Question 3

3. Are there any differential associations between implementation of the GBG (specifically fidelity, quality, dosage) and behavioural and/or academic outcomes for at-risk children?

The dataset utilised for this research question was reorganised prior to analysis to create two datasets, one for each year of the trial (year 3 and year 4). These consisted of pupils allocated to the GBG condition. Pupils were grouped by class teacher and teachers’ FidQual and dosage scores were added to the dataset. In order to establish if there was evidence of differential associations between implementation of the GBG and pupils at different levels of risk exposure, MLMs were utilised. Two sets of analyses were conducted, one with disruptive behaviour scores as the outcome variable, and one with reading point scores. Standardised
versions of the outcome variables were utilised, to produce effect sizes and allow for meaningful comparisons between models. Similarly to prior analysis for RQ1 and RQ2, empty models were calculated first; in the empty models, class- and pupil-level data were added as predictor variables, with either disruptive behaviour scores or reading point scores added as outcome variables. Empty models, including the ICC for each level, are presented below in table 7.36 for disruptive behaviour, and table 7.37 for reading attainment.

Table 7.36: Empty MLM for disruptive behaviour outcome variable

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β0ij = 0.066(0.049)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2*log likelihood = 3390.141</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class-Level</td>
<td>0.087</td>
<td>0.025</td>
<td>&lt;.001**</td>
<td>8.5%</td>
</tr>
<tr>
<td>Pupil-Level</td>
<td>0.935</td>
<td>0.039</td>
<td>&lt;.001**</td>
<td>91.5%</td>
</tr>
<tr>
<td><strong>Year 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β0ij = 0.053(0.054)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2*log likelihood = 3400.447</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class-Level</td>
<td>0.088</td>
<td>0.028</td>
<td>.001**</td>
<td>8.5%</td>
</tr>
<tr>
<td>Pupil-Level</td>
<td>0.950</td>
<td>0.040</td>
<td>&lt;.001**</td>
<td>91.5%</td>
</tr>
</tbody>
</table>

Table 7.36 shows that there was a significant association between disruptive behaviour at T2 and the year 3 class-level ($\mu_{0j} = 0.087, p < .001$) and pupil-level ($\mu_{0j} = 0.935, p < .001$). The year 3 ICCs demonstrate that the class-level accounted for 8.5% of the total variance in disruptive behaviour scores, whilst pupil-level variance accounted for the remainder (91.5%). There was also a significant association between disruptive behaviour at T2 and the year 4 class-level ($\mu_{0j} = 0.088, p = .001$) and pupil-level ($\mu_{0j} = 0.935, p < .001$). The ICCs were the same as the year 4 model.
Table 7.37: Empty MLM for reading attainment outcome variable

<table>
<thead>
<tr>
<th>Empty Model: Reading Attainment</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p value</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{0ij} = -0.038(0.055)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2*log likelihood = 3479.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Class-Level</strong></td>
<td>0.131</td>
<td>0.032</td>
<td>&lt;.001**</td>
<td>13.2%</td>
</tr>
<tr>
<td><strong>Pupil-Level</strong></td>
<td>0.859</td>
<td>0.035</td>
<td>&lt;.001**</td>
<td>86.8%</td>
</tr>
<tr>
<td><strong>Year 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{0ij} = -0.111(0.064)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2*log likelihood = 3479.793</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Class-Level</strong></td>
<td>0.150</td>
<td>0.040</td>
<td>&lt;.001**</td>
<td>14.7%</td>
</tr>
<tr>
<td><strong>Pupil-Level</strong></td>
<td>0.867</td>
<td>0.035</td>
<td>&lt;.001**</td>
<td>85.3%</td>
</tr>
</tbody>
</table>

Table 7.37 shows that there was a significant association between reading attainment at T2 and the year 3 class-level ($\mu_{0j} = 0.131$, $p < .001$) and pupil-level ($\mu_{0j} = 0.859$, $p < .001$). The ICCs demonstrate that class-level variance accounted for 13.2% of the total variance in reading scores, whilst pupil-level variance accounted for the remainder (86.8%). There was also a significant association between reading attainment at T2 and the year 4 class-level ($\mu_{0j} = 0.150$, $p < .001$) and pupil-level ($\mu_{0j} = 0.867$, $p < .001$). The ICCs demonstrate that class-level variance accounted for 14.7% of the total variance in reading attainment, whilst pupil-level variance accounted for the remaining 85.3%.

The next stage involved running two sets of analyses, one with disruptive behaviour scores as the outcome variable, and one with reading point scores. For each outcome variable, two models were created: one model to establish associations between implementation and outcomes for pupils at different levels of risk.
exposure in year 3, and one for year 4. Similarly to running the empty models, standardised disruptive behaviour and reading point scores respectively were first added to each of the full models, before adding the relevant baseline disruptive behaviour and reading point scores, to control for pupils’ prior behaviour and attainment. Both the categorical FidQual and dosage implementation variables were then also added, along with risk groups. Finally, interaction terms between the implementation variables of interest and risk groups were added to establish if there were any differential gains.

Similarly to RQ2, regarding behaviour problems, interaction terms that were significantly negatively associated with the disruptive behaviour outcome variable would suggest that high levels of implementation were predictive of lower levels of disruptive behaviour, as lower scores on the TOCA-C indicate fewer difficulties. For reading attainment, interaction terms that were significantly positively associated with the reading point score outcome variable would suggest that high levels of implementation were predictive of higher reading attainment, as higher point scores are indicative of better reading outcomes.

Due to the amount of missing data, MI procedures were conducted (see section 7.2). Pupil gender and the constant were added as auxiliary variables, whilst all pupil-level variables utilised in the multiple imputation for RQ2 were again added as outcome variables. Teacher-level variables including teacher gender, years’ experience, self-reported stress, self-efficacy and retention were also added as response variables. The same settings utilised in REALCOM-impute for RQ2 were utilised for this analysis.

The multiply imputed full models are shown in 7.38 and 7.39 for disruptive behaviour and 7.40 and 7.41 for reading attainment. Complete case analyses were also performed and are displayed in the appendices (see appendix 12), as there were some discrepancies between the models. However, the multiply imputed models are considered to be the main form of analysis, as bias is reduced (see section 7.2).
Table 7.38: Imputed year 3 MLM for implementation and behaviour

**Full Model: Disruptive Behaviour Year 3**

$\beta_{0ij} = -1.470(0.091)$

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>SE</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class-Level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.156</td>
<td>0.035</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>Yr3 FidQual</strong> (if high)</td>
<td>-0.016</td>
<td>0.154</td>
<td>.459</td>
</tr>
<tr>
<td><strong>Yr3 Dosage</strong> (if high)</td>
<td>0.172</td>
<td>0.144</td>
<td>.119</td>
</tr>
<tr>
<td><strong>Pupil-Level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.549</td>
<td>0.025</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>Behaviour_T1</strong></td>
<td>0.820</td>
<td>0.030</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>Risk Group:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-Risk</td>
<td>◇</td>
<td>◇</td>
<td>◇</td>
</tr>
<tr>
<td>Med-Risk</td>
<td>0.193</td>
<td>0.069</td>
<td>.004*</td>
</tr>
<tr>
<td>High-Risk</td>
<td>0.249</td>
<td>0.153</td>
<td>.055</td>
</tr>
<tr>
<td><strong>Yr3 High FidQual*Med-Risk</strong></td>
<td>0.067</td>
<td>0.123</td>
<td>.294</td>
</tr>
<tr>
<td><em><em>Yr3 High FidQual</em> High-Risk</em>*</td>
<td>0.539</td>
<td>0.320</td>
<td>.049*</td>
</tr>
<tr>
<td><strong>Yr3 High Dosage*Med-Risk</strong></td>
<td>-0.012</td>
<td>0.124</td>
<td>.462</td>
</tr>
<tr>
<td><strong>Yr3 High Dosage*High-Risk</strong></td>
<td>-0.374</td>
<td>0.262</td>
<td>.079</td>
</tr>
</tbody>
</table>
Table 7.39: Imputed year 4 MLM for implementation and behaviour

<table>
<thead>
<tr>
<th>Full Model: Disruptive Behaviour Year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{0ij} = -1.348(0.110)$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Coefficient</strong></td>
</tr>
<tr>
<td>Class-Level</td>
</tr>
<tr>
<td>(if high)</td>
</tr>
<tr>
<td>Yr4 FidQual</td>
</tr>
<tr>
<td>Yr4 Dosage</td>
</tr>
<tr>
<td>Pupil-Level</td>
</tr>
<tr>
<td>(if high)</td>
</tr>
<tr>
<td>Behaviour_T1</td>
</tr>
<tr>
<td>Risk Group:</td>
</tr>
<tr>
<td>Low-Risk</td>
</tr>
<tr>
<td>Med-Risk</td>
</tr>
<tr>
<td>High-Risk</td>
</tr>
<tr>
<td>Yr4 High FidQual*Med-Risk</td>
</tr>
<tr>
<td>Yr4 High FidQual*High-Risk</td>
</tr>
<tr>
<td>Yr4 High Dosage*Med-Risk</td>
</tr>
<tr>
<td>Yr4 High Dosage*High-Risk</td>
</tr>
</tbody>
</table>

As can be seen in tables 7.38 and 7.39 above, the multi-level analysis revealed no statistically significant associations between GBG implementation and pupils’ behaviour; neither FidQual nor dosage were found to be significant predictors of pupils’ TOCA-C disruptive behaviour scores at T2 in either year 3 (FidQual: $\beta_{0j} = -0.016$, $p = .459$; dosage: $\beta_{0j} = 0.172$, $p = .119$) or year 4 (FidQual: $\beta_{0j} = -0.170$, $p = .200$; dosage: $\beta_{0j} = -0.061$, $p = .368$).

There were also no statistically significant associations in the subgroup analysis at the $p<.05$ level between implementation and outcomes for pupils in any of the risk groups in either year 3 or year 4, with the exception of high FidQual in year 3 for pupils in the high-risk group. However, high levels of FidQual were associated with higher disruptive behaviour scores amongst high-risk pupils ($\beta_{0j} = 0.539$, $p < .05$). Regarding dosage, there was a marginal trend in the year 3 model indicative of
Table 7.40: Imputed year 3 MLM for implementation and reading

<table>
<thead>
<tr>
<th>Risk Group:</th>
<th>Coefficient</th>
<th>SE</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class-Level</td>
<td>0.063</td>
<td>0.015</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Yr3 FidQual (if high)</td>
<td>-0.073</td>
<td>0.127</td>
<td>.284</td>
</tr>
<tr>
<td>Yr3 Dosage (if high)</td>
<td>-0.162</td>
<td>0.170</td>
<td>.172</td>
</tr>
<tr>
<td>Pupil-Level</td>
<td>0.375</td>
<td>0.018</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Reading Score_T1</td>
<td>0.203</td>
<td>0.006</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Risk Group:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Med risk</td>
<td>-0.075</td>
<td>0.120</td>
<td>.267</td>
</tr>
<tr>
<td>High-risk</td>
<td>0.073</td>
<td>0.186</td>
<td>.348</td>
</tr>
<tr>
<td>Yr3 High FidQual*Low-Risk</td>
<td>0.064</td>
<td>0.118</td>
<td>.295</td>
</tr>
<tr>
<td>Yr3 High FidQual*Med-Risk</td>
<td>0.101</td>
<td>0.136</td>
<td>.230</td>
</tr>
<tr>
<td>Yr3 High FidQual*High-Risk</td>
<td>-0.170</td>
<td>0.195</td>
<td>.193</td>
</tr>
<tr>
<td>Yr3 High Dosage*Low-Risk</td>
<td>0.026</td>
<td>0.150</td>
<td>.432</td>
</tr>
<tr>
<td>Yr3 High Dosage*Med-Risk</td>
<td>-0.079</td>
<td>0.192</td>
<td>.341</td>
</tr>
<tr>
<td>Yr3 High Dosage*High-Risk</td>
<td>-0.102</td>
<td>0.289</td>
<td>.364</td>
</tr>
</tbody>
</table>
Table 7.40: Imputed year 4 MLM for implementation and reading

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>SE</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class-Level</td>
<td>0.064</td>
<td>0.018</td>
</tr>
<tr>
<td>Yr4 FidQual (if high)</td>
<td>0.302</td>
<td>0.185</td>
</tr>
<tr>
<td>Yr4 Dosage (if high)</td>
<td>0.159</td>
<td>0.176</td>
</tr>
<tr>
<td>Pupil-Level</td>
<td>0.369</td>
<td>0.016</td>
</tr>
<tr>
<td>Reading Score_T1</td>
<td>0.208</td>
<td>0.006</td>
</tr>
<tr>
<td>Risk Group:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Risk</td>
<td>◇</td>
<td>◇</td>
</tr>
<tr>
<td>Low-Risk</td>
<td>0.174</td>
<td>0.106</td>
</tr>
<tr>
<td>Med Risk</td>
<td>0.082</td>
<td>0.111</td>
</tr>
<tr>
<td>High-Risk</td>
<td>0.020</td>
<td>0.145</td>
</tr>
</tbody>
</table>

As can be seen in table 7.40 and 7.41 above, the multi-level analysis revealed a marginal trend whereby high FidQual was associated with higher reading scores at T2 in year 4 only ($\beta_{0j} = 0.302$, $p = .054$), with a small effect size of 0.302. This finding was supported by the descriptive statistics outlined in table 7.35 previously, which indicated higher overall mean reading scores in the high FidQual classrooms relative to low FidQual classrooms in year 4 (32.96 vs. 31.76). Dosage was not a significant predictor of reading scores in either model (year 3: $\beta_{0j} = -0.162$, $p = .172$; year 4: $\beta_{0j} = 0.159$, $p = .186$).
There were no statistically significant associations between implementation and risk group in either the year 3 or year 4 models.

**7.7.3 Summary of RQ3**

- Classes were categorised into one of two groups for FidQual (high FidQual and low FidQual) and one of two groups for dosage (high dosage and low dosage).
- Neither FidQual nor dosage were found to be significant predictors of pupils’ overall disruptive behaviour scores.
- There was a marginal trend indicative of higher reading point scores in high FidQual classrooms in the year 4 model. There were no statistically significant effects of dosage on reading point scores.
- There were no statistically significant beneficial subgroup associations between FidQual or dosage and disruptive behaviour scores in any of the risk groups.
  - There were *higher* disruptive behaviour scores among high-risk pupils in high FidQual classrooms in year 3.
  - There was a marginal trend in the year 3 model indicative of an association between lower disruptive behaviour scores and high dosage for high-risk pupils.
- There were no statistically significant subgroup associations between FidQual or dosage and reading scores in any of the risk groups.
**7.8 Sensitivity Analyses**

Whilst MI was utilised for the analyses in RQ2 and RQ3 to account for missing data (see section 7.2), in order to assess the sensitivity of these analyses to changes in modelling parameters and assumptions, complete case analyses were also conducted (see appendices 11 and 12). No substantive differences between the two analyses were evident for RQ2; however, for RQ3, some discrepancies did exist. Specifically, in the complete case analyses, there were statistically significant associations between:

- High year 3 dosage and lower disruptive behaviour scores for high-risk pupils
- High year 4 FidQual and lower disruptive behaviour scores for high-risk pupils
- High year 3 FidQual and higher overall reading scores
- High year 4 FidQual and higher overall reading scores
- High year 3 dosage and lower reading scores for high-risk pupils
- High year 4 FidQual and lower reading scores for low-risk pupils
- High year 4 dosage and lower reading point scores for low-risk pupils

However, once MI procedures were conducted, these associations were no longer statistically significant. It appears that while these sensitivities were restricted to interactions between implementation and high-risk status for disruptive behaviour outcomes, reading attainment outcomes were sensitive both to general interactions between implementation and reading scores, and specific subgroup interactions between implementation and risk status.

These analyses suggest that the findings from RQ2 were very robust indeed; however, for RQ3, they were less so. It is important to note that RQ3 was highly exploratory in nature, and it is likely that the differences between the complete case and MI analyses were due to the extremely small sample sizes of the risk
groups in the complete case models. This means that these complete case analyses were heavily underpowered, and so it is probable that the results were skewed or spurious. Therefore, the multiply imputed models were considered superior (Morris, Kahan, & White, 2014), and thus, while these discrepancies are acknowledged, the results from the complete case analyses are not discussed in any further detail in the present study.
8. Qualitative Findings – RQ2a

8.1 Chapter Overview

The following sections explore general perceptions of impact of the GBG, before moving on to perceptions of differential impact for pupils exposed to various risk variables. Hybrid thematic analysis was conducted, allowing for both inductive and deductive themes. Drawing on the quantitative findings in RQ1, deductive themes were organised by variables identified as risks for either disruptive behaviour or reading attainment; therefore, there were nine a priori organising themes included in the analysis. However, only some of these risk variables emerged in the qualitative dataset, namely gender (male), SEND status, EAL status (although not ethnicity), school-level behaviour, and both familial (FSM eligibility) and neighbourhood (IDACI score) deprivation. Somewhat broader themes of pupils with behavioural problems and pupils with low academic attainment also emerged inductively (see sections 8.3.1 and 8.3.2). Thus, eight organising themes were identified (see figure 8.1).

Figure 8.1: Thematic map of global and organising themes

Additionally, the dataset was analysed inductively to allow for unanticipated themes regarding the ways in which the GBG influenced outcomes for at-risk pupils. Several themes emerged through this analysis, many of which were discussed in relation to more than one risk variable. Thus, whilst there may be some repetition in the sections below, it is important that the mechanisms perceived to be producing change were allowed to develop organically for each risk variable, to account for nuances that may have emerged. Had themes been organised by
commonly identified mechanisms, some information regarding the specific mechanisms producing change for individual risk variables may have been lost.

8.2 Perceptions of Impact

Teachers felt that the GBG was benefiting the pupils in their class, although the ways in which they reported this varied. For instance, some teachers simply commented that the GBG “made a positive impact in the classroom” (school ci603, teacher vd8979) and that pupils were “getting much better, much better definitely” (school sa774, teacher pf2223), but did not elaborate on the outcomes or mechanisms through which these benefits occurred. Other teachers were more specific, noting that pupils’ behaviour and attainment had benefited from the GBG; for example, one teacher reported that they used the GBG to “teach them the skills not to disrupt their own or others’ learning” (school gy985, GBG lead). However, more often teachers’ perceptions of impact were attributed to other effects that, while not directly related to behaviour or attainment, could be seen to be mediating these outcomes. Therefore, while distal effects were not always discussed, other more proximal effects were explored by teachers, and therefore are also included in the further analyses reported below.

Several teachers commented on the beneficial effects of the GBG on pupils’ concentration and independence. Whilst these comments were not directly linked to attainment, if pupils are more focused on their work and less reliant on others, this may mean that their attainment improves (Chan et al., 2012); an increase in on-task behaviours may also mean that disruptive behaviours are reduced (Leflot et al., 2010). One teacher reported that the pupils found completing their work easier during the GBG as it encouraged them to concentrate: “there’s no noise at all because they’re concentrating... they’ve actually admitted that they find it easier when the Good Behaviour Game’s being played because they can concentrate better” (school hd887, teacher fw5212), whilst another commented that the biggest benefit they had seen was an increase in independence: “it’s not really been a behaviour management tool for me, it’s been an independence management tool”
Other teachers reported an increase in confidence, which in turn influenced pupils’ learning:

“I would say Year 4 has become much more independent because of it... they’re willing to risk now and that's... a start isn't it? If you can get a child to risk making mistakes then you’ve got a confident learner, you learn more from when they make mistakes than a page of correct ticks” (school gy9985, GBG lead)

Similarly for behaviour, teachers commented that pupils were “definitely staying in their seats more and putting their hands up more” (school gy985, teacher fb2388). Several teachers attributed this to the increase in accountability that the game placed on pupils, stating that it encouraged them to “take responsibility” (school sa774, teacher yp4362) for their own behaviour, particularly due to the team work element of the game:

“Normally if you [were] doing a group activity you’d get... the children constantly getting out of their seats... not doing what they should be doing, but if you’re saying ‘right we’re doing the Good Behaviour Game’ and they’re doing the team element of it, I feel like they stay on task, not getting out of their places” (school ci603, teacher gb6981)

This benefit seemed to be sustained over time, with this teacher still reporting this effect after a year of implementation: “they don’t get out of their places, they don’t come and ask me questions, and they just work through it as a... team together”. Several teachers reported similar benefits, noting that the “team work element’s been massive for them” (school ci603, teacher vd8979). Indeed, other teachers reported that it “has improved relationships” (school gy985, teacher fb2388), and that social skills were developed: “it’s also kind of allowed them to build their social skills because they’re on tables with other students who they won’t normally sit with, so they’ve... developed that kind of social element as well” (school tm517, teacher nk6311). Therefore it was not just disruptive behaviours that were discussed with
regards to perceptions of impact; it may be that other types of behaviour, such as pro-social and adaptive learning behaviours, were also thought to have improved as a result of the GBG.

8.3 Perceptions of Differential Impact

Whilst teachers generally reported that the GBG had a positive impact on pupils, some teachers also commented that they felt the GBG benefitted certain groups of pupils in particular; for example, one teacher commented that the game was benefiting “some more than others definitely” (school ci603, teacher iq8743). Teachers went on to identify pupils deemed to be at risk, such as pupils with an SEND and pupils speaking EAL, as benefiting from the game more than other pupils. Other teachers commented on the differential effects for ‘lower ability’ pupils, or pupils with ‘behaviour plans’, thus referring to the overarching problems more generally, as opposed to the underlying risk variables influencing these issues. Detailed accounts of these perceived differential effects, both for pupils with specific risk variables and those with overarching behavioural and academic difficulties, can be found in the following sections below.

However, not all teachers agreed that there had been differential effects of the GBG; some instead felt that “it’s been really beneficial to every single child in my class” (school ci604, teacher vd8979). Other teachers reported that whilst differential effects had been evident at first, that this had changed over time, with one teacher reflecting: “probably now at this stage in the year probably equally. I wouldn’t say anyone stands out in particular” (school sa774, teacher pf2223). One teacher who had initially commented on differential benefits for EAL pupils reported that whilst individual pupils may have benefited more than others, there were no differential effects for groups of pupils by the end of the trial, suggesting a change over time: “there are certain pupils that respond better than others but... I can’t say if it’s a certain group of children really, no” (school gy985, teacher fb2388).
8.3.1 Pupils with Low Academic Attainment

Whilst some teachers remarked on the differential effects of the GBG for pupils exposed to specific underlying risk variables for attainment, others commented more generally on pupils with low academic attainment, often referring to these pupils as ‘lower ability children’. Opinions were mixed regarding the effects of the GBG on this group of pupils; some teachers felt that there were no differential effects between pupils at different levels of ability: “even in terms of sort of attainment with highers, middles, lowers, they all have responded in a similar way” (school sa774, teacher yp4362), whilst others felt that differences were evident. However, whilst other teachers reflected that the GBG benefited these pupils, some felt that it hindered their progress. Conversely, some teachers noted differential effects for higher ability pupils.

Several teachers remarked that the GBG had a positive effect on lower ability pupils, although they did not always elaborate on the mechanisms through which the game helped these pupils. For instance, one teacher commented that the GBG helped these pupils’ attainment: “it does help them with their work, so they are doing well” (school sa774, teacher yp4362); whilst another felt that the GBG improved both attainment and behaviour:

“You’re faced with a lower ability class at the... beginning of the year and... you think like ‘how am I going to get them from there to there?’ but they've just made so much progress, even academic wise as well, and then like behaviour and as a class they’re just amazing, like any other teacher comes into the class and... they all just say that they’re fantastic, so yeah” (school ci603, teacher vd8979)

One teacher appeared to feel that fidelity to the manual regarding teacher-pupil interaction encouraged them to be more explicit with their directions during the pre-game stage, which in turn benefited the lower ability pupils: “…but I never move on in The Good Behaviour Game until a couple of them that I know might
struggle can tell me what they’re supposed to do, so... yeah” (school gy985, teacher fb2388); they suggested that this helped these children with their work as they better understood what was expected of them: “definitely helps sort of the lower... children who’ve struggled to understand so it’s especially good for them I think”. Other teachers felt that the team leader aspect of the GBG also helped these lower ability pupils as it allowed them to succeed in other areas, increasing their self-esteem and, consequently, their attainment:

“I can think of quite a few who... have the role to be a team leader and actually thriving on that... it gives them ability to show different aspects of themselves that are in there with them but they don’t always get the opportunity to let that come out. So while... they might not be in the top group for Maths or something they can be in the top team that’s winning the Good Behaviour Game, and that in themselves gives them self-esteem which then enhances their learning overall” (school hd887, GBG lead)

Teachers also felt that the responsibility that came with the team leader role increased these pupils’ “leadership skills” (school hd887, teacher td4967) and confidence: “I chose children who don’t excel academically... to give them more responsibility and a bit more a bit of a confidence boost” (school gy985, teacher fb2388). Another teacher commented that they thought the team element generally helped lower ability pupils with their work as it encouraged them to become independent learners: “I love the team work aspect... because... I’ve got the lower ability children and... they find it really hard to be an independent learner, never mind having to kind of learn within a group or help a group as well at the same time” (school ci603, vd8979).

However, not all teachers agreed that the GBG benefited lower ability pupils; contrary to other teachers’ remarks, one teacher felt that the team leader element of the GBG hindered attainment as it caused these pupils to “coast” and pass over responsibility for their work to the team leader:
“I’ve got the lower ability so... my kind of lower in my class as well would be particularly poor. I think sometimes they kind of coast a little bit whenever... we play The Good Behaviour Game... it’s based on a leader kind of taking the lead and helping one another and everything but I think sometimes... they switch off a little bit and kind of give up the responsibility hoping that the other team members will pick it up” (school ci603, teacher iq8743)

Other teachers commented that although the lower ability pupils “still enjoy the game”, they would “struggle a little but more with the team work” (school ci603, GBG lead). There were also issues raised by teachers regarding the lack of support and guidance that could be provided during the game; specifically, they felt that these pupils struggled to work independently, particularly in certain lessons:

“Very low children are just used to an adult being with them all the time and encouraging them or... boosting their confidence or even just kind of modelling through what they need to do, and I find it really difficult to do an independent Numeracy task because if not they just stare into space and... never following direction because they’re just not used to that independent learning” (school ci603, vd8979)

To overcome this, some teachers made adaptations to the game, and thus implemented it with lower fidelity in order to help these pupils: “initially I played the game rigidly but it became demotivating for the really low achieving students who, as much as I wanted to encourage to work independently, and that’s exactly what I wanted to do, some just couldn’t access it” (school tm517, teacher nk6311). Another teacher felt that lower ability pupils struggled to understand how to rectify their behaviour if teachers adhered to the strict “check-comment-redirect” (CCR) script outlined in the manual following an infraction, and so they adapted it to help these pupils:

“Because I’m in a lower ability class, sometimes if I just say ‘team number three you’ve broken rule number four, ‘we will follow directions”, and leave
it there, sometimes they don’t know what they’re doing wrong… so me and my mentor have decided that at times like that it’s better for me to… give a very small explanation as to what they might be able to do or what they are doing wrong just to help… the team… rectify the situation” (school ci603, teacher vd8979)

The type of task provided during the GBG was a common issue raised by teachers when discussing lower ability pupils, as some of these pupils found certain tasks more difficult. Most teachers attributed this problem to the lack of teacher-pupil interaction specified in the manual, for example, one teacher commented: “I can differentiate my work perfectly but they still might not understand it and it might still need that teacher input“ (school ci603, teacher iq8743). They went on to explain that they felt that the GBG was “a little bit too prescriptive at times… a little bit too black and white” for these pupils, and had thus adapted the game to address this, despite their coach’s advice: “I’ve been pulled up on that a couple of times with my feedback and… that’s just kind of what I’ve done… I’m not going to change that… I think… there does need to be a little room for… a grey area”.

One teacher felt that large tasks in particular left these pupils disengaged with their work, thus hindering attainment: “I think with the lower ability group as well they can be quite disengaged so if it is a big task like assessed writing, I find that they… do struggle” (school tr248, teacher hj8221), whilst another felt that pupils produced better work outside of the GBG if they were faced with a difficult or independent task:

“If it’s say literacy or numeracy and I want… a really good piece of work there's a lot of children in my class that can’t sit down and do that on their own without a teacher. No matter how much input I give to them before The Good Behaviour Game they just need that kind of constant ‘come on you need to carry on’ and they're... just that type of child, so if it’s an independent task and it’s writing... or a difficult numeracy task then I’d say

213
they do that better outside of The Good Behaviour Game” (school ci603, vd8979)

Conversely, some teachers focused on the higher ability pupils when discussing differential gains with the GBG, although, again, there were conflicting opinions over whether the GBG helped or hindered these pupils. They felt that the higher ability pupils benefited from the game, commenting, “being the higher ability group I think it just fits so well” due to these pupils’ “willingness to learn” (school ci603, teacher ka0624). One teacher felt that the GBG helped these pupils academically as it provided a better learning environment: “I think maybe some of the more independent children anyway, or the more academic, will have worked in a better working environment so I think maybe that be one of the pluses” (school gy985, GBG lead). However, another teacher felt that fidelity to the manual during the pre-game stage took “extra time” which hindered the higher ability children, as it prevented them from getting on with their work, leaving them frustrated:

“I guess some of the higher ability ones might be a bit sort of wanting to get on... so it maybe holds them back a little bit from getting started. ...Like I say it’s just the higher ability ones that I think might get a little frustrated at the beginning when they’re waiting to get started with their work” (school gy985, teacher fb2388).

8.3.2 Pupils with Behavioural Difficulties

Whilst some teachers remarked on the differential effects of the GBG for pupils exposed to specific underlying risk variables for behaviour, other commented more generally on pupils with behavioural difficulties. Some teachers remarked on the beneficial effects of the GBG for classes where behaviour was a particular problem; one teacher reflected that it “had such... a positive effect on the classroom as well ‘cos this class are a very difficult class and... it’s definitely working” (school tr248, teacher hj8221). Others commented on how the social element of the game had improved behaviour in these classes, finding that pupils were more conscious of
how their behaviour impacted others: “I think the fact that it’s made them realise that their behaviour impacts the other people on their team is kind of the next level up which is good” (school gy985, teacher fb2388). Teachers also noted how pupils with previous behavioural issues “really enjoy” (school tr248, teacher hj8221) playing the game, with one teacher commenting: “even the ones that have got, like, who struggle behaviourally, they’re quite enthusiastic about it” (school gy985, teacher fb2388).

Similarly, pupils remarked that behaviour had improved in the class, with one pupil commenting: “everybody’s behaviour’s improved a lot. Before we had The Good Behaviour Game other people started messing about and start acting like crazy maniacs but since... The Good Behaviour Game’s here... we’ve become like a bit more sensible I would say”. Another pupil felt that their own behaviour had improved as a result of playing the GBG: “I was always being silly, I was always singing in class, I was always humming, I was always being silly.... When we started doing The Behaviour Game I became more like, not getting out of my seat” (school gy985).

Teachers also commented on the benefits of the GBG to specific pupils with behavioural difficulties, or those with ‘behaviour plans’. Generally, teachers were positive regarding the outcomes of the game for these pupils, with one teacher remarking that it had “quite a profound effect”. They went on to explain:

“There’s been difficulties over the years for him and... he’s obviously on... bad behaviour plans for his behaviour and... at times be quite volatile if something’s not going his way... he’s often ran out of class and refused to do work... but that’s not happened for quite a few weeks now so obviously, you know it’s working” (school sa774, GBG lead)

Other teachers noted similar gains, with one teacher commenting: “the children on behaviour plans... one of them took to the game really well, he just got straight on with it and he was, he wasn’t the team leader for the first term but... he tried to take
on that role because he was really wanted to do well in the game” (school sa774, teacher yp4362). They explained that they thought the social element of the game was particularly beneficial to this pupil; although initially the teacher had adapted the game to provide the child with one-to-one support, “that’s gradually come away” as “he gets support from his group instead”.

One element of the game that several teachers found particularly beneficial to the pupils with behavioural difficulties was being able to allocate pupils the role of team leader. Teachers explained that they purposefully gave these pupils the role: “I chose some children... who do have behaviour problems to try and encourage them to lead their team” (school gy985, teacher fb2388) and to “boost their confidence and to make them think ‘well I’ve got a, a sort of important position here I need to make sure that I follow the rules as well’, and that’s worked so well, I didn’t expect it to but it has” (school hd887, teacher yt9266). One teacher explained how the responsibility that the role of team leader placed on pupils helped them to manage their behaviour:

“it’s affecting them all but... there’s a couple of kids with some serious behavioural issues in that class... One of them... since he got used to it he’s really turned it round, he’s been a little superstar recently... He was one that we put into... a small team because we didn’t think he could handle being in a team of six... but he quickly proved himself and he’s already, I’ve put him as a team leader now because he’s responsible. So it’s really having an effect on him... I mean you could see when he came in from the input he threw his book on the floor and... he didn’t really want, he wanted to stay in his Maths group, but as soon as he realised he was going into his Good Behaviour team where he has some responsibility... he sat back down and he got straight on with it” (sa775, teacher yp4362)
8.3.3 Gender (Male)

The majority of comments surrounding male pupils tended to focus more on the effects of the game in relation to other risk variables; for instance, some teachers reported that the game was having a positive effect on the boys with behavioural plans (see section 8.3.2), or boys with additional needs (see section 8.3.4). However, teachers did make comments regarding the GBG specifically in relation to the boys in their class, with some of these teachers reporting positive effects. For instance, one teacher commented that the GBG had improved behaviour: “it’s brilliant, it really is, and it works really well especially with my boys who can be often quite destructive... and hard work” (school hd887, teacher yt9266). Another commented that the majority of the boys in their class had taken to the game well: “I thought the girls had taken it on a bit more than boys but that’s only because the two boys we’ve got on behaviour plans, when I think about the whole class in general the boys... have stuck to it” (school sa774, teacher yp4362).

However, this teacher explained that they had experienced difficulties specifically with two of the boys in their class, and described how “trying to engage those two boys in particular was... really tricky... they all got involved but two of the children, it was quite a challenge to get some of them involved, and it... is still a challenge a little bit”. Other teachers also commented that they had faced issues engaging some of the boys in their class, meaning that behaviour sometimes got worse during the game, with several teachers describing a similar issue whereby the boys would “push it” (school sa774, teacher yp4362). Boys in their class would test the rules, purposefully building up infractions to “see what would happen” (school sa774, teacher yp4362); one teacher described how one pupil “got it right up to four and then just corrected his behaviour... he was just deliberately looking at the number going up and then when it got to four he stopped, it was quite annoying” (school gy985, teacher fb2388).

When the pupils were asked about differential responses to the GBG, some also described how certain boys were not engaging in the game, for example: “because
[pupil] doesn’t like… being quiet… [he] just like shouts, and he just always shouts out. He doesn’t follow directions… he just wants to do his own thing” (school sa774). However, some thought that the GBG had improved the behaviour of some of the boys; for example, one pupil felt that “[pupil] is bit more sensible” (school sa774). Conversely, other pupils thought that the girls had benefited from the game, although these comments focused more on social behaviours rather than disruptive behaviours; one pupil explained: “it helps girls, so like I used to be bossy but now… I just like wait for people to have their turn” (school gy985). One of the teacher’s comments mirrored this, as they had also noticed how the GBG improved girls’ social skills: “some of the girls that would have been a bit argumentative and like their way or no way type thing… they’ve realised that to listen to other opinions, other people’s opinions more” (school ci603, teacher gb6981).

8.3.4 SEND Status

Teachers from the majority of the case study schools discussed the differential effects of the GBG on pupils identified as having an SEND. Views were mixed as to the benefits of the GBG for these groups of pupils, with some teachers reporting positive effects, and others finding that the rigidity of the game could escalate situations for some SEND pupils. This rigidity led to some teachers adapting certain aspects of the game in order to help these pupils in their class. For instance, one teacher commented that they adapted the parameters of the rules when giving infractions to certain pupils due to their additional needs:

“If they were doing something I wouldn’t necessarily count it as an infraction as I might do possibly for someone else. I’ve got one boy in class who can make noises and can be quite loud but that’s just the way he is, that’s his nature, so if it was somebody else that would be a rule break but for him that’s, you know, that’s who he is, so for that, obviously, in some ways you make some allowances for those with additional needs” (school sa774, teacher pf2223).
Another teacher commented that they had adapted the GBG due to similar issues regarding the rigidity of the game and lack of interaction permitted in the manual:

“The only change that I’ve made a couple of times was when I’ve had to intervene… my children are… SEN so if they ever got stuck [at] any point or… didn’t understand what to do then I’d just… intervene in that way. [I: Yeah, so you’d talk to them?] Yeah basically… just to give them a little bit more direction or if they’re… way off with something that they’ve… started work on, I’ll just kind of point them in the right direction” (school ci603, teacher iq8743).

One aspect of the GBG that several teachers commented on was the lack of teacher-pupil interaction permitted during the game. Some teachers found that fidelity to this element was a particular issue with their pupils with an SEND; one teacher explained that not being able to interact with the teacher could be stressful for these pupils, which may in turn influence how well they worked during the game:

“I don’t like having rigidities that you can’t give a little bit of support… and some of our children do need that support when the game’s on… they do need that little bit of interaction from adults because sometimes they’re not able themselves to express how they’re feeling… and sometimes it might just be that they need the loo, they don’t want to ask, or they don’t understand the task and they don’t want to ask ‘cos they don’t want to get an infraction… for some children that five minutes, ten minutes, fifteen minutes of the game can be really quite stressful… they can find it quite stressful that they can’t ask and… they know nobody really is going to come over to them during the game.” (school sa774, teacher fe0140)

Another teacher also reported that the lack of interaction with pupils was an issue; they felt that high fidelity to the manual made managing SEND pupils with behaviour problems more difficult: “because we’re not able to go over… go to that
child and interact with that child, we therefore can’t deescalate the situation... the game just doesn’t allow us to do so” (school tr248, teacher sz6258). Other teachers found the required dosage could be particularly difficult for pupils with an SEND, as they needed support when working for longer periods of time: “we’ve got so much SEN that it’s difficult not to, and if you gave a longer stretch, for example, twenty minutes it would be unfair, well I... think personally it would be unfair to penalise really low achieving pupils for not being able to interact, especially with a teaching assistant” (school tm517, teacher nk6311). This led to some teachers deviating from the manual by allowing pupils with an SEND to have one-to-one support during the game. One teacher explained that for one pupil in their class they had initially incorporated a teaching assistant (TA) into their team; however, over time the pupil’s independence developed to the point that the TA was no longer needed:

“I have one child who has autism... At the beginning, I made [TA] part of the group but he is starting to work more and more independently now on his own so I haven’t had to continue with that. [I: Okay so the TA stays with him during the game but...?] Not anymore.” (school ci603, ka0624)

Other teachers agreed, finding that the GBG helped pupils with an SEND to develop their independence, to the point that one-to-one support was no longer needed, for instance: “they’re seeming to not be relying as much on the one to one support in the classroom... so they’ve not got an adult with them explaining the task and everything while the game’s being played and that seems to be increasing their independence” (school ci603, GBG lead).

Teachers commented on a range of benefits of the GBG for pupils with an SEND in their class. Contrary to some teachers’ reports of the rigidity of the game being an issue, others felt that fidelity to the highly structured nature of the game helped certain SEND pupils to “thrive” (school ci603, teacher vd8979), with one teacher commenting that “the children that it’s been more beneficial to are the are the children who struggle when they don’t have structure so like... the children who are quite clearly SEN in my class that I think it’s probably been better for them” (school
ci603, teacher vd8979). This teacher felt that the rigid structure and clear rules helped these pupils, as did the team work element, although they did not go on to explain the ways in which this benefited pupils:

“There’s a lot of children that thrive on structure and because it gives... such a basis of structure... it’s worked really well... Yeah again the structure really helps them so I’ve got... eleven children on the SEN register in my class but... just because the rules are so explicit and because it like the directions are so explicit they just really thrive under it and know exactly what to do and like the... team work element’s been massive for them”

Some teachers also commented on benefits for pupils with specific SEND diagnoses, namely autism. One teacher commented that the game had “worked really well” (school ci603, teacher gb6981) for a pupil in their class with an autism diagnosis, whilst another reported that although they had felt it necessary to adapt the game for one of the pupils in their class with autism, this pupil had benefited from the intervention:

“[He] doesn’t like praise... or, I suppose, sanctions either because he doesn’t like that focus, so if he is breaking a rule I’d make a more general comment rather than showing him that he’s broken the rule and then obviously with praise... he just... colours in his booklet. He doesn’t like stickers, he doesn’t like things that, but he does like the idea they won the game so we will show that... He’s responded really well that he’s now team leader, so yes he’s done really, really well with that” (school ci603, ka0624).

Although these teachers were reporting general benefits, they typically did not discuss specific outcomes of the game for these pupils. However, one teacher who felt that “the child I see the most benefit is the little girl with autism” and that it has been “a positive thing” for her did go on to explain that they thought the GBG had helped this pupil to control her behaviour:
“Usually she’ll start crying and having... a bit of a paddy and everyone knows about it and it’s everyone’s fault and stuff and... I kind of told her she’d broken a rule and I saw her about to do that but then she kind of kept it in herself and... didn’t say anything throughout the game, didn’t break any more rules... so she like kept it in for that whole time... she was able to kind of contain it a little bit more” (school ci603, teacher vd8979).

This teacher also discussed a pupil in their class with “attention problems”, suggesting that the game had helped to improve their behaviour. They commented that although they initially thought that playing the GBG “could be a bit of a challenge” for them, they instead found that “he seems to thrive under The Good Behaviour Game as well, he seems to be at his calmest while we’re playing the game”.

8.3.5 White EAL

Whilst teachers did not make reference to pupils’ ethnicity when discussing the benefits of the GBG, teachers from two schools did discuss the differential effects on pupils who were classified as EAL. One teacher reported benefits to pupils’ behaviour; they felt that it had improved due to the explicit nature of the GBG during the pre-game stage, and the fact that adherence to the manual required pupils to work independently, without interaction with their teacher:

“The couple of children that I’m thinking of are EAL children... and I think the... reasons for their behaviour before, were they weren’t working independently. I think the fact with the Good Behaviour Game you go through the task in so much detail, I think they’re a lot clearer about what’s expected of them, whereas in an ordinary lesson I wouldn’t go into as much specifics about what exactly they need to do and I wouldn’t check six times that everybody knows what they’re doing, so I think that’s why their behaviour has improved and they’re able to work more independently,
because... you make it so clear when you’re doing the Good Behaviour Game”
(school gy985, teacher fb2388)

This teacher was still reporting these differential benefits after one year of implementation, commenting that “it takes extra time but I think, the work, it’s worth it... it definitely helps sort of the lower and the EAL children who’ve struggled to understand, so it’s especially good for them I think”. They went on to discuss how high fidelity to the manual meant that they were clear with their directions during the pre-game stage, as instructions could not be clarified once the game was in progress; they thought that this helped EAL pupils to understand the task, which may in turn help them to complete their work, thus increasing their attainment. They also felt that the team aspect of the game encouraged EAL pupils to take responsibility for their own work and behaviour:

“Just because... I can’t intervene during the game I need to make absolutely sure that they know what they’re doing before they start and they need to take ownership of that as well because it’s their responsibility, if they don’t understand what they’re doing they’ll get strikes for the team so they’ve kind of switched on a little bit more and listening more.”

However, not all teachers agreed with these comments; a teacher in another school discussed difficulties they had experienced with EAL pupils who were new to the school, finding that “it’s been quite difficult integrating them into the game” (school tm517, teacher nk6311), suggesting that this teacher was not seeing any benefits for EAL pupils. They went on to report “it’s been quite challenging to... make sure that they understand it on a level that’s kind of applicable to them”. Whilst these comments are contrary to other teachers’ reports of benefits to EAL pupils, both comments reflect a similar theme whereby teachers found that the GBG required them to be clearer with the directions that they provided to pupils, particularly those that classified as EAL. However, whilst the teacher in the second school did not view this as beneficial to the pupils, they had “only started last week”, which may have affected their reports on this topic.
8.3.6 School-Level Behaviour

School-level behaviour was touched upon by some of the teachers in the case study schools, and was mostly discussed in terms of reasons for implementing the GBG. Whilst one school felt that school-level behaviour was not a problem, and instead implemented the GBG to develop their “understanding of managing behaviour” in order to keep it “fresh and improving” (school tm517, head teacher), most felt that the GBG would be useful for targeting school-level behaviour. One teacher appeared to be concerned that the disruptive behaviour the pupils were displaying would create problems for them later in life, and hoped that the GBG would help with this issue: “disruptive behaviour and low level behaviour... has an impact on the children’s learning and I think... in terms of... the outlook for the children long-term looking at the research it seems to have a positive impact in later life so it was something that we wanted... to try” (school gy985, teacher fb2388). Others hoped that the GBG could be utilised as a whole-school approach in order to improve poor behaviour in the school: “there's quite a lot of bad behaviour in the school, not necessarily in our year group which is unfortunate but... in the school... there can be some bad behaviour so... if there's something that works well then we wouldn't mind kind of using it across the whole school” (school ci603, teacher vd8979).

Alongside disruptive behaviours, some teachers also noted that the schools had problems with adaptive learning behaviours. One explained that the school “needed a shift on attitude to learning and for the children to be independent learners” (school gy985, GBG lead); whilst another commented “we definitely want to further develop learning behaviour not just behaviour around school” (school tm517, GBG lead). Teachers thought that the GBG might help to resolve some of these issues, by encouraging pupils to actively engage in their learning:

“There were issues around behaviour and actual engagement in what we would call the independent learning... we felt that this was the next thing that could explore to actually... ensure that children became engaged learners in their own rights rather than being... made to learn... It was a
good point at this school to enter into this because we still had pockets of issues that hadn’t really been sorted out... although it’s not bad behaviour it’s slightly disruptive all the while, low level... there was that underlying not quite focused on learning or understanding of learning and... we just felt that this Good Behaviour Game would teach them the skills not to disrupt their own and other’s learning.” (school gy985, GBG lead).

8.3.7 Deprivation (Familial and Neighbourhood)

Familial-level poverty and wider neighbourhood deprivation were often discussed interchangeably in interviews in the case study schools, and were most frequently cited by members of the senior leadership team when discussing reasons for implementing the GBG. For instance, one head teacher explained how they felt that family income influenced the types of behaviours pupils displayed in school, and hoped that the GBG would encourage these pupils to become more engaged in their learning:

“There is an issue with behaviour management in the school, it’s not what... people term ‘naughty’ behaviour but... when I... go to other schools so, for instance, we link with (primary school) where I would say their intake is probably more middle class, the actual behaviour in class is different because they arrive expecting to learn, whereas here there isn't that inward trained up, it’s not even training, there’s not that innate... engagement so although it’s not bad behaviour it’s slightly disruptive all the while, low level, you know, so they don’t see the point of it, it’s not been sort of pushed” (school gy985, GBG lead).

Deprivation was also often discussed in terms of broader proxy risk variables, such as parental education level or drug and alcohol problems. This means that multiple risk variables were mentioned collectively when teachers discussed pupil deprivation. For instance, one teacher described how both familial-level and neighbourhood poverty in the area were also associated with other risk variables:
“It’s very deprived area really… it’s mainly… Local Authority housing… lots of the children are coming from backgrounds where there's drug and alcohol… addiction and also it is very high in domestic violence… The children are coming in with lots of issues really from family background… it’s in the bottom nought point nought one per cent of deprivation… in the whole country… we’re forever referring through for… food banks… really the children they struggle, the families struggle… lots of single parents too” (school ci603, GBG lead).

Teachers commented on specific risk variables that they felt were accompanying deprivation, and explained that they had decided to implement the GBG to “combat” the issues associated with them. One teacher felt that pupils from deprived homes and neighbourhoods faced certain “barriers” that stemmed from parents’ education levels and parental engagement with the school, which in turn influenced pupils’ academic attainment; they thought that the GBG would help to overcome these problems:

“Not all of our parents are… not high earners or anything like that… our school has a lot of parents who have not got the best educational start themselves… may not be in high paid employment or employment at all, quite a lot of our parents are on benefits, we’re in that sort of social deprivation factor, we score quite high on that, and I felt the Good Behaviour Game would enable us to break down certain barriers. Some of our parents are not always the most supportive in education because they didn't value education themselves and therefore don’t turn up for parent’s evenings, don’t read with their children, or haven't got the academic skills to read with their children. Some of our parents do really struggle with their… reading and writing and Maths so the Good Behaviour Game is a way of trying to combat, if you will, not negative influences but influences that are not supportive for our for our kids” (school hd887, GBG lead)
Another teacher went on to discuss how being in “an area of high social deprivation” meant that the school was “dealing with parents that have issues around alcohol and drug abuse and unemployment”; similarly to the GBG lead in school hd887, they too hoped “if we could break that trend for the children that would make a significant difference... to the outcomes for their future life” (school gy985, GBG lead). Some teachers made the link between factors associated with deprivation and behaviour in school, commenting that structure of the GBG was needed to help pupils from deprived neighbourhoods to progress academically, for instance:

“We decided to do it because we come from, our school has a certain level of... deprivation and some of those factors of deprivation are about the influences that they have outside of the school and sometimes those behaviours can bring themselves into the school so we thought it would be a good... idea while trying to raise attainment and progress within the children to give them some structures within that to help them modify their own behaviours, and that seemed to fit that sort of ethos” (school hd887, GBG lead)

Other teachers did not necessarily comment on the link between deprivation and these proxy variables explicitly, but discussed issues associated with these proxy variables in the broader family and neighbourhood that meant pupils lacked structure, which could in turn influence their behaviour. For example, one teacher reflected that the structured nature of the GBG worked well in their class because their pupils lacked boundaries at home: “children respond to structure really well. Children go in search of boundaries if they’re not used to them, if they don’t have them at home” (school sa774, teacher pf2223). Other teachers’ comments mirrored this, suggesting that the highly structured nature of the GBG could give pupils a “sense of security” (school gy985, GBG lead) which could help overcome barriers to social and emotional wellbeing:

“Children need to feel safe and they need boundaries and... they’re almost begging, especially in a community like this where behaviour is so bad on the
outside, they're begging for boundaries and children need to know that no matter what adult they're talking to, no matter where they are in the school, those boundaries will be the same so that they know where... the line is... I mean really high pupil premiums... not many parents working at all... really high deprivation... awful behaviour in the communities and not got... good role models at all... we have to deal with... a lot of... social and emotional problems... before we can even look at education really in in the school”
(school tm517, GBG lead)

8.4 Chapter Summary

As noted previously, eight themes were identified in the qualitative dataset regarding differential effects of the GBG for pupils at-risk. In addition, several inductive themes emerged regarding both the GBG elements perceived to be underlying these differential effects, and the unanticipated proximal effects that may indirectly influence the more distal outcomes explored in the quantitative analysis (disruptive behaviour and reading attainment). To summarise these, a map is displayed below (see figure 8.2) highlighting the relationships between the risk variables and the emergent proximal outcomes, and the GBG elements perceived to be the mechanisms underlying these associations. For example, as shown on the map, teachers reported that being a member of a team encouraged EAL pupils to take responsibility for their behaviour, while the explicit nature of the game helped with their independence and understanding of the tasks.
Figure 8.2: Thematic map
9. Discussion

9.1 Chapter Overview

This chapter aims to provide a discussion of the quantitative and qualitative findings in the present study. The chapter begins with a summary of the findings, before discussing them in relation to the existing literature in the relevant fields; the quantitative and qualitative findings are also integrated. Conceptual and methodological limitations are discussed. Finally, the implications and directions for future research, alongside the study’s contribution to knowledge are examined.

9.2 Summary of Findings

9.2.1 RQ1

RQ1 explored the risk variables at the pupil- and school-level that were associated with higher disruptive behaviour scores and lower reading attainment scores.

Regarding behaviour, six variables emerged as significant predictors of higher disruptive behaviour scores, four at the pupil-level, and two at the school-level, specifically: being male, FSM eligibility (familial deprivation), having an SEND, being a looked-after child, being in a school with a higher proportion of conduct problems, and attending a school with a higher proportion of pupils speaking EAL.

Regarding attainment, seven variables emerged as significant predictors of lower reading point scores, six at the pupil-level and one at the school-level, specifically: being male, FSM eligibility (familial deprivation), having an SEND, younger relative age in the class (summer-born), being White and speaking EAL, having a higher IDACI score (neighbourhood deprivation), and being in a school with a higher proportion of pupils speaking EAL.
The results indicated that cumulative risk score was a significant predictor of both disruptive behaviour and reading point scores; thus, when a pupil was exposed to a higher numbers of risk variables, their disruptive behaviour scores were higher and reading point scores lower. The functional form of the relationships between cumulative risk and the outcome variables were tested. The squared term explained additional variance in disruptive behaviour and reading point scores, after accounting for the linear cumulative risk score, suggesting that a disproportionate, quadratic relationship was present for both outcome variables.

The squared terms of the cumulative risk scores were significant predictors of both behavioural and reading outcomes, even after controlling for the variance explained by the individual risk variables. This suggests that the cumulative risk score accounted for additional variance over and above that explained by the individual risk variables.

9.2.2 RQ2

RQ2 explored the differential effects of exposure to the GBG on disruptive behaviour and reading attainment for pupils at varying levels of risk exposure. RQ2a explored teachers’ and pupils’ perceptions of impact for pupils exposed to various risk variables.

MLMs revealed that there were no significant subgroup effects of the GBG on pupils’ disruptive behaviour or reading scores. However, there was a marginal, non-significant trend which indicated that being in the low-risk group and being exposed to the GBG was associated with higher mean reading scores. All findings were completely insensitive to the use of complete case analysis as opposed to MI.

Eight themes were identified in the qualitative dataset; namely gender (male), SEND status, EAL status (although not ethnicity), school-level behaviour, and both familial (FSM eligibility) and neighbourhood (IDACI score) deprivation. Somewhat broader themes of pupils with behavioural problems and pupils with low academic
attainment also emerged inductively. Teachers and pupils identified five key GBG elements (team leadership, team membership, the explicit nature of the GBG, lack of teacher-pupil interaction permitted, and the rigid structure of the game) that were theorised to be the mechanisms through which pupils’ outcomes were influenced. Ten prominent proximal outcomes emerged and related to a broad range of perceived benefits including self-esteem, social skills, and engagement with learning.

9.2.3 RQ3

RQ3 explored the differential associations between implementation variability, specifically FidQual and dosage, and disruptive behaviour and reading attainment, for pupils at varying levels of risk exposure.

Neither FidQual nor dosage were found to be significant predictors of pupils’ disruptive behaviour scores. There were also no statistically significant beneficial subgroup associations between FidQual or dosage and differential gains in disruptive behaviour scores for pupils in any of the risk groups. However, the model revealed that, contrary to predictions, high FidQual classrooms in year 3 were significantly associated with higher disruptive behaviour scores among high-risk pupils. Regarding dosage, there was a marginal trend in the year 3 model indicative of lower disruptive behaviour scores among high-risk pupils in high dosage classrooms.

There was a marginal trend indicative of an association between higher reading point scores and high FidQual classrooms in the year 4 model only. Dosage was not a significant predictor of reading attainment. There were also no statistically significant beneficial subgroup associations between FidQual or dosage and reading scores for pupils in any of the risk groups.

Some findings in RQ3 were sensitive to the use of complete case analysis as opposed to MI (see appendix 12). Specifically, these sensitivities were restricted to
interactions between implementation and high-risk status for disruptive behaviour outcomes, whilst reading attainment outcomes were sensitive both to general interactions between implementation and reading scores, and specific subgroup interactions between implementation and risk status.

9.3 Research Question 1: Risk

1. What are the risk variables associated with lower school functioning?
   a. For those variables identified, what is the magnitude of their association with behavioural (disruptive behaviours) and/or academic (reading) outcomes?

The following sections below outline the risk variables associated with disruptive behaviour and reading attainment in the present study, and discuss them in relation to the previous literature. However, in order to avoid unnecessary repetition, an explanation of the mechanisms through which these risk variables are thought to influence outcomes is not included here, as a detailed account can be found in section 2.5.

9.3.1 Findings Consistent Across Both Models

9.3.1.i Gender (Male)

Pupils’ gender, specifically being male, was significantly associated with both higher disruptive behaviour scores and lower reading point scores. This finding is consistent with a wealth of other studies that have found similar differences in outcomes between male and female students regarding their behaviour and academic attainment. Evidence in the literature suggests that boys are twice as likely as girls to display behaviour problems (Hansen et al., 2010), and disruptive behaviour amongst boys accounts for large proportions of exclusions from school in the UK (Department for Education & Skills, 2007). Research also suggests that boys consistently “underachieve” relative to girls in school-based examinations (DfE,
2017b), although others maintain that this academic “gender gap” is not as wide as it appears (Gorard et al., 2001). However, the finding in the present study contradicts some of these arguments; whilst some suggest that it is only boys of certain social classes and ethnic minority groups that are underachieving (Department for Education & Skills, 2007), being male was found to be a significant predictor of attainment even after accounting for these other variables. However, it is worth noting that that reading scores were utilised as a proxy for “academic attainment”, and so this finding may not be representative of boys’ achievement more broadly.

9.3.1.ii SEND Status

SEND status was the predictor variable with the strongest association with lower reading point scores, and was also the second strongest predictor of higher disruptive behaviour scores. This finding thus corroborates previous research in this area. SEND status has been identified as a major risk variable for behavioural difficulties, with a national study finding that over half of adolescents who met the clinical criteria for conduct problems were identified as having an SEND (Murray & Greenberg, 2006; Oldfield et al., 2015). Pupils with an SEND are also considered to be the most vulnerable group of learners regarding academic outcomes (Humphrey et al., 2013). However, SEND is a broad category, with individuals falling into one of several distinct subgroups, each with varying outcomes (DfE, 2015c). For instance, being diagnosed with ADHD is associated with an increased likelihood of both academic difficulties and disruptive behaviour problems (Loeber & Keenan, 1994; van Lier et al., 2004), whilst some diagnoses are more commonly associated with only one outcome (e.g. dyslexia and reading; Miles, Haslum, & Wheeler, 2001). Thus, it cannot be guaranteed that the findings from the present study are applicable to all SEND pupils equally; more research could be conducted to explore variations in outcomes between pupils with different categories of need.
9.3.1.iii Familial Socio-Economic Deprivation (FSM Eligibility)

Pupils’ FSM eligibility (used as a proxy for familial deprivation) was found to be significantly associated with both higher disruptive behaviour scores and lower reading point scores. The findings in the present study are consistent with results in the extant literature regarding behaviour, suggesting that pupils from poorer backgrounds are more likely to experience problems with conduct, aggression, hyperactivity and anti-social behaviour (DfE, 2012; Kellam et al., 1998). Concerns regarding the attainment gap between pupils of low and high SES are also frequently reported, with pupils eligible for FSM consistently experiencing lower grades throughout their education (Smith, 2003; Strand, 2015). Thus, the findings from the present study align with these issues.

9.3.1.iv School-Level Linguistic Diversity (EAL)

A high proportion of pupils speaking EAL (used as a proxy for school-level linguistic diversity) was the only school-level predictor found to be significantly associated with both disruptive behaviour and reading attainment. The effects of school-level linguistic diversity are an extremely neglected area of research, and so it was previously unknown whether this variable would have an effect on pupils’ school functioning. The limited number of studies that have been conducted in this area are contradictory; while some have found that having high numbers of pupils speaking EAL in the class has a detrimental effect on the test scores of non-EAL pupils, particularly regarding reading outcomes (Cho, 2012), others have found no negative impact on attainment (Strand et al., 2015). Therefore, this finding in the present study provides a valuable contribution to this area, helping to bring clarity to inconsistencies in the literature regarding attainment.

Regarding behaviour, one study conducted by Gottfried (2014) explored the relationship between EAL classmates and pupils’ socio-emotional skills; results suggested that pupils with a greater number of EAL classmates had fewer externalising behaviours and higher social skills, although effect sizes were
extremely small. The finding in the present study therefore contradicts these results. However, Gottfried’s study utilised younger children; thus it may be that the age of the pupils has an effect on outcomes for this risk variable; for instance, it is possible that these risk variables are more influential during the later years of schooling, or that a variable is a risk variable at a certain time point in development, but a promotive factor in another (Wright et al., 2013). This is something that should be explored further in future research.

9.3.2 Unique Predictors of Disruptive Behaviour

9.3.2.i Looked-After Child (CLA Status)

Being a looked-after child was the strongest predictor of higher disruptive behaviour scores. Based on the evidence available in the extant literature, CLA status is a well-established risk variable for school functioning; thus the findings in the present study regarding behavioural outcomes were expected. For instance, almost half of children in residential care have been reported to have difficulties with behaviour (Attar-schwartz, 2009). However, differential effects have been found dependent on the category of need (that is, the reason social services first became involved with the child). For example, 61% of children identified as CLA are categorised as being abused or neglected (Zayed & Harker, 2015). As previous studies have found that around 60% of abused or neglected looked-after children have conduct problems (Minnis & Devine, 2001), it may be that the findings from the present study are not applicable to all children identified as CLA. Indeed, abuse and neglect have previously been found to be risk variables, independently predicting negative outcomes regardless of CLA status (Christiansen et al., 1997; Luntz & Spatz Widom, 1994; OPRE, 2012; Sullivan, 2011), and thus it may be that CLA status was acting as a proxy variable for abuse and neglect in the present study. Further exploration may be warranted as to what exactly CLA status is truly measuring. It may be that CLA status alone is too broad, as it encompasses several other potential risk variables that could be inflating its effects on children’s
outcomes. This may explain why this variable was such a strong predictor of behavioural outcomes in the present study.

The attainment gap between looked-after and non-CLA pupils is persistent in both Literacy and Numeracy, at all Key Stage levels (DfE, 2014a); thus the finding in the present study that CLA status does not predict reading attainment contradicts previous research. However, two-thirds of looked-after children are also identified as having an SEND (ibid). As SEND status was already controlled for in the reading attainment model, this may explain why CLA status did not explain any unique variance, and hence why it was not a predictor of reading point scores.

9.3.2.ii School-Level Behaviour (Proportion of Pupils Borderline or At-Risk for Conduct Problems)

Attending a school with a higher proportion of pupils above the borderline threshold for conduct problems (teacher-rated) was associated with higher disruptive behaviour scores at the pupil-level. Research in this area suggests that attending a school with high levels of behaviour problems has a negative effect on pupils’ individual behaviour and thus the finding in the current study is consistent with the previous literature. However, the majority of studies exploring this topic utilise aggression as the outcome variable (Barth et al., 2004; Kellam et al., 1998; Thomas & Bierman, 2006; Thomas et al., 2011), and so the effect that school-level conduct problems would have on pupils’ lower-level behaviours was previously unknown. It is also interesting to note that school-level conduct problems significantly predicted a different type of behavioural difficulty, disruptive behaviour; the majority of studies typically only utilise one measure of behaviour. However, there was some overlap between the items on the two measures utilised in the present study, and so in reality they may have been measuring the same type of behaviour (e.g. SDQ conduct problems: ‘often fights with other children or bullies them’; TOCA-C disruptive behaviour: ‘fights’; Goodman, 1997; Koth, Bradshaw, & Leaf, 2009); indeed, scores were strongly correlated.
Previous research in the field has found that schools with high disorder, including aggression, have been shown to have lower school-level achievement (Barnes et al., 2006). Classroom-level measures of academic focus also appear to significantly predict individual academic focus (Barth et al., 2004). Thus, the results of the present study, suggesting that school-level behaviour does not predict reading attainment is somewhat incongruent with the results of previous studies. However, neither of these studies utilised conduct problems as the measure of school-level behaviour, nor did they look at individual pupils’ academic attainment.

9.3.3 Unique Predictors of Reading Attainment

9.3.3.i Relative Age (Summer Born)

Being born in the summer months was a significant predictor of lower reading point scores. This finding is consistent with previous research which suggests that young relative age is a risk variable for academic outcomes (Lien et al., 2005). However, studies that have explored this have also found that relative age is associated with an increase in the likelihood of being diagnosed with an SEND, particularly those that are linked to literacy attainment (Squires et al., 2012). Thus it may be that relative age is a risk variable for SEND diagnosis, and pupils’ academic outcomes are influenced through this mechanism. However, as SEND status was controlled for in the present study, this indicates that this was not the case. Indeed, SEND status and relative age both contributed unique variance regarding pupils’ academic outcomes.

Regarding behaviour, the findings were generally inconsistent with the extant literature; previous studies have typically suggested that being born later in the academic year is associated with an increase in behavioural disorders (Morrow et al., 2012; Polizzi et al., 2007). However, Lien and colleagues’ study (2005) found that relative age was only significantly associated with academic outcomes, and not behaviour, thus supporting findings in the present study. Additionally, Patalay and colleagues (2015) found that only internalising behaviours were significantly associated with relative age, and so it may be that the measure of behaviour
chosen influences results; perhaps certain types of behaviour are more susceptible to the relative age effect. As the theories for relative age include factors such as less mature central nervous systems, poorer social skills and lower self-esteem (Polizzi et al., 2007), it could be expected that certain internalising behaviours such as concentration problems, irritability, social withdrawal and loneliness may be more prevalent amongst these relatively younger children.

However, what is unique about the findings in the present study is that not only was younger relative age not a risk variable for disruptive behaviour, it was actually promotive, albeit with a small effect size. That is, pupils born in the summer months exhibited significantly lower levels of disruptive behaviour. This is contradictory to much of the research in the field and the reasoning is unclear. It may be that the binary measure of relative age (summer-born vs. other) amplified the effects. Further research could be conducted in this area to establish the point at which difficulties arise.

9.3.3.ii Ethnicity and Linguistic Diversity

Following issues surrounding multicollinearity, EAL and ethnic minority predictor variables were grouped, with pupils falling into one of four categories. This is consistent with Strand and colleagues’ argument (2015), whereby they suggested that EAL and ethnic minority status are very closely related, and that EAL simply acts as a proxy for ethnic minority status.

Being categorised as White and speaking EAL was associated with lower reading point scores, relative to the White and non-EAL category. Conversely, being identified as a member of an ethnic minority group and being non-EAL was significantly associated with higher reading point scores. This suggests that it is predominantly pupils from other European countries currently being educated in the UK (ibid; Tereshchenko & Archer, 2014) that are most at-risk for issues with attainment.
It is unsurprising that pupils whose first language is not English do not score as highly on reading tests as native English speaking pupils; this has frequently been shown to be the case in previous research, particularly with younger pupils (DfE, 2015a; Strand et al., 2015). However, with Government statistics indicating that pupils from ethnic minority groups are also less likely to achieve the GCSE benchmark (DfE, 2015a), the finding that White pupils specifically are significantly more likely to experience lower reading scores than ethnic minority pupils was unanticipated. Research typically indicates that pupils in ethnic minority groups experience lower grades than White British pupils (ibid). However, utilising a binary measure of ethnicity in the present study (White vs ethnic minority) may have biased the results, by not allowing for more nuanced differences between ethnic groups (Strand et al., 2015). Indeed, Strand and colleagues (ibid) reported that EAL pupils in the “White Other” category were 17.4% less likely than White British EAL pupils and 12.5% less likely than White British non-EAL pupils to achieve the KS2 benchmark at the end of primary school. In fact, only Traveller Irish and Black Caribbean EAL pupils were less likely to achieve this benchmark. This highlights the importance of both investigating pupils’ ethnic group in conjunction with their EAL status, and exploring more specific ethnicity categories, to gain a clearer indication of the groups of pupils most at risk for academic difficulties.

There has also recently been an increased focus in the media on “White working class boys”, with reports suggesting that they are the worst performing of the main ethnic groups in the UK (e.g. Rigby, 2016; Telegraph-Reporters, 2016; Weale & Adams, 2016); the finding in the present study regarding ethnic minority status being ‘promotive’ is therefore supported by these reports. However, these reports indicate that exposure to other mediating variables such as deprivation may be explaining the differences in outcomes amongst pupils of different ethnic groups (Department for Education & Skills, 2007; Strand et al., 2015). Although several of these factors were controlled for in the present study, the inclusion of potential risk variables was not extensive, and so it is possible that another variable was influencing the findings. An alternative explanation for the incongruous findings between the present study and earlier research is the recent increase in pupils
emigrating from other European countries to the UK. Indeed, pupils speaking Eastern European languages are the fastest growing group in English schools (Evans et al., 2016; Tereshchenko & Archer, 2014), and Strand and colleagues recent work (2015) found a similar issue regarding attainment for White EAL pupils. Thus it may be that there were not such high numbers of White EAL pupils from other European countries in earlier studies, making their findings less comparable with those of the current study.

However, it is interesting to note that the same effect was not present for disruptive behaviour scores, although the previous literature suggests that ethnicity and EAL status are risk variable for behavioural problems (Hansen et al., 2010). In the current study, being identified as a member of an ethnic minority group and speaking EAL was significantly associated with lower disruptive behaviour scores. However, similarly to arguments surrounding attainment, other factors may be moderating the differences in behavioural outcomes between different ethnic groups (Guttmannova et al., 2007; Leadbeater & Bishop, 1994). For example, Black African and Black Caribbean pupils are more likely than other groups to live in areas of high deprivation (Department for Children Schools and Families, 2009); therefore, it may be that after controlling for other variables such as familial SES and gender, ethnic minority pupils have the most favourable outcomes. Furthermore, variations in parenting practices and attitudes towards schooling can differ across ethnic groups; for instance, Chinese pupils consistently achieve above the NA in the UK, regardless of their FSM status (Kirby & Cullinane, 2016), and this is thought to be due the cultural value placed on education (Francis & Archer, 2005). It may be that these variations are lost through the use of the binary measure of ethnic minority status utilised in the present study, such that being in the overall ethnic minority group actually resulted in favourable outcomes.

9.3.3.iii Neighbourhood Deprivation (Higher IDACI Score)

High IDACI scores (used as a proxy for neighbourhood deprivation) were associated with lower reading attainment in the present study. Interestingly, IDACI score was a
significant predictor of reading attainment, even after accounting for FSM eligibility, suggesting that neighbourhood-level and familial-level deprivation contribute unique variance to pupils’ reading attainment. This finding is consistent with previous literature that suggests that pre-schoolers living in affluent areas have more positive gains in IQ and verbal ability (Kohen et al., 2009). There is also research to suggest that this effect is sustained over time, with evidence indicating that differences in attainment are still present at GCSE (DfE, 2015a).

However, contrary to previous research, effects of neighbourhood deprivation were not found in the present study for disruptive behaviour scores. Studies generally suggest that living in a low-income neighbourhood is associated with higher levels of CD (McCulloch, 2006) and externalising problems (Kohen et al., 2009). One explanation for these incongruent findings is simply that after accounting for familial deprivation, IDACI scores did not explain any additional variance in pupils’ disruptive behaviour scores. Conversely, it may be that other risk variables not measured in the present study also associated with high deprivation neighbourhoods (e.g. parental engagement) were mediating the effects (McCulloch, 2006).

9.4 Research Question 1: Cumulative Risk

9.4.1 Cumulative Risk – Assumption 1

b. *Is there evidence of a cumulative effect of risk exposure on these outcomes?*

Pupils’ cumulative risk scores were significantly positively associated with disruptive behaviour scores; this suggests that as the number of risk variables that a pupil is exposed to accumulate, the likelihood of higher disruptive behaviour scores also increases. Additionally, pupils’ cumulative risk scores were significantly negatively associated with reading point scores, thus as the number of risk variables accumulated, the likelihood of lower reading scores also increased. This supports
the first assumption of cumulative risk theory: the greater the number of risk variables, the greater the prevalence of problems.

A wealth of previous research conducted in this area, utilising a variety of samples and outcomes, has found similar effects (e.g. Appleyard et al., 2005; Deater-Deckard et al., 1998; Gerard & Buehler, 2004a, 2004b). For instance, Oldfield and colleagues (2015) identified a cumulative risk effect for behavioural outcomes amongst children with SENDs, whilst Flouri and Kallis (2007) found that the number of adverse life effects was associated with psychopathology and adjustment. Whipple and colleagues (2010) also noted a significant association between cumulative risk across multiple ecological domains and school-wide achievement. However, academic achievement is explored relatively infrequently in comparison to behavioural outcomes, and so the findings from the present study contribute to the evidence base in this area by documenting cumulative risk effects in relation to reading.

The findings of the current study are consistent with the theory posited in cumulative risk literature that risk variables are not independent of each other and they work in a cumulative manner; exposure to each additional risk variable results in an increase in negative outcomes, irrespective of the nature of the individual risks (Oldfield et al., 2015). It is important to note that the likelihood of negative outcomes increased as the number of risk variables also increased in the present study, regardless of the nature of the specific risk variables, or in which ecological domain they were situated. This supports the argument that risk variables should not be investigated in isolation, and in fact, they may interact both within and across multiple ecological domains (Lanza et al., 2010; Whipple et al., 2010). Thus, examination of singular risk variables likely does not accurately describe their effects; instead, risk research needs to acknowledge the clustering of risk variables that occurs when considering pupils’ outcomes (Whipple et al., 2010).
9.4.2 Functional Form

c. What is the functional form of the risk-outcome relationship?

The squared term of pupils’ cumulative risk scores was significant when added to the disruptive behaviour and reading models, explaining additional variance after accounting for the linear cumulative risk score, thus indicating a quadratic relationship. This means that there was a disproportionate change in disruptive behaviour and reading scores as the level of risk increased. There are currently an equal proportion of cumulative risk studies evidencing linear and quadratic relationships. However, in the majority of these studies, no formal tests of linearity were conducted; instead, the functional form of the relationships were inferred from interval changes in the data (Evans et al., 2013). These findings therefore contribute to the literature in this field, providing more robust evidence of a quadratic risk-outcome relationship for both disruptive behaviour and reading attainment.

The quadratic relationships between risk and pupils’ outcomes indicated that there was a disproportionate increase in behavioural and academic problems as pupils’ risk levels increased. This effect has been termed “mass accumulation” in the cumulative risk literature, and suggests that that total effect of individual risk variables on pupils’ outcomes is greater than the sum of its parts (Flouri & Kallis, 2007; Oldfield et al., 2015). This is consistent with Rutter’s original work on cumulative risk (1979), which found that the combination of stressors had much more than an additive effect on children’s psychiatric outcomes; instead, there was an interactive effect whereby the risk associated with several concurrent stressors was more than the sum of the individual effects. Rutter (ibid) posited that this effect was due to the presence of one stressor potentiating the damage caused by another. Indeed, quadratic relationships are thought to provide evidence that risk variables not only cluster and commonly co-occur, but also interact with each other in a multiplicative manner, having a disproportionately negative effect on outcomes (Gerard & Buehler, 2004b; Oldfield et al., 2015). Thus, the findings from the present
study support the notion that risk variables for behavioural and academic problems
do not occur in isolation, are not independent of each other, and should not be
treated as such (Flouri & Kallis, 2007).

In addition to the uncertainties in the literature regarding the function of the risk-
outcome relationship, it is also unclear whether a certain level of risk needs to be
reached before the effect is evident (Stouthamer-Loeber et al., 2002). Therefore, in
the present study, graphs displaying the risk-outcome relationship for both
disruptive behaviour and reading attainment were examined. Regarding the specific
nature of the risk relationship with disruptive behaviour, there appeared to be a
threshold effect after exposure to one risk variable, whereby risk was related to
outcomes through an accelerative function, resulting in an exponential increase in
difficulties. A similar effect was also present in the reading model; after exposure to
two risk variables there was a sharp decrease in reading point scores. This threshold
effect is consistent with previous research into the quadratic risk-outcome
relationship that suggests that there is a critical point after which children’s coping
strategies become overwhelmed, resulting in a sudden increase in problems (Flouri
& Kallis, 2007).

In the behaviour model only, this initial dramatic increase in behavioural problems
was followed by a plateauing after exposure to three risk variables. This “saturation
effect”, as it is often referred to, implies that the addition of risk variables did not
compound the influence of other stressors to any great extent, likely due to the
already highly compromised development of the child (Gerard & Buehler, 2004b).
Interestingly, this is consistent with several other studies that also noted an effect
at around three or four risks (Biederman et al., 1995; Jones et al., 2002; Rutter,
1979); however, whilst they typically note a threshold effect, there was evidence of
a saturation effect in the present study.

Whilst it is theorised that both threshold and saturation effects result in protective
factors and other coping mechanisms becoming overwhelmed and subsequently
exhausted (Flouri & Kallis, 2007), there is little research into the specific ways in
which this phenomenon occurs, and hence how it can be addressed. Although the findings from the present study support the arguments in the literature that it is important to utilise the number risk variables when exploring pupils’ outcomes, as opposed to the presence or absence of particular risk variables (Morales & Guerra, 2006), it is also important that more research is conducted in this area, to further establish the ways in which cumulative risk exposure has a disproportionate and deleterious effect on children’s coping mechanisms.

9.4.3 Cumulative Risk – Assumption 2

d. Is the number of risk variables a better predictor of outcomes than their relative independent strength?

The evidence from RQ1c that cumulative risk exposure was having an accelerative detrimental effect on school functioning appears to support the argument that the total effect of the number of risk variables is greater than the sum of its parts (Flouri & Kallis, 2007). However, it is imperative to establish whether cumulative risk score is a superior predictor of outcomes, after accounting for the individual risk variables, before it can be determined that the number of risk variables is more important than their nature.

Cumulative risk score was a significant predictor of both disruptive behaviour and reading attainment, even after accounting for the variance explained by the individual risk variables. Thus, the number of risk variables explained additional variance above and beyond that explained by the individual risk variables. Results therefore support the second assumption of cumulative risk theory: the accumulation of risk variables is more important than the presence or absence of particular risk variables. This finding is consistent with previous literature in this field, and is thought to be due to the principal of equifinality (Oldfield et al., 2015). This theory proposes that negative outcomes do not occur via one specific route, but rather several distinct pathways (ibid; Dodge & Pettit, 2003); there are
transactional effects between risk variables, whereby exposure to one risk increases the likelihood of the occurrence of others (Rutter, 1979).

By utilising a cumulative risk methodology, the natural co-variation of risk variables can be accounted for (Flouri & Kallis, 2007), meaning that this approach has superior power to explain more variance in outcomes. This counters previous criticisms of cumulative risk research that discuss the loss of potentially important information on risk variable intensity (Evans et al., 2013; MacCallum et al., 2002); instead, this finding is consistent with the argument that no one risk variable is more important than another. Indeed, it is the confluence of risk variables, rather than any singular risk, regardless of context, that leads to dysfunction (Flouri & Kallis, 2007).

The model of allostatic load, where more frequent exposure to stressors accelerates wear and tear on the body by engaging multiple bodily response systems, provides one possible explanation for the superior predictive power of cumulative risk (Evans et al., 2013). This suggests that it is not simply the summation of the individual risk variables that is having a negative effect on children, but it is the interactions between these risk variables that are placing additional stress on children’s bodily response systems. This explains why the cumulative risk scores in the models still accounted for additional variance, even after controlling for the effects of the individual risk variables, and therefore supports arguments in the literature for utilising a multiple model of risk (Evans et al., 2007). However, research into this phenomenon, particularly focusing on non-biological outcomes, is lacking, and so the validity of this explanation, and the exact mechanisms through which children’s response systems are affected, are unknown. With further research in this area, perhaps more could be done to counteract the disproportionately negative effects of cumulative risk exposure for children.
9.5 Research Question 2: GBG

9.5.1 Impact of the GBG

2. Are there differential intervention gains in behavioural and/or academic outcomes among children at different levels of cumulative risk exposure?

Being allocated to the GBG condition in the trial was not found to be a significant predictor of pupils’ disruptive behaviour scores; it was also not a significant predictor for pupils in any of the risk categories. Although the descriptive statistics suggested that pupils at high levels of risk exposure experienced a small decrease in mean scores, this reduction was not found to be significant. In other words, being exposed to the GBG for two years was not associated with any improvements in disruptive behaviour scores for pupils, regardless of their risk status. This finding is incongruent with the majority of GBG research, which generally suggests that pupils’ behaviour improves as a result of participating in the game. Flower and colleagues (2014) reported in their meta-analysis of the GBG that six of the eight studies where a measure of disruptive behaviour was utilised found the game to be effective. The GBG also has a long history of success regarding other behavioural outcomes including aggression (e.g. Dolan et al., 1993; Kellam et al., 1994, 1998) and conduct problems (e.g Ialongo et al., 2001; Reid et al., 1999; Spilt, Koot, & van Lier, 2013).

Additionally, intervention research typically suggests that behaviour improves for pupils exposed to certain risk variables (Wilson & Lipsey, 2007). A key finding in GBG literature is that boys benefit from the game significantly more than girls, noting reductions in aggressive behaviours (Dolan et al., 1993; Petras et al., 2008) and antisocial behaviours (Kellam & Anthony, 1998). Additionally, pupils with SEND also appear to benefit from the game, with improvements reported in a variety of behavioural outcomes (Huizink et al., 2009; Johnson et al., 1978; Spilt et al., 2013; Sy et al., 2016). However, although many behavioural outcomes were measured in these studies, few were specifically disruptive behaviour, and so accurate comparisons cannot be drawn between these studies and the findings from the
present study. It may be that the GBG improves other aspects of behaviour, such as aggression, that were not measured in the present study; perhaps had a different measure of behaviour been utilised, different results would have been obtained. However, although aggression and disruptive behaviour have different definitions (Wilson & Lipsey, 2007), the TOCA-C contains a “disruptive behaviour” subscale that is based on the “disruptive/aggressive behaviour” subscale of the TOCA-R, and so there is some overlap between measures of aggression and disruptive behaviours utilised in GBG studies. It is therefore not possible to accurately identify the types of behaviours that benefit most from the GBG.

Furthermore, of the studies that have utilised disruptive behaviour as an outcome variable, consistent measures have not been employed. For instance, whilst Sy and colleagues (2016) reported positive associations between GBG exposure and the disruptive behaviour of pupils with special educational and health needs, they only measured improvements in “target responses”, that is, behaviours they identified in individual pupils at baseline that they felt required improvement. This therefore limits the generalisability of the results, and so improvements in disruptive behaviour amongst pupils with SEND cannot be compared across studies. Additionally, while these studies reported benefits for pupils exposed to a single risk variable, outcomes for pupils exposed to multiple risk variables were not explored, and so it was unclear if the present study would find similar effects for pupils at varying levels of risk. GBG studies that have claimed to examine the outcomes for “at-risk” pupils have reported positive results in terms of these pupils’ conduct problems, anti-social behaviour and problem behaviour (Ialongo et al., 1999; Petras et al., 2008). However, they tend to use the term “at-risk” as a proxy for a single risk variable, such as children living in neighbourhoods with high levels of juvenile delinquency (Reid et al., 1999), or pupils deemed to be aggressive (Kellam et al., 2008). Studies have also reported the benefits of the GBG for pupils displaying “early risk factors” for substance abuse, smoking and achievement (Ialongo et al., 1999; Kellam & Anthony, 1998), but have not gone on to look at levels of risk. Furthermore, most studies have not explored differential subgroup
effects, failing to utilise a comparison group when examining the outcomes for at-risk pupils.

Thus, while the results from the present study are not consistent with the extant GBG evidence base, this study was the first to examine the differential subgroup effects of the GBG on pupils at varying levels of risk, and so it was unclear what the effect would be, if any. Muthén and colleagues (2002) hypothesised that the GBG would have the largest effect on pupils at medium levels of risk, due to the universal nature of the intervention not being strong enough to benefit those at the highest levels of risk, and not required for members of the low-risk group. However, the results of this study have indicated that this is not the case. Indeed, the GBG did not have any significant effect on pupils’ behaviour, regardless of risk status, although from the quantitative results alone it is not possible to determine why this was the case. It may be that the wrong outcome measure was utilised, that the GBG was not suited to UK schools, or that it was not implemented as intended.

Regarding reading, being allocated to the GBG condition in the trial was not found to be a significant predictor of pupils’ reading point scores; it was also not found to be a significant predictor for pupils in any of the risk categories. In other words, being exposed to the GBG for two years did not result in any improvements in reading point scores for pupils, regardless of risk status. However, there was a marginal, non-significant trend indicative of increased reading scores for low-risk pupils exposed to the GBG; this was also supported by the descriptive statistics. It therefore may be that while the GBG was having an impact on attainment for pupils in the low-risk group, the magnitude of this change was not practically meaningful; indeed the effect size identified was only .115. However, considering the effects of the GBG on pupils’ attainment are deemed to be secondary to behavioural outcomes and are thought to be a more medium-term outcome (Chan et al., 2012), these results are promising for low-risk pupils; indeed, it is possible that these effects may attenuate over time in line with the GBG logic model. Therefore, a longer-term follow-up is warranted to establish if these effects are sustained or amplified in the longer-term.
There is very little literature available regarding the effects of the GBG on pupils’ reading, and academic attainment more generally, and so it was unclear if the GBG would have an effect on reading point scores in the present study. However, as it is only hypothesised that the GBG will influence pupils’ academic attainment indirectly, by improving their adaptive learning behaviours (Coombes et al., 2016), it is unsurprising that there were no significant benefits to at-risk pupils’ reading in the present study. Furthermore, of the studies that have utilised academic outcome variables, the GBG has often been implemented alongside another academic curriculum, and so the specific effects of the GBG cannot be ascertained. For instance, while the results from the present study contradict Bradshaw and colleagues’ findings (2009), suggesting that the GBG was associated with increased scores on standardised tests, they implemented the GBG as part of a wider CC intervention, and did not test the individual effects of each programme. It therefore cannot be determined that the improvements in attainment were due to the GBG directly. However, as the GBG was indicative of a small effect on low-risk pupils in the present study, perhaps Bradshaw and colleagues were right to pursue a combination of the GBG with another academic intervention, in order to enhance the effects of the intervention on pupils’ attainment.

Conversely, the lack of rigor inherent in many of the previous studies may explain the incongruous findings with the present study. For example, failure to test specific aspects of a broader intervention (e.g. Bradshaw et al., 2009; Reid et al., 1999) means that the mechanisms of change could not be attributed to the GBG specifically. Furthermore, many studies did not utilise a control group (e.g. Donaldson, Fisher, & Kahng, 2017), or allocated pupils within one school to the different arms of the trial (e.g. Ialongo et al., 1999), meaning that accurate comparisons could not be conducted. Different versions of the GBG were also utilised in some studies (e.g. Dion et al., 2011), and fidelity to the manual was rarely monitored. Thus it may be that several of these studies involved a Type III error, that is, the inaccurate attribution of the cause of the results, due to weaknesses in
their methodologies (Lendrum & Humphrey, 2012). Had these studies conducted more robust evaluations of the GBG, different results may have been found.

Alternatively, it may that cultural transferability was an issue regarding the implementation of the GBG in the present study. Only one other small-scale pilot study conducted by the intervention developers has explored the effects of the GBG in a UK context. Qualitative data from this study indicated that teachers expressed concerns regarding several aspects of the GBG, including the time required to implement the game, the inflexibility, and the lack of teacher-pupil interaction allowed (Chan et al., 2012). If these issues were perceived by teachers to be incompatible with school culture in the UK, the GBG may not have been delivered as intended, thus diluting its effects (Wigelsworth et al., 2016). However, the results from the UK pilot were promising, indicating that scores fell on all nine TOCA-R behaviour subscales (Coombes et al., 2016), suggesting that the GBG can be successfully implemented in the UK and that cultural transferability was not an issue affecting the results. However, effect sizes typically decrease once an intervention is moved from the efficacy to the effectiveness stage, often due to the reduced involvement of the intervention developers (Beelmann & Lösel, 2006; Wigelsworth et al., 2016); although the present study was somewhat of a hybrid efficacy/effectiveness trial (it was implemented at scale but costs to schools were subsidised), this may explain the difference in findings between the two UK-based studies. Additionally, as the Oxfordshire research was a pilot study, the sample size was extremely small and no control group was utilised, limiting the extent to which conclusions could be drawn regarding the effects of the GBG. Thus it may be that once applied to a “real world” setting, with a more representative sample and without on-going guidance from the intervention developers, the GBG failed to be imported and implemented successfully in UK schools.
9.5.2 Perceptions of Differential Impact

a. In what way does exposure to the GBG influence behavioural and/or academic outcomes for at-risk children?

Teachers generally perceived the GBG to be benefiting pupils in their class, and felt that it had differential effects for pupils at-risk. Eight organising themes were identified in the qualitative dataset regarding these perceptions; six related to pupils exposed to specific risk variables, while two focused more broadly on pupils with behavioural and academic difficulties. Teachers and pupils also typically identified five key GBG elements that they perceived to be the mechanisms underlying intervention effects, the majority of which were consistent across several risk variables. Whilst participants infrequently commented on academic and behavioural outcomes specifically, they did refer to benefits regarding a range of other outcomes that they felt resulted from exposure to the GBG.

Teachers also made a variety of comments regarding issues surrounding fidelity to the manual and the level of dosage required, specifically for pupils at-risk. While they did not always elaborate on these in great detail, it was interesting to note that teachers felt that some aspects of the GBG protocol were not suitable for the at-risk pupils in their class, and so made adaptations to accommodate them. They also often felt that the length of time and frequency of games required was problematic, and so intentionally implemented with (presumed) sub-optimal levels of dosage. However, teachers did not always go on to expand on the ways in which differences in implementation influenced outcomes for at-risk pupils, and so more research is warranted in this area.

9.5.2.i Pupils with Low Academic Achievement

Whilst several teachers made comments around the effects of the GBG on ‘low ability’ pupils, there was no consistent narrative as to the types of effects observed; some felt that the GBG had no effect, others thought that it was beneficial, and
some felt that it was detrimental. Three elements were commonly discussed: team leadership, team membership, and the explicit nature of the GBG. Proximal outcomes included enhanced self-esteem, confidence, responsibility, independence and understanding of the task.

In terms of perceived benefits, some teachers did feel that the GBG helped low ability pupils’ behaviour and attainment. This is in line with the GBG logic model (see section 4.4), which suggests that pupils will evidence improvements in on-task behaviours and literacy and numeracy outcomes, and reductions in disruptive and aggressive behaviours. Furthermore, broader GBG research discussed previously in section 4.5 also suggests that the game will produce benefits regarding these outcomes. However, GBG research has not previously explored the differential effects of the GBG on low ability pupils, and so it was not previously known how pupils of differing abilities would respond to the game. Indeed, some teachers did note that the GBG caused problems for these pupils; it appeared that the lack of teacher-pupil interaction permitted during the game meant that teachers struggled to find tasks that low ability pupils would be able to complete independently, leading to pupils becoming demotivated and disengaged from their learning. This was also identified as an issue in the Oxfordshire pilot of the GBG, whereby teachers felt that the minimal teacher-pupil interaction had a negative effect on pupils, and meant that teaching opportunities were missed (Chan et al., 2012). Previous research more generally into teaching low ability pupils also indicates that key beneficial strategies include increased scaffolding, more feedback, and positive teacher-pupil relations (Dunne et al., 2007), all of which are absent from the GBG protocol.

Other elements of the GBG, such as the explicit nature and rigid structure of the game, also had conflicting reports as to the perceived benefits on low ability pupils. Factors such as the additional structured pre-game teacher input were thought to be beneficial, whilst the strict CCR script was thought to provide inadequate feedback for pupils that struggled to work without support. Independence featured frequently in this theme, and was discussed almost exclusively negatively; this is
somewhat unsurprising considering that additional resources, including TAs and smaller class sizes, are often provided for low ability pupils (Betts & Shkolnik, 2000; Dunne et al., 2007). Indeed, research suggests that additional support for low ability pupils has a direct positive impact on pupils’ progress (Webster & Blatchford, 2012). This may explain why some teachers felt the need to adapt the GBG by adding supplementary comments to the CCR script, or by talking to pupils during the game.

However, there is little evidence in the literature of any substantial gains as a result of increased support in terms of pupils’ broader behavioural, emotional and social development (ibid). This may explain why some teachers felt that the reduced teacher interaction and enhanced support from the team were beneficial for other outcomes such as self-esteem. There is a wealth of previous research that, contrary to popular opinion, suggests that grouping pupils by attainment is detrimental to low ability pupils (Betts & Shkolnik, 2000; Francis et al., 2017; Ireson et al., 2002) and that peer support is valuable (Dunne et al., 2007). Thus it may be that the mixed-ability grouping emphasised in the GBG manual (Ford et al., 2014) provides low ability pupils with positive role models and increased peer support, meaning that they are modelling adaptive learning behaviours and becoming less reliant on the teacher.

Furthermore, the team leadership aspect of the game was also thought to enhance pupils’ proximal outcomes such as self-esteem and confidence. The label of being a low ability pupil, and the experience of struggling in relation to other more able pupils, can be harmful to pupils’ self-concepts (Francis et al., 2017), and so this may explain why being allocated the role of team leader was perceived by teachers to be particularly beneficial to these pupils. Teachers felt that increasing low ability pupils’ self-concepts in this way had an indirect beneficial effect on their attainment; similar findings have been evidenced in previous literature, indicating that there is an association between higher overall self-esteem and increased academic achievement (Bahrami & Bahrami, 2015). However, not all teachers viewed the team leadership element positively, one commented that having a team leader caused low ability pupils to pass over responsibility and become less engaged in
their work. Thus, it appears that the team leadership element was only perceived to be beneficial when the low ability pupils themselves were allocated the role. This is consistent with previous research suggesting that giving low ability pupils autonomy in their learning, by providing them with some responsibility, can be beneficial when attempting to motivate these pupils (Dunne et al., 2007).

9.5.2.ii Pupils with Behavioural Difficulties

Generally, when teachers were discussing differential effects of the GBG for pupils with behavioural difficulties, they reported positive outcomes for these pupils. Teachers perceived the GBG to be beneficial, both for classes where behaviour was a problem, and for individual pupils with behavioural difficulties. The majority of perceived benefits appeared to be related to the social elements of the game, predominately team membership and team leadership. Proximal outcomes included social skills, confidence, leadership, and responsibility.

Teachers felt that the team membership element of the game improved pupils’ social skills by increasing their awareness of how their behaviour impacted on other pupils in their team. The GBG logic model (see section 4.4) outlines increased social awareness and improvements in functional skills, including concern for others, as immediate impacts of the game, and so teachers’ reports were consistent with the intervention’s underlying theory of change. It is interesting to note that while the logic model details improvements for all pupils, teachers reported these benefits specifically for pupils in classes with high levels of behavioural problems. However, this is consistent with previous research regarding pupils with behavioural difficulties, whereby these pupils are often characterised by a tendency to act on negative impulses, without any apparent attention to the effects that this behaviour has on others (Hastings et al., 2000). Therefore, the team membership element of the GBG may be helping to address these issues, by rewarding pupils for maintaining behaviour standards in their team (Ford et al., 2014). It has also been suggested that the relationship between behavioural difficulties and lack of concern for others can increase with age, resulting in anti-social behaviour issues in
adolescence (Hastings et al., 2000); thus implementing the GBG with these pupils from a young age may also have increased benefits in the long term.

Several teachers also commented on the team leadership element of the GBG, reporting that this was particularly beneficial for pupils with behaviour plans. They felt that allocating these pupils the role of team leader encouraged them to take responsibility for their own and others’ behaviour, and was a way of boosting their confidence. Previous literature suggests that disruptive behaviour can stem from issues surrounding anxiety or low self-esteem (Demirdag, 2015), and that the self-esteem of pupils with behavioural difficulties can be boosted by allocating these pupils responsibilities in the classroom (Roper, 2017). This may explain why teachers felt that pupils with behavioural difficulties responded so well to the role of team leader, and why they noted an improvement in confidence. It is noteworthy that neither confidence nor responsibility are listed as outcomes in the GBG logic model (see section 4.4), nor is team leadership cited as an underlying mechanism of change; thus this finding in the present study contributes to the evidence base in this area.

9.5.2.iii Gender (Male)

While the effects of the game on boys were discussed occasionally, the majority of comments were in relation to the intersection between gender and other risk variables, such as boys with behaviour plans, or those with additional needs. When teachers did discuss the benefits of the GBG for boys, they typically used generic statements, and did not go on to discuss the ways in which the game benefited them, or the mechanisms through which these benefits occurred.

Several teachers reported that they experienced difficulties encouraging the boys in their class to participate in the GBG, explaining that it was a challenge to get them involved. Some went on to describe how boys’ behaviour was disruptive when implementing the GBG, as they viewed it as an opportunity to test the boundaries of the game. Thus, these reports appear to suggest not that the GBG failed to
improve the behaviour of boys, but that it failed to engage boys sufficiently, and so the mechanisms of change detailed in the logic model (see section 4.4) could not be triggered. This suggests there was an issue with participant responsiveness when implementing the GBG, although it was beyond the scope of the present study to measure this, and so it cannot be determined that this was the cause of the teachers’ concerns. However, previous research has found boys’ desire to display ‘laddish’ behaviours in school, such as the intentional withdrawal of effort and overt rejection of academic activities, to be an issue in the classroom (Jackson, 2010); thus boys may purposefully have sabotaged the GBG in an attempt to display behaviours consistent with social constructs of gender.

This finding is incongruous with the majority of quantitative GBG research, which consistently demonstrates that the GBG has beneficial effects for boys, with reductions in aggressive and antisocial behaviours frequently reported. Additionally, not only does research suggest that the GBG benefits boys, but that it benefits boys significantly more than girls (Dolan et al., 1993; Kellam & Anthony, 1998; Petras et al., 2008). Conversely, in the present study, teachers and pupils actually reported differential beneficial effects for girls, commenting that social skills improved as a result of playing the GBG. However, GBG research tends to focus on aggressive and disruptive behaviours, and so the differential effects on other outcomes such as social skills have not been explored in as much detail. Nevertheless, the GBG logic model (see section 4.4) does indicate that the GBG will have an immediate impact on pro-social behaviours, and so this finding for girls is consistent with the underlying theory of change.

9.5.2.iv SEND Status

SEND status was one of the risk variables most frequently discussed by teachers when asked about differential effects of the GBG. However, views were mixed; whilst some perceived benefits for pupils with SEND, others commented that the game created problems for them. It is possible that this is a result of the way SEND status was treated in the present study. SEND is a broad term, and encompasses a
wide range of needs (DfE, 2015c), it may be that had specific SEND diagnoses been explored in more detail, a more consistent picture regarding the effects of the GBG would have been obtained. Additionally, teachers did not tend to describe specific outcomes when discussing the effects of the GBG for pupils with an SEND, but instead made more general comments, simply stating that the GBG benefitted or hindered pupils. However, social skills, general wellbeing, and independence were mentioned as proximal outcomes. Teachers theorised that these were influenced through several game elements including team membership, the explicit nature of the GBG, the rigid structure, and the restricted teacher-pupil interaction.

A frequently mentioned element of the GBG was its rigid and highly structured nature, with teachers often explaining that this was unsuitable for the pupils in their class with an SEND. In an attempt to overcome this issue, some intentionally adapted the GBG for these pupils. For instance, some teachers adjusted the parameters of the game, so that SEND pupils would not receive infractions for displaying behaviours that typically constituted a rule break. Others commented more specifically on the inflexibility of the game regarding the limited amount of interaction allowed with pupils. They felt that this was impractical for pupils with an SEND as they required more support, and so they adapted the game accordingly. Teachers would interact with pupils during the game, provide additional directions, allow pupils to ask questions, and have one-to-one TAs sat with these pupils. They would also implement with lower levels of dosage than is recommended in the GBG manual, as it was felt that the required game length was too long for pupils with an SEND to work without support.

There is limited literature available regarding the effects of the GBG on those with an SEND, although it typically suggests that it is beneficial to such pupils, and so these findings are inconsistent with previous research (Johnson et al., 1978; Sy et al., 2016). However, in some studies, effects were only found for boys, and were mediated by peer relations (Witvliet et al., 2009). Therefore, the other risk variables to which a pupil was exposed may have influenced whether teachers perceived the GBG to be beneficial. Furthermore, these studies did not examine levels of fidelity
and dosage when investigating the effects of the GBG for these pupils, and so it may be that these teachers also adapted the game to meet pupils’ individual needs; however, the quantitative nature of these studies did not allow for any adaptations to be identified. In addition, research argues that in-class support is the single most important factor in enabling pupils with an SEND (Clarke et al., 1999), and so it is unsurprising that teachers felt that these pupils experienced difficulties regarding the lack of interaction permitted during the game.

However, other teachers did report benefits of the limited teacher-pupil interaction for pupils with an SEND, explaining that they had noticed an increase in independence, as pupils learned to rely on their team rather than an adult to support them. This eventually led to the removal of one-to-one TA support with these pupils. Therefore it may be that adaptations were required initially, but that over time independence increased, making procedural fidelity more feasible. Indeed, lower demand for SEND support is listed as a more medium-term impact of the game in the GBG logic model (see section 4.4), and so it may be that more time is required before benefits are seen with SEND pupils. Thus, the timing of the interviews may explain the incongruous findings in the present study.

For some teachers, the highly structured nature of the game was also perceived as beneficial for pupils with an SEND, although one teacher did explain that it was the pupils who typically required high levels of structure who benefitted most. Certain SENDs, such as autism, are characterised by a need for routine, and advice is often given to teachers to provide additional structure for these pupils (Manikiza, 2015). This finding is therefore consistent with previous research in the field. Thus it may be that the type of SEND diagnosis influenced whether the GBG was perceived to be beneficial by teachers. Indeed, GBG research conducted in this area tends to focus on one specific SEND, such as pupils with emotional disorders (e.g Sy et al., 2016) or foetal alcohol syndrome (Wiskow et al., 2018). However, not enough research has been conducted in this field to ascertain the differential effects for pupils in the various SEND areas of need and support. It is also possible that the effects of the GBG vary based on the individual; even within a broad area of support,
pupils’ needs can vary greatly (DfE, 2015c), hence why each pupil will have their own individual education, health and care (EHC) plan. This would explain why although the same few game elements were discussed by teachers, there were no consistent findings regarding the perceived benefits or necessary adaptations.

9.5.2.v White EAL

The effects of the GBG on EAL pupils were discussed infrequently, although the majority of comments reflected positive differential effects. Teachers described how the explicit nature of the game, the lack of teacher-pupil interaction permitted, and the team membership elements of the game influenced EAL pupils’ independence, understanding of the task, and responsibility.

The main GBG element discussed by one teacher was the explicit pre-game stage of the GBG, whereby teachers are required to clarify tasks with the pupils, and go through any directions in detail, as they cannot provide any further instructions to pupils once the game has begun. The teacher felt that this enhanced EAL pupils’ understanding of the task, as instructions were not normally provided in so much detail. In the National Association for Language Development in the Curriculum’s (NALDIC) guidelines (South, 1999), developing learner independence is considered to be an underpinning principle for good practice when working with EAL pupils. NALDIC suggested that the teacher has a key role in encouraging EAL learners to become more independent, through the application of strategies that develop self-reliance. As the GBG logic model (see section 4.4) indicates that the game develops functional skills, such as independent work, previous research regarding effective pedagogical strategies for EAL pupils is in line with the GBG’s underlying theory of change.

This teacher also felt that the team membership element of the GBG encouraged EAL pupils to work independently. They felt that the GBG increased pupils’ responsibility, as they were accountable for their behaviour, and so they learned to work independently to avoid incurring an infraction for their team. Team
membership is considered to be one of the four core elements underpinning the GBG (Ford et al., 2014), and so the findings from the present study are consistent with this. This finding is also supported by NALDIC’s guidelines (South, 1999) which suggest that teachers need to implement strategies that encourage social-affective awareness, such as co-operation.

9.5.2. vi School-Level Behaviour

This theme was not frequently discussed by teachers, and was mainly explored in terms of reasons for implementing the GBG, as opposed to the observed benefits of the game for pupils. Teachers explained that they implemented the GBG to improve pupils’ longer-term outcomes, improve current behaviour in the school, and to encourage adaptive learning behaviours. However, teachers did not typically comment on the mechanisms through which they hoped the GBG would benefit pupils.

One teacher appeared to be concerned that the disruptive behaviour displayed in the school would be a problem for pupils’ learning and outcomes later in life, and so wanted to implement the GBG in an attempt to target this issue. Indeed, the GBG logic model (see section 4.4) details benefits for pupils exposed to the GBG in childhood into youth, adolescence and adulthood. Research into the GBG has also found that individuals exposed to the GBG in school report reduced rates of service use from mental or medical health providers, drug treatment services, the juvenile or adult justice system, and social services in adulthood (Poduska, Kellam, Wang, Hendricks Brown, et al., 2008). However, these reductions were only evident in certain groups of the sample, predominately males, and not all schools had high levels of behavioural problems; thus beneficial effects for all pupils cannot be assumed.

Conversely, some teachers hoped that the GBG would improve adaptive learning behaviours in the school, such as independence, and encourage pupils to engage in their learning. Previous GBG literature suggests that the game is effective in
improving on-task behaviour, appropriate classroom behaviours, and functional skills such as independent work (Coombes et al., 2016). It is thought that these behaviours are improved through the psychological principles underpinning the GBG, including social learning theory, social field theory, and operant conditioning (see section 4.3). However, teachers did not comment on whether the GBG was successful in improving these outcomes, and so it is unclear if the GBG was beneficial for schools with high levels of behavioural difficulties.

9.5.2.vii Deprivation (Familial and Neighbourhood)

Issues surrounding pupils experiencing poverty, and the associated proxy risk variables, were mainly explored by interviewees in terms of the reasons for implementing the GBG, as opposed to the observed benefits of implementation. These issues were also primarily raised by members of the senior leadership team, as opposed to class teachers implementing the game.

Interviewees hoped that the GBG would improve pupils’ engagement with their work and develop adaptive learning behaviours. This is unsurprising considering that the GBG logic model (see section 4.4) emphasises the intervention’s impact on functional skills such as independent work and consistent on-task behaviour, which in turn leads to improved curriculum progress and attainment in literacy and numeracy. However, teachers did not go on to identify if these changes in pupils’ behaviour actually occurred, and so the success of the GBG for pupils experiencing poverty in the present study cannot be established.

Interviewees also expressed a desire for the GBG to target issues associated with deprivation, such as parental education level, substance abuse and other issues in the wider community. Teachers typically reported hoping for improvements in social and emotional competencies, as well as pupils’ general wellbeing. Previous research suggests that pupils’ experiences of poverty can lead to anxiety, unhappiness and insecurity (Mazzoli-Smith & Todd, 2016), while pupils eligible for FSM are more likely to have worse social-behavioural outcomes, including lower
levels of self-regulation and pro-social behaviour, and higher levels of anti-social behaviour (Sammons et al., 2014). Therefore, teachers’ concerns regarding the outcomes for pupils experiencing poverty were generally consistent with previous literature in the field.

One way in which they thought the GBG may be beneficial was through the structured and rigid nature of the game. While teachers reported that this was detrimental to some pupils, such as those with an SEND or low academic achievement, some teachers commented that the strong and consistent boundaries provided by the GBG would help to overcome the issues associated with deprivation. Indeed, previous research suggests that structure can be beneficial for pupils experiencing poverty (Estyn, 2012), and the logic model (see section 4.4) details longer-term health and social outcomes similar to those described by teachers, such as increased productivity and school completion, reduced likelihood of substance abuse, lower rates of anti-social behaviours and criminality, and reduced need for access to mental health services. However, from the present two year study alone, it is not possible to determine whether these longer-term outcomes will be achieved.

9.6 Research Question 3: Implementation

3. Are there any differential effects of implementation of the GBG (specifically fidelity, quality, dosage) on behavioural and/or academic outcomes for at-risk children?

As noted previously (see section 6.6.1.iii), fidelity and quality were combined for analyses, following an EFA conducted on the observation schedule; thus, they are discussed here as one variable (FidQual).

No overall associations between high FidQual and pupils’ disruptive behaviour outcomes were found. Whilst this finding is inconsistent with the very limited number of GBG studies that typically find fidelity to be an important moderator of
outcomes (Ialongo et al., 1999; Lannie & McCurdy, 2007), studies of other interventions have found similarly null effects (Berry et al., 2015; Humphrey et al., 2017; Social and Character Development Research Consortium, 2010). Specifically regarding pupils’ risk status, there were no significant beneficial associations between implementation and outcomes for pupils at any risk level. Whilst no GBG research exists regarding the differential gains for pupils at varying levels of risk associated with implementation, research into other interventions has found that high levels of implementation are important for high-risk pupils (Abbott et al., 1998; O’Donnell, 2008).

Only one interaction was significant in the behaviour model, and was contrary to the direction hypothesised; being in a year 3 class with high levels of FidQual was found to be positively associated with disruptive behaviour scores for high-risk pupils. This means that, relative to pupils in the low-risk and low FidQual group, high-risk pupils’ behaviour worsened, with a medium effect size. This is inconsistent with the limited amount of GBG research that has explored the relationship between implementation and pupils’ outcomes. For instance, Ialongo and colleagues (1999) found that implementing the GBG with low fidelity produced similar outcomes to the control group, while Lannie and McCurdy (2007) found that on-task behaviour increased when feedback regarding fidelity was provided to teachers. However, Ialongo and colleagues utilised a lower cut-off when measuring high fidelity, and Lannie and McCurdy investigated the effects of feedback on delivery by individual teachers, and explored the effects of increased levels of fidelity as opposed to high fidelity. Furthermore, neither of these studies explored the differential effects of fidelity on pupils at different levels of risk exposure and so any potential outcomes were previously unknown.

Although, no previous GBG research has demonstrated such an effect, one study of the school-based intervention PATHS, conducted by Humphrey and colleagues (2017) did explore the effects of implementation, and found that higher levels of dosage predicted worse outcomes. This finding was incongruous with both the intervention’s theory of change and other studies of PATHS (e.g. Faria et al., 2013;
Schonfeld et al., 2015). Humphrey and colleagues speculated that methodological differences, such as the use of teacher self-report, between this study and previous research may have accounted for the apparent incongruence in results. Indeed, there are multiple ways to measure FidQual, and to treat it during analysis, and so this may be why different results were obtained in the present study. Furthermore, other factors not measured in the present study, such as participant responsiveness or reach, may have been influencing the results. For instance, the GBG is a highly structured and prescriptive intervention that does not allow teachers to interact with pupils during the game. Perhaps high-risk pupils who were in classrooms where the GBG was implemented with strict adherence to the manual found that there was not enough support available to meet their needs. Additionally, it is possible that teachers utilised the time that the GBG was played to withdraw high-risk pupils for other more targeted interventions or nurture groups (Askell-Williams, 2015). Thus while these pupils were technically ‘in’ the high FidQual classrooms, they were not necessarily present for the intervention. Indeed, in the evaluation of PATHS, participant reach was the only factor of implementation significantly associated with outcome variability across all of the analyses (Humphrey et al., 2015). Had more factors of implementation been explored in the present study, a more comprehensive picture may have been built, helping to better explain this seemingly incongruous finding.

As only half of the sample was utilised for RQ3, and pupils were categorised both by risk status and classroom implementation, sample sizes were extremely small. Indeed, only nine pupils were in the high-risk, high year 3 FidQual group for behavioural outcomes, and this number had almost halved by year 4; thus, it is highly likely that the results were skewed or spurious. The analysis for RQ3 was exploratory in nature, and results should be interpreted tentatively. However, the findings do suggest that levels of FidQual may be associated with changes in pupils’ behavioural outcomes, particularly high-risk pupils, although more research needs to be done in this area with a larger sample size, both with regards to the GBG and school-based interventions generally, to gain a clearer idea of the importance of FidQual for at-risk pupils’ outcomes.
Contrary to behavioural outcomes, high year 4 FidQual was found to be a marginally non-significant predictor of pupils’ overall reading attainment, indicating that pupils in high FidQual classes had higher reading scores by the end of the trial, relative to the low FidQual group, with a small effect size. This finding was supported by the descriptive statistics, which demonstrated that reading scores for pupils in the high FidQual group were 1.2 points higher on average than for pupils in the low FidQual group. Whilst this was not a statistically significant difference, it is nevertheless potentially important as research into the effects of GBG implementation, and the implementation of school-based interventions generally, is rarely conducted, and is even more infrequent when exploring academic outcomes such as reading. Instead, research typically focusses on behavioural outcomes (e.g. Abbott et al., 1998; Lannie & McCurdy, 2007) or social and emotional competencies (e.g. Askell-Williams et al., 2012). Thus, this finding highlights that further research needs to be conducted in this area to better understand the potential role that FidQual may play in pupils’ academic outcomes.

To the author’s knowledge, only one GBG study has examined the associations between implementation and pupils’ achievement, conducted by Ialongo and colleagues (1999) almost 20 years ago in the USA. Whilst the present study is in line with Ialongo and colleagues’ findings, in that greater fidelity was associated with higher reading scores, the applicability of Ialongo’s study could not be assumed; it was important that more recent research was conducted in this area, specifically in a UK context. Furthermore, the high/low fidelity cut-off in Ialongo and colleagues’ study was extremely low at 50%, meaning that a true measure of high fidelity may not have been obtained. It is questionable whether teachers implementing with only 50% fidelity were triggering all of the mechanisms of change considered to be crucial in the GBG logic model, or whether there were other factors influencing outcomes; indeed, 70-80% is generally recommended as the cut-off for high fidelity (Savignac & Dunbar, 2014). Thus, the present findings contribute to the evidence base in this area by providing a more current and accurate exploration of the
association between high levels of GBG implementation and pupils’ academic outcomes.

Regarding subgroups, there were no significant associations between high FidQual and pupils’ reading in any of the risk groups. Thus, while being in a class with high levels of FidQual was potentially associated with higher reading scores, there were no differential subgroup effects for pupils at varying levels of risk. Therefore it may be that high FidQual was equally important to all pupils, regardless of their risk status. While research typically suggests that fidelity and quality are important predictors of pupils’ outcomes (Askell-Williams et al., 2012; O’Donnell, 2008), differential effects for pupils at-risk have not previously been explored; thus this finding contributes to the evidence base in this area. It is interesting that while high FidQual was marginally associated with potential overall benefits to pupils’ reading outcomes (although this was only approaching significance), FidQual was only associated with high-risk pupils’ behavioural outcomes. This is particularly noteworthy as the GBG is primarily intended to improve pupils’ behaviour, regardless of their risk status, while academic outcomes are generally considered secondary. It would therefore be expected that the critical steps outlined in the GBG manual would be tailored towards behavioural outcomes, when in fact, they may be more important for attainment. Indeed, pupils’ reading scores were not associated with GBG exposure alone, but were potentially influenced only when the GBG was implemented as intended by the programme developers.

Regarding dosage, no overall associations between high dosage and pupils’ behavioural outcomes were found. There were also no statistically significant differential associations between high dosage and behaviour for pupils in any of the risk groups. However, there was a marginal non-significant trend, indicative of a potential association between lower disruptive behaviour scores and high dosage in year 3 among high-risk pupils. The descriptive statistics supported this, indicating that disruptive behaviour scores for high-risk pupils were 0.21 points lower on average relative to pupils in the low dosage classroom. Similarly to the FidQual models, this research is highly exploratory, and sample sizes were extremely small,
which may explain why some results were only approaching statistical significance. Nevertheless, these results suggest that there may be some differentials effects occurring, and that more robust research is warranted in this area.

However, based on the descriptive statistics, teachers were only implementing the GBG for approximately 25 minutes per week. As it is advised in the manual that teachers initially play 10 minute games three times a week from the outset, and increase this over the course of the year, it is likely that even teachers in the high dosage category were implementing with (presumed) suboptimal levels of dosage. Indeed, qualitative data from the UK pilot of the GBG indicated that teachers were concerned about the frequency and duration of the games required, particularly in the latter part of the year (Coombes et al., 2016). Had teachers been implementing with the intended frequency, different outcomes may have been evidenced. However, of the handful of GBG studies that have measured and reported dosage it seems that lower than intended levels of dosage are the norm (Pas et al., 2015; Tanol, 2010; Hagermoster-Sanetti & Fallon, 2011; Domitrovich et al., 2015). Unfortunately, these studies did not go on to assess the association between dosage and pupils’ outcomes; as Becker and colleagues (2013) explained, it is currently unknown whether a certain dosage is necessary to bring about student gains. However, the findings from the present study would suggest that this is potentially the case for high-risk pupils regarding behavioural outcomes.

Similarly for reading attainment, no overall associations with dosage were found, although the same issues apply regarding the suboptimal levels evidenced in the present study. Furthermore, there were no statistically significant subgroup associations between dosage and reading for at-risk pupils. While no GBG studies have investigated the associations between dosage and pupils’ outcomes, as previously noted, an evaluation of the PATHS intervention (Humphrey et al., 2017) did explore this, although a different outcome measure was utilised. They not only found that high dosage had no benefits regarding pupils’ pro-social behaviour and social-emotional skills, but that it actually had a significant detrimental effect. Humphrey and colleagues posited three key hypotheses to explain these
incongruous findings, which are also applicable to the present study. Firstly, they suggested that high dosage was at the expense of quality; in other words, teachers implemented the intervention often as opposed to well. Further research could be conducted in this area to establish if there is an association between quality and dosage. Alternatively, this may be due to some classes having large numbers of pupils with behavioural or academic difficulties, resulting in teachers implementing the GBG more often as a perceived solution to these problems. Indeed, this may be the case in the present study, where it is likely that classes had behavioural difficulties above the NA due to their voluntary participation in the trial (see section 9.8.2). Finally, it may be that the GBG was not as effective as other activities or interventions that were displaced by the more frequent implementation of the game.

It is clear that the level at which the GBG is implemented may be associated with pupils’ outcomes, and that this association may differ depending on the outcome of interest. Indeed, Durlak (2015) noted that it should not be assumed that each dimension of implementation is equally important for all possible outcomes. Pupils’ reading scores were not associated with GBG exposure alone, but potentially were when the GBG was implemented with the high levels of FidQual intended by the programme developers (although this was only approaching significance). Conversely, pupils’ overall behaviour was not associated with GBG exposure or variations in implementation at all. Meanwhile, results suggested that high levels of FidQual had no differential associations with pupils’ reading, but that high-risk pupils’ behaviour was associated with both high FidQual and potentially with high dosage (again, this was only approaching significance), although these associations were not always beneficial. Thus the results from this exploratory analysis indicate that implementation variability is important and should be evaluated when examining the outcomes of an intervention. However, unfortunately, this is rarely done, meaning that evaluations of interventions are susceptible to Type III error, resulting in the inaccurate attribution of the cause of the results (Lendrum & Humphrey, 2012); evaluating implementation data can therefore help to determine whether an intervention failed due to poor programme design, or to poor
implementation of a well-designed programme (Askell-Williams et al., 2012). Furthermore, it is evident that implementation does not have the same outcomes for the overall class as it does for pupils at certain levels of risk. Hence, it is important that further research is conducted in this area with a larger sample to establish the optimal levels of implementation in order to benefit all pupils.

9.7 Integration of Findings

The present study was a concurrent embedded mixed methods design (QUANqual); the quantitative strand was considered dominant, and comprised the larger portion of the study, whilst the qualitative strand provided a supportive, explanatory role. Integration is considered to be the hallmark of MMR, and so it is essential that the two strands of the present study be integrated, in order to access the insights that would be unavailable had a quantitative study been undertaken independently (Moseholm & Fetters, 2017). This allows for a greater understanding of the quantitative results regarding differential effects of the GBG for pupils at-risk, and furthers understanding of the elements of the GBG that may have influenced these results. The integration technique utilised in the present study is one recommended by Fetters and colleagues (2013), which involves comparing the quantitative and qualitative findings after separate analyses of the data.

9.7.1 Convergence of Findings

Results from RQ2 indicated that the GBG had no statistically significant differential subgroup effects for pupils at varying levels of risk, for either disruptive behaviour or reading attainment. The qualitative findings in RQ2a did not typically suggest that the GBG improved school functioning for at-risk pupils, with teachers reporting that certain GBG elements, such as the rigid structure, were unsuitable for them. Thus, the two datasets were aligned regarding these results. However, the datasets also diverged, with the qualitative findings suggesting that the GBG improved proximal outcomes such as confidence, social skills and independence for at-risk pupils. Thus it may be that the quantitative measures were targeting different
outcomes; had more specific measures been utilised regarding these proximal outcomes, different results may have been obtained. Furthermore, while these proximal outcomes are not directly related to disruptive behaviour or reading attainment, they could be seen to be mediating these outcomes, and so it may be that in the longer term, improvements in the proximal outcomes will evidence change in these more distal outcome variables.

However, there was a non-significant trend indicative of a beneficial effect on reading attainment for pupils exposed to the GBG in the low-risk group. The qualitative findings regarding differential effects for ‘low ability’ pupils were mixed, and so the datasets both converge and diverge regarding this result. This result is inconsistent with much of the qualitative findings, as teachers frequently suggested that certain core elements of the GBG, such as the rigid structure and lack of teacher-pupil interaction permitted, were detrimental to low ability pupils. Specifically, the lack of support available for pupils was problematic, and led to teachers having difficulty finding tasks that these pupils would be able to complete during the game. For instance, teachers commented that these pupils struggled to complete large pieces of work, independent literacy tasks, and difficult numeracy tasks during the game without any input from a member of staff, resulting in them disengaging from their learning.

However, other teachers disagreed, with some suggesting that the explicit nature of the game helped these pupils to understand the task, which would in turn better equip them to complete their work, and thus enhance their learning. This may explain why the quantitative results demonstrated a marginal improvement in reading attainment for low-risk pupils. It is possible that these inconsistent qualitative findings regarding low ability pupils were due to other potential risk variables that these pupils were exposed to. Risk variables do not typically occur in isolation (Evans et al., 2013), and ‘low ability’ can be an overarching term for a myriad of issues. Thus it may be that the low ability pupils who struggled with the GBG were those exposed to multiple risk variables; this would also support the
quantitative results regarding the null effects of the GBG for pupils at higher levels of risk.

While the qualitative analysis in RQ2a was focused on explaining the quantitative results from RQ2, the emergent qualitative findings also have applicability to the results from RQ3 regarding implementation. For instance, while GBG exposure generally was not a predictor of behavioural outcomes for at-risk pupils, exposure under high FidQual conditions in year 3 was associated with poorer disruptive behaviour scores for high-risk pupils. The qualitative findings provide some insight into this seemingly incongruous result. For instance, teachers reported that boys, who are typically at higher risk of behavioural difficulties (Department for Education & Skills, 2007; Hansen et al., 2010; Kellam et al., 1998; Storvoll & Wichstrom, 2002), failed to engage with the GBG, and instead tried to test the boundaries of the game. Other teachers commented that due to the rigidities of the game, they could not intervene with SEND pupils with behavioural problems, meaning that the game escalated situations instead of helping these pupils learn to control their behaviour. In an attempt to overcome their concerns regarding the negative effects of the game on at-risk pupils’ outcomes, teachers made a variety of adaptations; they provided at-risk pupils with additional CCR feedback, interacted with them during the game, and allowed them to have a one-to-one TA with them for support. In fact, teachers only typically reported benefits of the GBG for at-risk pupils if they made adaptations to the game. This suggests that certain core elements of the GBG outlined in the manual as critical to successful implementation were actually considered to be detrimental to at-risk pupils. This therefore explains the quantitative finding that high FidQual had a negative effect on behaviour, as the teachers in the high FidQual group were likely those who adhered to the manual and made fewer of these seemingly necessary adaptations.

Finally, there was a non-significant trend indicative of a potentially beneficial association between high dosage in year 3 and high-risk pupils’ disruptive behaviour scores. While some elements of the qualitative findings contradicted this (for instance one teacher reported that the increasing length of the games required was
unfair for the SEND pupils in their class), other findings did provide an explanation. Teachers commented that the length of the games helped SEND pupils to control impulsive behaviours, and that an extended period of time without input from staff encouraged at-risk pupils to develop their independence and take responsibility for their own behaviour. Therefore, pupils in classes with teachers who adhered to the recommended levels of dosage had more opportunities to develop these skills.

Based on the detail provided above, it appears that the use of MMR in the present study achieved its intended aim. While there are some diverges evident, the qualitative findings generally provide an additional insight into the quantitative results, offering possible mechanisms for any effects identified, and helping to explain any seemingly incongruous results.

### 9.7.2 Unique Contributions of the Qualitative Findings

While the data collected for RQ2a was intended to help explain the quantitative results from RQ2, and the a priori elements of the qualitative analysis were designed so as to complement the quantitative analysis, additional emergent themes were identified in the qualitative dataset. Although the analysis for RQ2 focused on levels of risk, regardless of the nature of the specific risk variables, the emergent findings in the qualitative dataset typically related to the effects of the GBG for pupils exposed to individual risk variables. The qualitative analysis thus provided an interesting insight into some unanticipated differential effects of the game. While GBG research has typically focused on male pupils or those with behavioural difficulties (e.g. Dolan et al., 1993; Kellam & Anthony, 1998; Petras et al., 2008), it appears that pupils exposed to other specific risk variables also respond differently to the game. For example, teachers noted differential effects for EAL pupils and those identified as having an SEND, as well as the female pupils in the class.

Furthermore, while research typically focuses on behavioural outcomes for students, the qualitative analysis identified 10 alternative, more proximal outcomes.
Although some of these outcomes are cited in the GBG logic model (see section 4.4), others (e.g. responsibility, leadership) have not previously been identified in GBG research. Finally, the qualitative findings enhance understanding of the mechanisms through which these outcomes are influenced, and the ways in which these mechanisms differ for each risk variable. These findings therefore pave the way for future quantitative GBG research regarding the differential effects of the game.

9.8 Limitations

Whilst the present study attempted to address several criticisms of previous research in the fields of prevention and implementation science (e.g. through the use of a larger sample size, employing mixed methods, and measuring multiple dimensions of implementation), a variety of limitations still exist. Although no psychological or educational study can be without flaws, it is important that these limitations are addressed, in order to allow the results to be interpreted more reliably.

9.8.1 Methodology

As discussed in section 6.4, while cluster-RCTs are considered by many researchers to be the “gold standard” of research design, (Maughan, 2013), they are not without their criticisms. It is argued that RCTs understate other data sources and factors such as context, and ignore or over-simplify the complex underlying processes taking place (Morrison, 2001). In an attempt to address these concerns, the present study adopted a mixed methods design, utilising qualitative data collection techniques to further explore the effects of other factors and processes that influenced pupils’ outcomes. However, there are also arguments that quantitative and qualitative methods should not be mixed, as this leads to the introduction of potential treatment bias and an unequal balance of evidence (Creswell, 2009; Creswell & Plano Clark, 2007). Indeed, the use of a concurrent embedded design dictates that one methodology should be given less priority. Therefore, while the qualitative data may have helped to explain the quantitative
results, it may mean that some potentially important findings in the qualitative data were lost.

9.8.2 Sample

While all state schools in the desired regions were targeted during the recruitment phase of the trial, the majority of the schools participating in the RCT were situated in Greater Manchester, a densely populated region known for its ethnic diversity and socio-economic issues (CLEs, 2012; Jivraj, 2013). Indeed, schools were typically larger than average schools, with higher rates of pupils with an SEND, eligible for FSM, and speaking EAL (DfE, 2017c). Furthermore, schools were invited to participate in the trial, and so there is also a potential bias as to the types of schools that chose to take part. The types of schools interested in participating in this study were likely those where there was a greater perceived need for an intervention targeting behaviour, whilst uninterested schools had no behavioural issues or were already implementing effective behaviour management strategies. As such, the sample of schools participating in the trial may not have been representative of those in the UK overall.

This raises an additional concern regarding the schools allocated to the UP condition; while these schools also perceived a need for a behaviour management intervention, they did not receive the GBG, which may have led to compensatory rivalry. Unlike other fields that utilise RCTs, all variables cannot be controlled for in the UP condition, and so it is possible that schools not implementing the GBG may have adopted other interventions over the course of the trial to address this need. This could mean that while the GBG was potentially having a beneficial effect on pupils’ disruptive behaviour and reading attainment, outcomes were also improving in the UP condition, resulting in no significant difference in outcomes at the end of the trial. If this were the case, this could have resulted in a type II error, falsely indicating that the GBG was ineffective. However, it is likely that all schools will implement some form of behaviour management strategy, regardless of their
participation in RCTs, and so this is not an issue relating to the design of the current study.

Finally, the present study experienced significant attrition at the pupil-level, with 12.6% leaving the school over the course of the trial, and either behaviour or reading data missing for an additional 11.6%. That said, missing data resulting from this issue was investigated, found to be MAR and subsequently accounted for through the use of MI, and so should not be a major concern. However, almost one quarter of schools (23% of classes) in the intervention arm had ceased implementation by the end of the trial. This is particularly pertinent considering that this trial was conducted under efficacious conditions, with intervention costs subsidised and developer support provided for the delivery team. As data were still collected from all of these schools, and ITT analyses were conducted, it is therefore possible that the lack of effect found was due to implementation failure, and thus results were not truly representative of optimal GBG delivery.

9.8.3 Measures

9.8.3.i Disruptive Behaviour (TOCA-C)

Another potential limitation of the present study is the use of the TOCA-C as a measure of pupils’ disruptive behaviour problems. This measure was originally designed by a developer of the GBG, and has primarily been used in GBG studies. Thus, while it may be that this measure was well suited to the present study, there is also a concern that the TOCA may be considered inherent to treatment, whereby it is tailored to best measure the aspects of behaviour targeted by the GBG, rather than disruptive behaviour more generally, potentially biasing results. However, as this issue would have favoured outcomes in the intervention arm of the trial, and no effects were found for the GBG, it is unlikely that this was an issue. Additionally, while previous psychometric validation has found the measure to be sound (Bradshaw et al., 2015; Koth et al., 2009), studies exploring this are limited and, again, are most frequently conducted by developers of either the TOCA or the GBG
(e.g. Bradshaw et al., 2015; Chan et al., 2012; Kellam et al., 1994). Furthermore, the TOCA-C has not previously been validated in a UK context, and so the utility of the measure in the present study is unknown.

In addition, the TOCA-C is a teacher response scale which can also lead to several issues. The developers of the measure advise that a variety of factors, such as the demographic characteristics of the child and their grade level, can influence teachers’ reports of behaviour problems. Furthermore, it is thought that the timing of administration can also influence reports, with a notable difference in reports between the beginning and end of the school year (Koth et al., 2009). Specifically, significantly higher reports of behaviour problems have been found for boys in spring as opposed to autumn when utilising the TOCA-R (Dolan et al., 1993). As the present study relied on data collected utilising the TOCA at similar time points, this may have influenced whether a significant effect was found for the GBG. However, as this issue would have occurred in both the intervention and UP arms of the trial, results should not have been affected.

It is also acknowledged that no single informant can provide a comprehensive picture of a pupil’s behaviour, and so collecting similar data from other informants such as the pupil or parents could potentially have provided a more valid disruptive behaviour score (De Los Reyes et al., 2015). Indeed, inter-rater correlations between teacher- and self-report behaviour measures have previously been found to be weak (Goodman, Meltzer, & Bailey, 1998); discrepancies in scores are thought to be due to biases, access to different information, and psychometric differences across informants (Olino & Klein, 2015). It is also argued that there are divergences between direct observational measures of pupil behaviour and informant-report measures, with observations considered to be more robust (Merrell, 2001). However, the additional financial and human resources required, as well as the increased data burden for teachers, meant that only one form of informant-report measure was utilised for pupils’ disruptive behaviour scores.
Finally, Greenberg and Abenavoli (2017) argue that it is critical to consider characteristics of the population and the outcome of interest when deciding which measure to use, otherwise the effects of an intervention may be underestimated. In the present study, the majority of the pupils in the trial did not have disruptive behaviour problems at baseline, and so average scores prior to implementation were already low. Thus there was not a lot of scope on the scale for pupils’ behaviour scores to fall over the course of the trial, meaning that any effects may have been lost. However, the exploratory subgroup analysis in the present study meant that the outcomes for pupils with more extreme behaviour scores were also examined, and so this issue has been partially addressed.

9.8.3.ii Reading Attainment (HGRT and KS1 Data)

Similarly to limitations surrounding the TOCA, whilst the HGRT was validated by the developers utilising a large sample of pupils (Vincent & Crumpler, 2007), the validity of the test has not been more widely explored by independent researchers. Furthermore, the HGRT is not designed for large-scale research use and so its utility in the present study is not known.

The HGRT was chosen as the outcome measure for the current study as previous studies have found it to produce scores that correlate with NC reading test scores (ibid), the measure utilised to assess baseline reading attainment. However, these studies were conducted over ten years ago in 2006; as KS1 national tests were overhauled in 2016, it is unclear whether the HGRT and national tests still produce comparable scores. If the concurrent validity of the HGRT is no longer as high, this may affect whether pupils’ reading attainment appeared to change over the course of the trial, potentially influencing the interpretation of the results. Additionally, teachers spend time preparing their pupils for national tests, and so pupils have experience of the types of questions that they will be asked, which may bias their scores. Conversely, the majority of pupils had no prior experience of the HGRT; this could mean that reading scores were inflated at baseline, and thus that there appeared to be no significant improvement in attainment over the course of the
trial. However, as this effect would have occurred in both the intervention and UP arms of the trial, this should not have had a detrimental effect on the results.

9.8.3.iii Implementation (Scoreboard and Observation Schedule)

Two measures were utilised for assessing teacher implementation of the GBG, an online scoreboard (dosage) and an observation schedule (FidQual). Whilst the online scoreboard meant that detailed dosage data were collected with no additional burden to the teacher, accurate data depended on teachers using the scoreboard during every game, and filling in the details accurately. However, there was no real way to monitor this, and dosage data were not triangulated in any way.

Regarding the observation schedule, this was completed by a researcher from the evaluation team, to reduce the bias associated with self-report measures (Domitrovich et al., 2010; Lillehoj et al., 2004). Although observations are considered to be the most valid method for assessing implementation (Humphrey, Lendrum, et al., 2016) and there were no concerns regarding IRR or researcher effects, only one observation was conducted for each teacher. This therefore only provides a single snapshot of implementation, and does not account for contextual factors that may have been influencing implementation on the day of the observation. Repeated observations over multiple time points would have been desirable to provide a more representative average rating (ibid), had the constraints of the trial allowed this. Additionally, whilst every effort was made to reassure the teachers that the research team were not monitoring fidelity, social desirability may have influenced the way that teachers implemented the GBG when they were being observed, thus not providing a representative picture of day-to-day delivery.

9.8.3.iv Risk

The majority of the data on pupils’ risk variables came from the NPD, although some data were also collected from the surveys completed by teachers in the
schools at baseline. One drawback of the measures of risk used in the present study is that the data were only collected from a single time point following a cross-sectional design. There is evidence to suggest that specific risk variables can influence outcomes at different ages, and that sustained exposure to risk variables can have an additional deleterious effect (Stouthamer-Loeber et al., 2002); however, this was not accounted for in the present study. Furthermore, whilst there are an abundance of risk variables that can influence outcomes for children, in various ecological domains, only a limited number of risk variables and domains were explored. Whilst it would not have been feasible to explore all possible risk variables, and the present study was dependent upon the data available from the wider trial, it is important that other factors including biological influences such as testosterone levels, parental and familial issues such as maternal mental health and parenting style, and the influence of peer groups (Deater-Deckard et al., 1998; Lösel & Farrington, 2012) are acknowledged here.

Several risk variables were also measured through the use of a proxy variable; for instance, familial deprivation was measured utilising FSM status. Whilst this is a technique commonly used in risk research, FSM status is only one way of measuring an aspect of deprivation; it does not provide additional context, demonstrate the level of deprivation, or indicate whether deprivation has been sustained over time. It is also dependent on parents applying for FSM for their children (Hobbs & Vignoles, 2010). Whilst there may not be one single complete measure for these risk variables, had the scope of the study allowed, multiple aspects could have been explored concurrently, to provide a more accurate representation of pupils’ risk status.

9.8.3.v Interview Schedules

A subset of the questions from the interview schedules designed for the wider trial were utilised for the qualitative strand of the current study. Thus, there were few questions focusing on differential effects of the GBG, specifically for pupils with varying levels of risk, and so the extent to which this topic could be explored was
limited. Furthermore, as the interview schedule was semi-structured, and different interviewers were utilised across the case study schools, it is possible that not all interviewees were asked questions focusing on differential effects. Additionally, as noted in section 9.7.1, issues surrounding fidelity and dosage emerged in the qualitative analysis for RQ2a, with some teachers describing adaptations that they had made to the game to make the intervention suitable for at-risk pupils in their class. However, as the wider trial was not focused on the associations between implementation variability and outcomes for at-risk pupils, rich data surrounding this issue was not obtained. Thus, the ways in which implementation influenced at-risk pupils’ outcomes could not be fully ascertained.

It is possible that questions regarding differential effects of the GBG led to social desirability bias, or that teachers displayed demand characteristics. Teachers may have felt that the GBG should have been producing differential effects, or that this was what the interviewer was expecting, and so provided responses in accordance with this. However, efforts were made to ensure that the interviewees were aware that the research team were independent evaluators of the intervention. Questions were also phrased carefully so as to avoid leading the participant, and prompts and probes were determined in advance to reduce bias.

9.8.4 Analytic Strategy

9.8.4.i Multi-Level Models

MLMs were the analytic strategy chosen for the present study, as they account for clustering of pupils within schools and allow for variance at both the pupil- and school-level (Twisk, 2006). However, in the present study, no classroom-level was included in the analysis for the first two research questions, and no school-level was included in RQ3 analysis. Therefore, for RQ1 and RQ2, variance that appeared to be explained at the school-level may have actually been accounted for at the classroom-level, and variance at the classroom-level in RQ3 may have been accounted for by differences between schools. This is particularly pertinent for
pupils in multi-form entry schools, where not all pupils in a year group are in the same class or have the same teacher.

However, whilst fitting two-level models to three-level data can lead to the misattribution of response variation (Leckie, 2013), including three levels in the model was beyond the scope of the present study, and the interpretation of results would have been limited by power and sample size. MI procedures would also not have been possible with a three-level model in REALCOM, meaning that the amount of missing data would have been an issue. Furthermore, the levels utilised directly related to outcomes of interest for a given analysis. For example, school was utilised as the second level for the GBG impact analysis, as schools were the level at which randomisation occurred, whilst the exploratory implementation analysis utilised the class as the second level as this was where implementation occurred.

9.8.4.ii Treatment of Predictor Variables

Specifically regarding analysis for RQ2 and RQ3, the ways in which pupils were grouped into risk categories may have been problematic. Pupils were allocated to one of three risk groups for behaviour, and one of four for reading, and these were based on calculations of effect size between risk levels. However, this meant that sample sizes were uneven between the groups, and that the higher risk groups had smaller sample sizes. This was particularly pertinent for RQ3 analysis, where the overall sample size was half that of the analysis for RQ1 and RQ2. It is therefore possible that skewed or spurious results may have been found in the subgroup analysis. Indeed, this may explain why differences emerged between the complete case and multiply imputed models. Thus, results from some of the subgroup analyses should be interpreted with caution, and the exploratory nature of RQ3 in particular should be acknowledged.

For RQ3, FidQual and dosage were examined to establish whether variation in implementation was associated with at-risk pupils’ outcomes. However, it is recommended in the literature that all eight aspects of implementation should be
included in analyses, in order to gain a more complete picture (Durlak, 2015; Durlak & DuPre, 2008; Humphrey, Lendrum, et al., 2016). While this was beyond the scope of the present study, including all aspects of implementation may have helped to explain some of the seemingly incongruous findings. Furthermore, although FidQual and dosage were continuous variables, they were re-coded into high and low categories for the purposes of the present study, using an external cut-off of 80% for FidQual and the 75th percentile as a cut-off for dosage. While the external 80% cut-off for FidQual is often recommended (Savignac & Dunbar, 2014), this does cause some information to be lost, and can lead to uneven sample sizes, influencing the reliability of the results.

Regarding dosage, as this evolves over the course of implementation, there was no external cut-off or accepted level that could have been used in the present study, and so the top 25% of the sample were categorised as high dosage. However, based on the descriptive statistics, it is evident that dosage was not increasing over time, suggesting that the level required by the programme developers was not reached. Therefore, while the high dosage teachers in the present study may have been implementing more frequently relative to the low dosage teachers, their dosage levels may still have been sub-optimal, and thus not representative of the levels of implementation required in the GBG manual. This brings into question whether the categorisation of dosage was truly representative, and hence means that the results may not provide an accurate indication of the effects of optimal dosage.

9.8.4.iii Thematic Analysis

Some qualitative researchers argue that thematic analysis is a generic skill and is not a specific approach in its own right. It is thought that thematic analysis lacks rigour due to the absence of clear and concise guidelines for conducting it (Braun & Clarke, 2006; Ryan & Bernard, 2000). However, in an attempt to overcome this issue, Braun and Clarke’s (ibid) six-phase guide for thematic analysis was utilised in the present study, to ensure that the analysis was undertaken in a way that was theoretically and methodologically sound.
Other methods of qualitative analysis do exist, and could have been utilised in the present study to overcome some of the concerns of utilising thematic analysis as a method in its own right. For instance, interpretative phenomenological analysis (IPA) is often utilised to gather the experiences of specific individuals dealing with specific situations or events in their life (e.g. exposure to risk variables) (Larkin, Watts, & Clifton, 2006). However, IPA seeks to uncover how an individual constructs their situation, as opposed to uncovering information about the situation itself (ibid); this would have limited the extent to which information regarding the perceived benefits of the GBG to pupils exposed to various risk variables could have been established. Therefore, thematic analysis was the most suitable approach for the present study.

9.8.5 Conceptual Limitations: Cumulative Risk Theory

One of the criticisms of cumulative risk theory is that the required treatment of the risk variables means that potentially important information can be lost (Evans et al., 2013). When identifying the risk variables associated with disruptive behaviour and reading attainment, all categorical predictor variables, with the exception of ethnicity and EAL category, were treated as binary variables. This means that more nuanced effects may not have been identified. For instance, SEND status was treated as a yes/no dichotomy, and encompassed all levels of SEND support, and all SEND categories. Whilst SEND status was found to be a significant predictor, in reality it may be that only certain types of SEND category were risk variables, or that these differed between outcome variables.

Furthermore, while continuous predictor variables were left unchanged for RQ1 analysis, they were converted to binary variables for future analyses, as is consistent with cumulative risk theory (Gerard & Buehler, 2004b; Hebron et al., 2016). To achieve this, pupils in the top 25% were categorised as at-risk, while the remaining 75% were treated as not at-risk. While this is consistent with much of cumulative risk research, it does have its limitations. For example, pupils were only
deemed to be at-risk relative to the rest of the sample, and it is unknown whether the scores in the sample were representative of the wider population. Therefore, pupils may have been falsely identified as at-risk or not at-risk when comparing their scores to the wider population.

There is also a loss of information on the intensity of the risk variables due to the dichotomisation of continuous variables. Information about the degree of risk exposure is lost, resulting in weaker predictive power and resulting in less sensitive estimates of covariation. Further criticisms include the lack of attention to temporal parameters such as pupils’ age at exposure to risk, and the duration of exposure; this can influence the likelihood of exposure to other risk variables and the impact of these variables on pupils’ outcomes (Evans et al., 2013).

9.9 Implications and Directions for Future Research

The present study utilised data from the first national large-scale trial of the GBG, and was largely exploratory in nature, and so while the results have implications for schools and intervention developers regarding the risk variables that pupils are exposed to, and the differential effects of interventions for at-risk pupils, they also highlight several gaps in the literature that require further investigation. These are discussed in more detail below.

9.9.1 Risk Variables and Cumulative Risk Exposure

9.9.1.i Risk Variables

All of the significant variables in RQ1 were either fixed risk factors, in that they cannot be changed, or were malleable risk markers, but it is beyond schools’ capabilities to reduce exposure to them (e.g. FSM eligibility). Therefore, schools could implement interventions to address the influences of these risk variables for these groups of pupils, rather than targeting the risk variables themselves. Conversely, schools could implement interventions that aim to develop protective
and promotive factors that enhance resilience, in order to provide pupils with the coping strategies necessary to overcome the adversity associated with the risk variables to which they are exposed (Brooks, 2006). Strategies could also be put in place by the school specifically to support pupils’ reading attainment, such as the provision of specialist resources, one-to-one support, or additional small-group literacy sessions (Gottfried, 2014). The use of effective universal behaviour management strategies could also be utilised to target both school-wide and individual behaviour problems.

Some risk variables, such as deprivation and CLA status, can have a host of proxy risk variables associated with them. Therefore, instead of targeting the risk variables themselves, schools could attempt to address the effects of the proxy variables, and the mediating mechanisms through which these effects occur (e.g. lack of stability, abuse and neglect, disrupted attachment to primary caregivers) (Kraemer et al., 2001; Leyro et al., 2010; Wright et al., 2013). For instance, schools can implement programmes such as nurture groups that can provide these pupils with an opportunity to learn early nurturing experiences and equip them with the social and emotional skills they need to function effectively in school (Nurture Group Network, 2017). They can also increase their engagement with the community by providing additional provision for parents, such as parenting classes, on-site counsellors and family support workers to target the proxy variables associated with deprivation (Sharples et al., 2011; Webb et al., 2012).

Regarding future research, there is scope to explore the risk variables identified in the present study in greater detail. For example, some risk variables, such as SEND status and CLA status, were treated in the broadest sense, meaning that more nuanced information may have been lost. For example, SEND encompasses a wide range of needs (DfE, 2015c), and so it may be that not all pupils with an SEND will be as likely to experience the influences of certain risk variables. Furthermore, some risk variables such as familial deprivation were measured utilising a single proxy variable, thus resulting in only a small amount of information being obtained on a complex issue (Hobbs & Vignoles, 2010). Therefore, a more detailed analysis,
exploring categories within risk variables and other factors associated with them, could be conducted in the future. This would help to establish whether all pupils in a group are equally at-risk, and would provide a more accurate representation of pupils’ risk status. Further research could also pay closer attention to the relative strength of individual risk variables, to identify the variables that need to be targeted; this could be done by standardising the outcome variables of interest to allow for meaningful comparisons between risk variables, as was done in the present study.

Additionally, some risk variables have the capacity to change over time (Kraemer, 1997; Leyro et al., 2010), and there is evidence to suggest that their detrimental effects can be exacerbated following prolonged exposure (Stouthamer-Loeber et al., 2002). However, only a single time-point was utilised in the present study, and so further research could be conducted to establish if these risk variables have greater influence for certain age groups, whether the age of onset can influence their effects, and whether the length of exposure affects pupils’ outcomes. This would further understanding of the ways in which risk variables influence pupils’ outcomes, in order to help to tailor interventions in the future. There is also scope for research to explore additional potential risk variables for pupils’ school functioning. A small number of variables were examined in the present study, in only a few ecological domains. However, Bronfenbrenner (1986) argued that to fully understand an individual’s behaviour, multiple ecological domains need to be explored. Thus, more risk variables in a wider variety of ecological domains could be examined in future research. There is evidence to suggest that variables in a variety of domains can influence pupils’ outcomes, including in the neighbourhood, such as crime levels (Greenberg, Domitrovich, & Bumbarger, 2000), and familial variables, such as parental education and mental health (Deater-Deckard et al., 1998), and so their associations with pupils’ school functioning could be explored. Furthermore, biological variables within the child, such as hormonal factors (ibid; Lösel & Farrington, 2012), could also be incorporated into future educational research.
Relative to pupil-level variables, school-level risk variables have been explored infrequently in the previous literature. However, the present study indicates that school-level variables have a smaller but still statistically significant effect on pupils’ behavioural and academic outcomes. Thus, it is important that these risk variables are identified, so schools can implement the appropriate strategies to negate their effects where possible. Therefore, more research is warranted in this area, to establish any other school-level risk variables that may influence pupils’ outcomes. Similarly to pupil-level risk variables, any varying effects at different time-points also need to be established.

9.9.1.ii Cumulative Risk

Findings from the present study indicated that there was a cumulative risk effect for both disruptive behaviour and reading attainment; with the likelihood of negative outcomes increasing as the number of risk variables accumulated, irrespective of the individual risks. Thus, consistent with previous literature this suggests that risk variables are not independent and can interact with each other (Gerard & Buehler, 1999). Furthermore, the risk-outcome relationships for both the disruptive behaviour and reading attainment outcomes were non-linear, indicating a disproportionately negative increase in outcomes as the number of risk variables increased. The finding that the number of risk variables that a child is exposed to is a better predictor of outcomes than the nature of the specific risk variables also supported this.

This has implications for the future of risk research; by continuing to conduct only single risk research, studies could fail to account for the clustering of risk variables and hence produce biased estimates regarding the effects of an isolated risk variable (Evans et al., 2013). Future studies need to acknowledge the unique and complex profile of individuals when conducting risk research. Risk variables do not occur in isolation, but interact with each other to have an additional deleterious effect (Appleyard et al., 2005); thus research should focus on exploring multiple risk variables across different ecological domains, and the interactions between them,
when investigating children’s outcomes. Indeed, as Flouri and Kallis (2007) suggested, only investigating exposure to one extreme risk variable when assessing the prevalence of high-risk youths can bias estimates, and results in those pupils at higher risk due to exposure to multiple medium-level risks being neglected. However, the risk variables that are most likely to occur in conjunction are unknown, as are the ways in which they interact with each other. Indeed, it is likely that severely negative outcomes will have multiple causal chains (Kraemer et al., 2001). Therefore, future research may benefit from exploring the risk variables that frequently occur together, and examining the multiple underlying interactions taking place between them.

Kraemer and colleagues (ibid) proposed that future risk research needs to consider proxy and overlapping risk variables, as well as the mediating and moderating mechanisms occurring between them. They argued that this was necessary in order to be able to accurately identify the high-risk individuals in need of interventions, and to ensure these interventions are effective; when there are chains of causal risk variables mediating outcomes, addressing only one aspect of that chain will result in negligible benefits to the individual. Hebron and colleagues (2016) also suggested that future research should consider including variables not identified as individual risk variables in a pupil’s cumulative risk score, to ascertain if there are previously unknown interactions occurring between variables that are increasing pupils’ risk levels. This would provide further support for the superior predictive power of cumulative risk, and would mean that those pupils at the highest levels of risk would be more accurately identified.

The findings regarding cumulative risk also have implications for the types of interventions that are utilised in schools when attempting to target concerns surrounding behaviour and attainment. Similarly to the issues regarding risk research, interventions need to target multiple risk variables across various domains (Oldfield et al., 2015), in an attempt to address as many of the risk variables as possible. Based on the findings from the present study, it appears that the specific risk variables that these interventions target is not important; indeed, in
order to lower a pupil’s risk level, interventions should aim to reduce all the risk variables that they can. Interventions also need to take account of the different types of risk variable, so as to target the effects of any fixed risk variables, whilst intervening to modify those that can be changed (Kraemer, 1997; Leyro et al., 2010). One solution to this may be to utilise universal interventions that have a strong logic model and theory of change, targeting a range of both proximal and distal factors associated with the outcome variable of interest.

Additionally, schools may need to provide targeted interventions for pupils at high levels of risk. Checklists of risk variables across multiple domains could be used to help identify high-risk pupils (e.g. Shepler, 2009). Schools could then implement more tailored interventions with these pupils in an attempt to reduce their risk level. Again, it would be impractical to target each and all of the specific risk variables that a pupil was exposed to, but instead these interventions could attempt to reduce the overall number of risk variables as far as they can. Furthermore, an increased awareness of the risk variables could aid earlier identification of the pupils more likely to experience negative outcomes; this means that they could be targeted before problems develop, as opposed to conducting the more difficult task of addressing them once the issues are already evident (Greenberg et al., 2005).

Conversely, instead of attempting to reduce the number of risk variables, or address their resultant effects, schools could attempt to develop promotive and protective factors; these enhance resilience, and subsequently buffer the effects of exposure to risk variables (Powers, 2010; Stouthamer-Loeber et al., 2002). The findings from the present study support suggestions in the literature that after exposure to a certain number of risk variables, pupils’ coping strategies become overwhelmed and subsequently exhausted, resulting in an exponential increase in difficulties (Evans, 2003; Gerard & Buehler, 2004b; Oldfield et al., 2015). By enhancing pupils’ resilience, they will be able to handle exposure to more risk variables, without their risk level having a negative effect on their outcomes. Indeed, Morales and Guerra (2006) emphasise the importance of being able to incorporate
training in coping strategies into interventions for pupils, particularly those at the highest levels of risk. A variety of strategies exist in the literature regarding the ways in which schools can enhance resilience, such as through the provision of a safe and nurturing school environment, stability, supportive relationships with teachers and peers, high expectations of pupils, and opportunities for pupils to engage in positive activities (Alvord & Grados, 2005; Corcoran & Nichols-Casebolt, 2004; DfE, 2016b). Additionally, both universal and targeted promotive interventions could be used.

However, in order to do this effectively, more research needs to be conducted into the promotive and protective factors, particularly those that can be developed in schools, that have beneficial effects for pupils at varying levels of cumulative risk exposure, regardless of the nature of the specific risk variables. The mechanisms through which these factors interact with risk variables at various levels also need to be further examined, to enhance understanding of the best ways to reduce the negative effects of risk exposure, and to help pupils cope with exposure to multiple risk (Oldfield et al., 2015). It is possible that promotive factors operate in the same was as risk variables, in that they are unlikely to occur in isolation and can interact with one another (Stoddard et al., 2013; Ostaszewski & Zimmerman, 2006). Indeed, Ostaszewski and Zimmerman (ibid) suggested that enhancing individual promotive factors may not be sufficient to achieve a successful outcome when facing certain constellations of risk; thus, further research is required to identify if there is a cumulative promotion effect, in order to utilise this successfully when targeting at-risk pupils.

Finally, Evans and colleagues (2013) argue that one of the primary limitations of cumulative risk theory is the lack of a theoretical explanation for its effects. While three rationales exist (ibid), little research has been conducted to establish which, or any, of these provides support for its predictive power. An understanding of the mechanisms through which cumulative risk influences pupils’ outcomes may aid in the development of effective interventions or strategies to reduce the harmful effects of cumulative risk exposure. One compelling rationale is allostatic load, a
model of chronic stress that has gained increasing attention in recent years (Evans & Kim, 2012). However, research into this phenomenon from an educational perspective is lacking; more inter-disciplinary studies are needed, to examine the effects of cumulative risk in educational outcomes from a biological and health perspective. Once the underlying mechanisms responsible for the superior predictive power of cumulative risk are identified, more may be able to be done to prevent the disproportionately detrimental effects on pupils’ school functioning.

9.9.2 The GBG and the Implementation of School-Based Interventions

9.9.2.1 Cultural Transferability

Results pertaining to RQ2 and RQ3 indicated that there was no overall effect of the GBG on pupils’ disruptive behaviour or reading attainment, nor were there any significant beneficial associations between different levels of implementation and pupils’ school functioning. There were also no subgroup effects for pupils at varying levels of risk. These results therefore suggest that the GBG is not beneficial for UK schools seeking an effective behaviour management strategy for pupils at-risk, regardless of the way it is implemented. Although the GBG has a strong evidence base spanning several decades in a number of countries around the world (see section 4.5 for an overview), it was not found to be effective in this UK study; thus, the cultural transferability of this intervention cannot be assumed. Lendrum and Wigelsworth (2013) raised concerns regarding the success of interventions that have been exported to countries with different education systems, cultural beliefs and expectations, and the findings from the present study do indeed suggest that this may be a problem with the GBG. This also has implications for the implementation of imported interventions more broadly in schools; it highlights the importance of evaluating imported interventions to assess their cultural transferability before implementing them to scale, and schools should be wary of implementing imported interventions that have not been tested in this way. Thus, future research should seek to examine the cultural transferability of both the GBG
and other universal interventions before importing and implementing them in schools, in order to ensure that they are able to attain their intended outcomes.

While this originally American intervention has been found to be effective in other countries, it was heavily adapted prior to implementation to suit the school culture in the Netherlands (Huizink et al., 2009; Spilt et al., 2013; van Lier et al., 2004), France (Dion et al., 2011) and Spain (Ruiz-Olivares, Pino, & Herruzo, 2010) (see section 4.5 for an overview). This may explain why the cultural transferability of the intervention has not previously been found to be a concern when evaluating its success. However, in the UK the GBG was implemented in its original format, and indeed, the qualitative findings indicate that cultural transferability may have been an issue for teachers. Several teachers reported that they felt the need to adapt the game to make it suitable for their class. However, those who did make these adaptations often reported that the game was beneficial for at-risk pupils; this has implications for the intervention developers, as it may be that the GBG needs to be altered before it is suitable for implementation in the UK.

Lendrum and Wigelsworth (2013) noted in their review of future directions for the evaluation of school-based interventions that further research is needed to explore the transferability of programmes implemented outside of their country of origin, in order to establish if any adaptations are required to achieve the expected outcomes. It is thought that local adaptations that enhance ownership or support ‘goodness-of-fit’ can be beneficial to the successful implementation of interventions (ibid), and research suggests that a major factor in the successful transferability of interventions is their adaptability (Castro, Barrera, & Martinez, 2004). While adaptations are not currently permitted in the GBG manual, it appears that they may be necessary prior to importation; thus future research needs to identify the appropriate adaptations that are in line with an intervention’s theory of change before importation, to ensure that the intervention is suitable for the context, while the components considered to be critical are still in place (Sharples, Albers, & Fraser, 2018). One way to achieve this may be to conduct further qualitative research, to establish the types of adaptations perceived to be necessary by UK teachers, and
the components of the intervention that are considered acceptable to the local context. Intervention developers can then tailor the manual accordingly, or provide guidance on acceptable adaptations, in order to increase the likelihood of achieving successful outcomes.

9.9.2.ii Differential Effects for At-Risk Pupils

Regarding reading attainment specifically, there was a marginal, non-significant trend indicative of improved reading scores for pupils in the low-risk group. Thus, while the GBG was potentially having an impact on attainment for pupils in the low-risk group, the magnitude of this change was not practically meaningful. However, considering the effects of the GBG on pupils’ attainment are considered to be more medium-term (see logic model – section 4.4), it is possible that they may strengthen. Indeed, sleeper effects were noted in Kellam and colleagues’ (1994) study of the GBG at a six-year follow-up. Therefore, a longer-term follow-up is warranted, to establish if these effects are amplified over time, thus making the GBG a valuable intervention to schools.

Additionally, it is possible that the measures chosen for the present study influenced the academic effects identified. More research could be conducted regarding other aspects of Literacy, or could be further extended to explore the effects on other academic outcomes such as Mathematics. Furthermore, there was an effect on reading attainment (albeit marginal) when there was no effect on disruptive behaviour, even though academic outcomes are considered to be secondary to behavioural ones (Coombes et al., 2016). It is possible there were improvements in behaviours pertinent to improved attainment that were not measured in the present study, which may explain why these academic outcomes did not follow on from improved behaviour scores. Had an alternative measure of behaviour been utilised, different results may have been obtained. Indeed, the GBG logic model (see section 4.4) indicates that academic attainment is improved through the development of adaptive learning behaviours rather than the reduction of disruptive behaviour. Future research should therefore seek to establish the
effects of the GBG on other types of behaviour, in order to ascertain the pathways through which pupils’ outcomes are influenced.

Previous studies that have found positive results for academic outcomes have often combined the GBG with another academic curriculum (e.g. Bradshaw et al., 2009). If the GBG is indeed successful at improving pupils’ adaptive learning behaviours, it may be that while the GBG intervention alone is not strong enough to influence pupils’ attainment, implementing it alongside another intervention may enhance the benefits to pupils, by targeting both the general learning behaviours and specific reading skills. This may be a useful strategy for schools to utilise when implementing interventions to improve pupils’ academic outcomes. Thus, future research needs to explore the types of intervention that can be successfully combined with the GBG, and the combinations that will have the greatest benefits to pupils’ academic outcomes.

9.9.2.iii The Fidelity-Adaptation Debate for At-Risk Pupils

The finding that the GBG had no statistically significant beneficial effects for pupils at any risk level is incongruous with the qualitative findings from RQ2a, which suggested that teachers felt that the GBG could be beneficial for at-risk pupils when adaptations were made. However, based on the qualitative findings, teachers typically made these adaptations for pupils on a case-by-case basis, and so the qualitative results are not generalisable to all at-risk pupils. It may be that universal interventions are not intensive enough to benefit at-risk pupils (Muthén et al., 2002). Greenberg and colleagues (2001) argued that universal interventions do not provide a sufficient duration or intensity to alter the developmental pathways of pupils already at significant risk. Thus schools may need to focus their efforts on providing additional targeted interventions with a more individualised approach for these pupils. Intervention developers may also need to address some of the concerns raised by the teachers in the present study that led to them making adaptations, such as the problems associated with the lack of teacher-pupil interaction, the reduced support available to at-risk pupils during the game, and the
levels of dosage required. Similarly to the Dutch studies of the GBG (e.g. Huizink et al., 2009; Spilt et al., 2013; van Lier et al., 2004), the game may need to be reformulated before it is can have a significant beneficial effect for at-risk pupils in the UK.

Whilst the qualitative analysis did not set out to identify the ways in which implementation influenced at-risk pupils’ outcomes, the importance of adaptations was a key emergent finding in RQ2a. The findings suggested that teachers felt different adaptations were necessary for different pupils, depending on the risk variables that they were exposed to, and their individual needs. It therefore may be that the GBG needs to provide more flexibility to teachers, and allow them to use their professional judgement to decide which adaptations are necessary, and for which pupils. This could be facilitated by the coaches that are provided to schools implementing the GBG; with their knowledge of the underlying mechanisms triggering change and the steps in the manual that are considered to be critical to this, they could work more closely with teachers to develop the game so that it is feasible to implement in their classrooms. As well as increasing the acceptability of the intervention, which may in turn increase the likelihood of sustained implementation (Wehby et al., 2011), it is also more likely to be beneficial to the at-risk pupils most in need of a programme that addresses school functioning.

The results pertaining to RQ3 appear to support the qualitative finding that the core elements of the GBG were not beneficial for high-risk pupils. Results indicated that while there was a marginal, non-significant trend indicative of an association between high FidQual in year 4 and higher overall reading scores, high FidQual in year 3 was in fact associated with higher disruptive behaviour scores for high-risk pupils. This means that their behaviour worsened, relative to those pupils not at risk. This suggests that while implementation may have differential associations with outcomes for pupils at varying levels of risk, and hence it is important that it is examined, strict adherence to the manual (that is, high fidelity) may not necessarily be the answer. The results from the present study have important implications for schools seeking effective universal interventions, as they suggest that a “one size
fits all” approach regarding implementation may actually be hindering progress for certain groups of pupils (NEA, 2014), and so schools will need to be cautious regarding the interventions they choose to implement, and the ways in which they do this. The inflexibility of these prescriptive interventions seemingly fails to account for the varying needs between groups of pupils, and the individual differences amongst pupils. Indeed, in the qualitative analysis in RQ2a, no one element of the GBG was reported to be equally detrimental for all pupils with an SEND, but varied depending on the individual needs of the pupils.

Further research needs to be conducted to evaluate other aspects of implementation, such as participant reach, which may help to identify the groups of pupils to whom the intervention is failing to attend (Durlak & DuPre, 2008). One way to ensure that this is achieved would be to make thorough comprehensive implementation and process evaluations of imported interventions standard practice in future research. This would help to provide a clearer picture regarding the ways in which pupils are engaging with and benefiting from an intervention (Dane & Schneider, 1998; Humphrey et al., 2016). Indeed, monitoring all aspects of implementation and examining how they relate to different outcomes are now considered to be essential aspects of all programme evaluations (Durlak, 2015; Durlak & DuPre, 2008; Humphrey, Lendrum, et al., 2016). Further, more explicit, research could also be conducted to identify the aspects of the GBG that are suitable for at-risk groups, in order to ensure that the intervention is benefiting all pupils. The adaptations that teachers make for specific subgroups of pupils, the factors influencing their decisions to make them, and the perceived need for these adaptations, should also be clearly explored when evaluating school-based interventions. Indeed, Durlak (2015) highlighted the need to assess the influence of adaptations as a key research priority. The GBG developers and coaches associated with the intervention could then use this information when assisting teachers with implementation. They may consider placing less emphasis on fidelity, and instead focus on aiding teachers in making appropriate adaptations where necessary for different groups of pupils.
This emphasis placed on fidelity may also be a broader issue relating to the implementation and evaluation of school-based interventions more generally. Fidelity is often the focus for intervention designers, with strict adherence to the manual encouraged, and the other aspects of implementation neglected (Durlak & DuPre, 2008), even though there is no agreement in the literature as to whether strict adherence to the manual is always beneficial (Dusenbury et al., 2003). The ‘fidelity-adaptation’ debate is ongoing (Lendrum & Humphrey, 2012), with tensions between intervention developers’ desire for strict adherence to the programme, and implementers’ wishes to adapt the intervention to suit the context. Indeed, it has been suggested that interventions that are not flexible enough to meet the needs of the context, or those that cannot be adequately adapted to do so, are at risk of failing (Greenberg et al., 2005). Therefore, intervention developers may need to take care to avoid over-emphasising the importance of fidelity, and in fact, may want to work on establishing the adaptations that can be implemented, to ensure that these interventions are truly universal. Lendrum and Humphrey (2012) distinguish between adaptations that are considered to be modifications to existing components, and those that are additions to an intervention, arguing that these do not equate to a lack of fidelity. Therefore, further research could explore the additions that can be made to the GBG, as opposed to modifications that interfere with the intervention’s underlying theory of change; this may help to ensure that fidelity to the critical components remains high, whilst also reducing any detrimental effects for at-risk pupils.

Based on the quantitative and qualitative findings, table 9.1 below outlines adaptations perceived by teachers to be necessary for at-risk pupils. The feasibility of these regarding the GBG’s critical components is also discussed in more detail. Based on the information available, this study cannot advise teachers to implement these specific adaptations; however, they do provide a foundation for the intervention developers to conduct future research, to establish if these adaptations are in line with the logic model and underlying theory of change. They also provide guidance as to the types of adaptations that teachers feel are
necessary, so that the GBG can be developed accordingly for use in the UK and with at-risk pupils.

Table 9.1: Discussion of adaptations

<table>
<thead>
<tr>
<th>GBG Component</th>
<th>Suggested Adaptation</th>
<th>Supporting Data</th>
<th>Comments*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team leadership</td>
<td>At-risk pupils as team leaders</td>
<td>• Section 8.3.1</td>
<td>The manual recommends that a variety of pupils should be team leaders, including shy pupils and those with behavioural difficulties. Thus, it would likely be acceptable to select at-risk pupils to be team leaders.</td>
</tr>
<tr>
<td>Team membership</td>
<td>TA as a team member (particularly at the beginning of implementation)</td>
<td>• Section 8.3.1</td>
<td>Team membership is considered a core element in the manual, as it allows pupils to interact with positive role models. It therefore could be viewed that TAs are simply an additional role model. While no teacher-pupil interaction is specified in the manual, and so this adaptation may interfere with this, this is limited to repeating directions, helping students by answering questions, or giving praise or reminders. Thus, so long as the TA did not provide additional help with the work, this core component would not be interfered with.</td>
</tr>
<tr>
<td>No teacher-pupil interaction permitted</td>
<td>One-to-one support from TA permitted during game</td>
<td>Similarly to above, the manual recommends no teacher-pupil interaction. However, this is limited to certain criteria. The manual also states that there are special circumstances in which it is appropriate to interact with pupils. Thus, if TAs were there simply to keep these pupils on task, or if a pupil’s needs dictated additional support, then this adaptation would not violate this element of the game.</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| Provide task-specific guidance to pupils during game | • Section 8.3.1  
• Section 8.3.4 | As the manual dictates that no additional help should be provided to pupils regarding the task, this adaptation would not be in-line with the critical components of the game. However, if a pupil’s needs dictated additional support, then this adaptation may be acceptable. |
<p>| Pupils can ask questions during game | • Section 8.3.4 | The manual states that the teacher cannot help individual students by answering questions and so this adaptation would not be in keeping with the core components of the game. |
| Teacher intervene to | • Section 8.3.4 | The manual states that there are special instances in which it is |</p>
<table>
<thead>
<tr>
<th>Rigid structure of rules and infractions</th>
<th>Prevent escalating behavioural issues</th>
<th>Appropriate to intervene, and so this would be permitted.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Option of additional parameters for rules</td>
<td>Section 8.3.4</td>
<td>The manual states that rules should always be enforced consistently, with no changes to expectations. However, teachers are also encouraged to outline their expectations of each rule with students, and in special circumstances to select appropriate courses of action that address the underlying cause of behaviour problems. It therefore may be feasible to have different expectations for certain pupils, as long as they are implemented consistently.</td>
</tr>
<tr>
<td>Additional sentence explaining how to rectify behaviour in CCR</td>
<td>Section 8.3.1</td>
<td>There is a strict CCR script in the manual. However, teachers are only told that they cannot repeat directions outside of CCR, and so it may be appropriate to add an additional sentence during CCR.</td>
</tr>
<tr>
<td>Avoiding specific attribution of infraction to certain pupils</td>
<td>Section 8.3.4</td>
<td>The manual advises that teachers should use a non-verbal gesture to identify the pupil that has received an infraction and so this adaptation is not consistent with the critical components of the</td>
</tr>
<tr>
<td>Increasing length of games</td>
<td>Shorter time/smaller tasks</td>
<td>GBG.</td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------------------</td>
<td>------</td>
</tr>
</tbody>
</table>
|                           |                           | • Section 8.3.1  
|                           |                           | • Section 8.3.4  
|                           |                           | Whilst the manual dictates that the length of game play should increase over the course of the year to 40 minutes, it does say that this 40 minute period can be broken down into more frequent, smaller time periods. It also emphasises that the tasks do not have to be the same for all students, and so it would not interfere with the critical components if teachers gave at-risk pupils simpler tasks and played shorter games, as long as they played with increased frequency. |
| TA available as game length increases |                           | • Section 8.3.4  
|                           |                           | Similarly to above points regarding teacher-pupil interaction during the game, the appropriateness of this adaptation may depend on the type of interactions occurring, and whether there are special circumstances surrounding a pupil’s individual needs. |

* All references to core components are from the AIR manual (Ford et al., 2014)

9.9.2.iv Dosage

Regarding dosage, there was a marginal, non-significant trend indicative of an association between high levels of dosage in year 3 and lower disruptive behaviour.
scores for high-risk pupils. Whilst this association was not statistically significant, this analysis was highly exploratory and stronger relationships may have been evidenced had sample sizes been larger, hence this finding still has implications for GBG implementation. This finding supports the intervention’s underlying theory of change regarding frequent implementation (Ford et al., 2014), and in particular suggests that teachers should take care to provide high-risk pupils with high levels of exposure, to ensure that they benefit from the game. It is also in keeping with risk research generally, which typically suggests that children experiencing multiple forms of risk are likely to benefit from more intensive exposure to interventions (e.g. Abbott et al., 1998). However, more research needs to be conducted with larger samples to establish if a significant relationship exists between high levels of dosage and outcomes for at-risk pupils. Furthermore, while this marginal association was evident for high-risk pupils, there were no overall associations between dosage and pupils’ school functioning. This suggests that not all pupils will always respond in the same way to the GBG, and thus highlights the importance of establishing the differential effects of implementation variability for different groups of pupils when evaluating universal interventions, to ensure that they are benefitting all pupils.

However, no external cut-off was utilised to categorise high and low dosage, and it may be that teachers in the high dosage category were still implementing at suboptimal levels, relative to the guidelines outlined in the manual. Indeed, previous GBG research has found suboptimal levels of dosage to be a concern (Hagermoser Sanetti & Fallon, 2011). However, as outlined by Becker and colleagues (2013), no empirically validated benchmark exists regarding the levels of dosage required to achieve the intended outcomes, and so it may be that no overall effect was found because the wrong cut-off for high dosage was used. This has implications for the measurement of dosage when evaluating the effectiveness of interventions, indicating that more research needs to be conducted to determine the levels of exposure necessary to trigger the intended mechanisms of change for specific interventions. This may also have implications for intervention developers; if more work was conducted in this area, patterns may emerge regarding the levels of dosage commonly implemented. If suboptimal dosage was frequently identified
as an issue, this may be an indication that unfeasible levels of implementation are being required from teachers, and thus intervention designers may need to address this when developing their programmes. Hence, further work needs to be done to establish the level of dosage necessary to trigger change, and the levels of dosage that are acceptable to teachers. This may in turn help intervention designers find the optimal level, leading to empirically validated benchmarks for high and low dosage. This means that more accurate research will then be possible, to better determine the importance of dosage, and to establish any differential subgroup effects.

9.9.2.\textit{v} \textit{Summary}

In summary, the findings from the present study demonstrate that implementation is a complex issue that does not affect all pupils equally, and so should not be ignored when evaluating the effectiveness of school-based interventions. While the GBG had no overall effect on behaviour for pupils at varying levels of risk, its implementation appeared to be potentially influential for high-risk pupils. Conversely, the GBG possibly had greater benefits for at-risk pupils’ reading attainment (although these results were marginally non-significant), regardless of the way it was implemented. This is useful information for the developers of the GBG, as it appears that not only could the GBG have differential gains for different groups of pupils, but that the importance of implementation for these groups varies between outcomes.

This also has broader implications for the development and evaluation of school-based interventions generally; whilst these interventions are often considered to be universal, subgroup explorations need to be conducted before this can accurately be determined. Future research needs to be conducted to establish any differential associations between outcomes and other aspects of implementation, and should explore the differential effects of implementation for at-risk pupils on other outcomes variables (Durlak, 2015). Similar analyses could also be conducted regarding other school-based interventions, to establish the importance of
implementation in achieving the intended outcomes for different groups of pupils. This in turn can inform the development and implementation of these universal interventions, in order to help ensure their success for all pupils.

Intervention designers should consider that there may be varying mechanisms of change required to achieve different intended outcomes, and that these will need to be identified before being incorporated into the critical steps outlined in the manuals. Furthermore, as these mechanisms of change do not appear to act in the same way for all groups of pupils, guidance on appropriate adaptations to address issues with certain groups of pupils could also be provided.

9.10 Contribution to Knowledge

The present study makes several distinct contributions in the areas of prevention and implementation science. Firstly, the study has contributed to the evidence base regarding risk variables associated with pupils’ school functioning. Of the pupil-level risk variables that are well-established in the extant literature base, the present study has confirmed their associations with disruptive behaviour and/or reading attainment in a current UK dataset. However, several risk variables in the literature have previously been met with some ambiguity, and so the present study helped to bring clarity to these areas by identifying the variables that are associated with pupils’ school functioning, and the magnitude of their association.

By exploring 16 potential risk variables in various ecological domains, confounding effects of other proxy variables are more likely to have been controlled for. For example, studies exploring pupils’ looked-after status often fail to address the higher levels of SEND diagnoses that is prevalent amongst this group (DfE, 2017a), while familial and neighbourhood deprivation are frequently conflated in risk research (McCulloch, 2006). However, the present study has identified the individual and unique contributions of each variable to pupils’ outcomes. Finally, comparatively fewer studies conduct research into the school-level risk variables associated with pupils’ outcomes, and so the present study has contributed to the
evidence base by identifying the variables at this level that are associated with pupils’ school functioning.

In addition to contributing to the evidence base on individual risk variables, the present study has also made several contributions regarding the cumulative risk effect identified. Whilst most risk research typically explores individual risk variables, the present study has accounted for the clustering that tends to occur between them (Flouri & Kallis, 2007), and so has examined the effect of multiple risk variables on pupils’ school functioning. Results have confirmed that it is important that risk research is conducted in this way. Furthermore, cumulative risk research typically only incorporates risk variables from a single domain (Oldfield et al., 2015), and so this study has contributed to the evidence base by examining the cumulative risk effect across multiple domains. These findings have increased understanding of how risk exposure influences pupils’ behavioural and academic outcomes; they also have ramifications for the way future research is conducted and the way that school-based intervention target at-risk pupils.

Additionally, the present study has contributed to the understanding of how cumulative risk exposure influences pupils’ outcomes, by examining the functional form of the risk-outcome relationship. Previously, an equal number of studies had found a linear as opposed to a non-linear relationship (Evans et al., 2013), leading to ambiguity in this area. However, this study has confirmed that a quadratic relationship exists for both outcomes in the present dataset. This finding provides support for the theory of mass accumulation in the extant literature (Flouri & Kallis, 2007; Oldfield et al., 2015). Furthermore, the threshold and saturation effects identified contribute to the evidence base, providing further information regarding the ways in which exposure to multiple risk variables leads to negative outcomes. Finally, the study offers a further contribution to knowledge by investigating both assumptions of cumulative risk theory, as previous research often only focuses on one. This improves the rigour of the resultant findings, providing an additional insight into the mechanisms through which risk variables operate, and the best ways to treat risk variables in future research.
Regarding the GBG, the present study is, to the best of the author’s knowledge, part of the largest RCT conducted of the intervention worldwide to date, thus providing a comprehensive and rigorous examination into the effects of the GBG. Furthermore, this is the first RCT of the intervention in a UK setting, providing important contributions regarding the cultural transferability of the game. Typically, when the effects of the GBG are explored, only one aspect of pupils’ school functioning is examined, and so the present study further contributes to the evidence base by examining the effects on both pupils’ behavioural and academic outcomes. Additionally, the present study utilised a design whereby the effects were tested in isolation; often when academic outcomes are examined, the GBG is combined with another intervention (e.g. Bradshaw et al., 2009). Thus, this study has also built on previous research by exploring the unique contributions of the GBG to pupils’ reading attainment.

Furthermore, the study has advanced understanding of differential programme benefits in preventive interventions (specifically, the GBG), an area highlighted as a priority for future research in Durlak and colleagues’ (2011) meta-analysis. The limited amount of previous research into differential intervention gains has found that the outcomes of interventions will likely vary as a function of individual differences (Humphrey et al., 2013); however, differential effects of the GBG among pupils at different levels of risk exposure have not previously been assessed. There is little research conducted into the differential gains of the GBG, and other school-based interventions more generally, and so the present study contributes to this area by exploring the overall effects of the intervention alongside any differential subgroup effects. Of the GBG studies that have examined differential effects of the game, they typically only explored outcomes for one subgroup of pupils, or pupils exposed to a single risk variable, such as male pupils (e.g. Dolan et al., 1993; Kellam & Anthony, 1998) or pupils eligible for FSM (Ialongo et al., 1999; Petras et al., 2008). However, the present study examined the differential effects of the game for pupils exposed to multiple risk variables, whilst also exploring the effects for pupils at
varying levels of risk, thus providing a greater insight regarding the true universality of GBG.

The used of mixed-methods in the present study ensured that the research questions regarding differential gains were answered as comprehensively as possible, meaning that a deeper exploration could be conducted into not just “what works” for pupils, but also how and why it works under certain conditions and circumstances (Bonell et al., 2012). The qualitative findings identified several previously unexplored subgroups of at-risk pupils for whom the GBG had differential effects. While GBG research typically focuses on boys or those at-risk of behavioural difficulties (e.g. Dolan et al., 1993; Kellam & Anthony, 1998), the findings from the present study suggest that there are pupils exposed to other specific risk variables that also respond differently to the game. Furthermore, while research typically focuses on behavioural outcomes, 10 alternative, more proximal outcomes also emerged in the qualitative dataset. These findings pave the way for future research in this area regarding groups of at-risk pupils, and highlight the importance of conducting mixed-methods research when exploring the differential subgroup effects of an intervention.

Finally, only a handful of studies have explored variability in implementation of the GBG, and even fewer have linked these variations to pupils’ outcomes (Lendrum & Humphrey, 2012). By failing to measure implementation when evaluating the effectiveness of an intervention, studies are susceptible to Type III error, as the outcomes and processes of an intervention are interrelated (Askell-Williams et al., 2012). The present study therefore contributes to the evidence base in this area by examining the link between implementation variability and successful outcomes. Furthermore, of the studies that have incorporated implementation into their analysis, it is typically only fidelity is examined (Durlak & DuPre, 2008); thus, this study has a addressed a major gap in the literature by examining three different aspects of implementation, and their individual associations with both disruptive behaviour and reading attainment. Indeed, Durlak (2015) identified specifying the
aspects of implementation that are related to specific program outcomes as apriority for future work.

The study will also advance current research in the field of implementation science as it is amongst the first to examine differential gains of the GBG among at-risk children as a function of implementation variability. While there were suggestions in the literature that variability in implementation may be differential associated with outcomes for pupils at varying levels of risk (Abbott et al., 1998; CPPRG, 1999), this had not previously been explicitly explored, and so the findings from the present study contribute to this area of the research. Finally, the adaptations that were perceived to be necessary for different groups of at-risk pupils also emerged in the qualitative dataset. These findings have contributed to the evidence base by identifying the elements of the GBG that are perceived to be inappropriate for pupils, helping to explain any differential effects of the game. This also means that future work can be conducted to tailor the intervention so that it is beneficial to all pupils.

9.11 Summary and Conclusions

In summary, the present study aimed to (1) increase knowledge and understanding of the risk variables for poor school functioning in academic and behavioural domains among children in primary school, (2) assess the impact of the GBG on children at different levels of risk exposure, and (3) determine the extent to which differential benefits among at-risk children vary as a function of implementation.

The aims were achieved through a mixed-methods study of 3,084 pupils in 77 schools. The study was a longitudinal RCT following a hybrid concurrent embedded design. 16 predictor variables were utilised to identify the risk variables for school functioning, and measures of disruptive behaviour and reading attainment were utilised as the outcome variables. Analysis was carried out using MLM for the quantitative strand. Significant risk variables were identified in the first analysis, before being used to generate a cumulative risk score. This was then used to test
the underlying assumptions of cumulative risk theory and to identify the functional form of the risk-outcome relationships. The second analysis involved categorising pupils into risk groups and testing the interactions between trial group allocation and risk status. Finally, the third analysis involved half of the dataset, whereby the interactions between FidQual/dosage and risk status were tested.

Results revealed six risk variables for disruptive behaviour, and seven risk variables for reading attainment, both at the pupil- and school-level. Both assumptions of cumulative risk theory were met, and quadratic relationships were identified for both outcome variables. Regarding the GBG, there were no statistically significant overall effects of the game on school functioning, or subgroup effects for pupils in any of the risk groups. High levels of FidQual were significantly associated with worsening behaviour scores for high-risk pupils; however, there were no differential subgroup associations between implementation variability and pupils’ reading attainment.

Thematic analysis was used for the qualitative strand of the study. Semi-structured interviews were conducted in six case study schools implementing the GBG twice a year over the course of the trial. Perceived differential gains were identified for pupils exposed to six different risk variables; 10 proximal outcomes also emerged, along with five common GBG elements considered to be the mechanisms through which these outcomes were influenced.

Limitations to the study included the attrition rates, and the number of GBG schools ceasing implementation, alongside the cross-sectional design through which the risk variables identified, and the behaviour, reading, and implementation measures utilised. However, many of these limitations were addressed, and appropriate methods were used where possible to overcome these issues. The implications of the results are far-reaching, and are relevant to intervention designers, teachers and members of the senior leadership team in schools, and other researchers examining risk and school-based interventions.
Finally, the present study provides several demonstrable theoretical, practical and methodological contributions to knowledge in the fields of prevention and implementation science. The results emphasise the importance of addressing the unique and complex interactions that occur between risk variables, whilst also highlighting the need for further research into the implementation of universal interventions. Due to the exploratory nature of the study, particularly regarding differential gains and implementation variability, more research is required in order to ensure the best possible outcomes for pupils experiencing multiple levels of risk. Further research will help with the development of school-based interventions that effectively target the negative consequences of multiple risk exposure and the mechanisms through which they occur, while also equipping at-risk children with the necessary coping skills to protect them against any adversity that they may face.
References


Askell-Williams, H., Dix, K., Lawson, M., & Slee, P. (2012). Quality of implementation of a school mental health initiative and changes over time in students’ social and emotional competencies. *School Effectiveness and School Improvement, 24*, 357–381.


Barker Bausell, R., & Li, Y. (2002). *Power analysis for experimental research: A
practical guide for the biological, medical and social sciences. Cambridge University Press.


EEF. (2013). *Pre-testing in EEF evaluations*.


Francis, B., & Archer, L. (2005). British-Chinese pupils’ and parents’ constructions of


Huffman, L., Mehlinger, S., & Kerivan, A. (2000). Risk factors for academic and behavioral problems at the beginning of school. In *Off to a good start: Research on the risk factors for early school problems and selected federal*
policies affecting children’s social and emotional development and their readiness for school (pp. 1–93). Chapel Hill, NC: Frank Porter Graham Child Development Center.


Jackson, C. (2010). “I’ve been sort of laddish with them … one of the gang’: teachers’ perceptions of “laddish” boys and how to deal with them. Gender and Education, 22, 505–519.


Mazzoli-Smith, L., & Todd, L. (2016). *Poverty proofing the school day: Evaluation and development report*.


McRobbie, J. (2001). *Are small schools better? School size considerations for safety*


NEA. (2014). *Positive behavioral interventions and supports: A multi-tiered framework that works for every student.*


Ofsted. (2014). *Below the radar: Low-level disruption in the country’s classrooms.*


330


Prelow, H., & Loukas, A. (2003). The role of resource, protective, and risk factors on academic achievement-related outcomes of economically disadvantaged


333


Songer, N., & Gotwals, A. (2005). Fidelity of implementation in three sequential curricular units. In S. Lynch (Chair) & C. L. O’Donnell, “Fidelity of implementation” in implementation and scale-up research designs: Applications from four studies of innovative science curriculum materials. Montreal, Canada: AERA.


https://www.telegraph.co.uk/news/2016/11/10/white-working-class-boys-perform-worst-at-gcse-research-shows/


Appendix 1: TOCA-C

Has this child been present in your classroom for a majority of the time (i.e., at least 8 days) during the last three weeks?

Yes | No

In the last three weeks, would you say the following statements were never, rarely, sometimes, often, very often, or almost always true of this child . . .

<table>
<thead>
<tr>
<th>No.</th>
<th>Statement</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Very often</th>
<th>Almost always</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Concentrates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Is friendly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Pays attention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Breaks rules</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Is liked by classmates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Doesn’t get along with others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Works hard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Harms others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Shows empathy and compassion for others’ feelings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Gets angry when provoked by other children</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Stays on task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Yells at others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Is easily distracted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Is rejected by classmates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Fights</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Lies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Has many friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Harms property</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Completes assignments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Teases classmates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Learns to up ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 2: GBG Observation Schedule

### 2. Classroom
- **Number of children**
  - Number of children in the classroom
  - Number of children who are not participating in the activity

### 4. During Game
- **Player Management**
  - Check, Convey, Respond
  - Player management is not addressed

### 5. Post-Game
- **Team Management**
  - Team management is not addressed
  - Team management is addressed

### 6. Overall - Teacher
- **Quality of Environment**
  - Quality of environment is not addressed
  - Quality of environment is addressed

---

<table>
<thead>
<tr>
<th>Description</th>
<th>Procedure</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huddle time</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle location</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle structure</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle duration</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle content</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle participation</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle leadership</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle assessment</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle interpretation</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle evaluation</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle feedback</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle adaptation</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle interpretation</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle management</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Huddle maintenance</td>
<td>Yes</td>
<td>1</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Rule</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule 1</td>
<td>No</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Rule 2</td>
<td>No</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Rule 3</td>
<td>No</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Rule</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule 1</td>
<td>No</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Rule 2</td>
<td>No</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Rule 3</td>
<td>No</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Appendix 3: Teacher Interview Schedule

The Good Behaviour Game - Year 3 Teacher Interview Schedule (semi-structured)

<table>
<thead>
<tr>
<th>Interview data: aims</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To explore, understand and explain the processes of implementation of the GBG in English educational contexts</td>
</tr>
<tr>
<td>2. To triangulate with and support interpretation of observation and impact data</td>
</tr>
</tbody>
</table>

Data needed for:
(a) Examining processes of implementation
   - Fidelity – the extent to which the school is adhering to the intended treatment model
   - Dosage - how much of session delivered; number of sessions
   - Quality – how well different the GBG is delivered
   - Participant responsiveness – the degree to which children and their parents engage with the intervention
   - Programme reach – rate and scope of participation
   - Monitoring of control conditions
   - Adaptation – the nature and extent of changes made to the intervention

(b) Identification of context specific factors affecting implementation
(c) Evaluating the feasibility of the future implementation of the GBG in English educational contexts

Preamble

1. Check that the interviewee has received the information sheet and consent form and understands the project and his/her role in it.
   Ask: Have you any questions about the project?

2. Emphasise that:
   - The research team is speaking to a range of people involved in the GBG e.g. senior management, teachers, pupils at all of our the GBG schools
   - We are interested in individual experiences and thoughts about the GBG, both positive and negative... “this is your opportunity to make your voice heard on the GBG ... your comments may be helpful to others in your position at other schools at a later date”
   - However, we combine all the data we collect to provide an overall picture of the GBG and its implementation and any comments in the report are attributed very generally, for example, as “A (Year 3) teacher commented that...” . Any comments/opinions will not be reported back to schools
Ask: Have you any questions about how we use your comments?

**Ethics:**
Remind interviewee:
- The interview will take about 30 minutes.
- You do not have to answer any questions that you are not comfortable with
- You can stop at any time, no explanation needed
- If any question doesn’t make sense, ask for an explanation

Ask: Is it alright to record the interview? The transcript will only be seen by those working on the project. I will send you a copy too if you wish.
Ask: Are you able/willing to sign the consent form?

**Explain procedure:**
I will begin the interview with my name, the date, time and the identifying code we have assigned to your school - this is just to keep the recordings organised. All your details will be anonymised when the data is transcribed.

The first question will be about your role in school, followed by general questions about social and emotional learning in school, then moving on to the GBG more specifically

Ask: Have you any questions before we start?

Ask: Is it okay for me to start recording now?
Interview schedule

State researcher’s name, date, time, school identifying code *(for data management)*

Can I just ask you to confirm your roles at school........

........and in relation to the GBG *(e.g. Y3 teacher etc.)*

A. Usual practice (Implementation - programme differentiation)

\[\text{Aims: to clarify foundations for the GBG and school ethos around behaviour; perceptions of benefits of the GBG/behaviour interventions; perceptions of need for the GBG; previous practice around behaviour management, whether starting the GBG has been integrated or resulted in changes to this.}\]

1. How would you describe the overall profile of the GBG in your school?
   - Is it just classroom teachers in Y3 (Y4) that are involved?
   - How involved is the Head Teacher? Senior management team?

\[\text{Looking for information about:}\]
   - Type of HT/SMT support
     - Verbal only?
     - Active e.g. training time allowed, curriculum time allowed, included in planning etc.

2. How does the GBG compare to other behavioural management strategies you have used?

B. Implementation - What progress have you made playing the GBG?

\[\text{Aims: clarify implementation dosage and fidelity; modifications or adaptations and reasons for them; generalisation (link to quality)}\]

3. How long have you been implementing the GBG?

4. How often do you play the GBG? *Ask for example.*

5. When do you typically play the GBG? Where do you typically play the GBG?

\[\text{Looking for:}\]
   - Certain days, times, particular lessons, activities etc.
   - Certain locations etc.
6. How long do you typically play the game for? Has this changed over the year?

7. Have you had to make any changes to the GBG in order to suit your class or your teaching?
   - How have you dealt with pupils with additional needs, or need support from teaching assistants? Have you had to adapt the GBG to meet their needs?
   - Have you had to make any changes to the way GBG is played?

From time to time, teachers may encounter situations that require multiple responses or different courses of action as a result of GBG. For example, they may find that the teams are not winning GBG or individual students frequently cause their teams to lose GBG. Have you encountered such situations?

Aims: Establish whether teams are balanced across gender, behaviour, and learning, and, if not, rearrange the teams or temporarily shorten GBG time.

8. Are all pupils in the class present for the GBG?

Looking for:
- Participant reach
- Is the GBG session used as withdrawal time? If so, do these pupils have the GBG at another time?
- Do some pupils have a more targeted approach? Is this in addition or instead of the GBG?

9. What sorts of things have you used a reinforcers?
   - Have you made changes to the sorts of things offered as reinforcers/rewards?
   - If so why? Give examples.

Looking for:
- Tangible
- Intangible
- Tokens
- Immediate
- Delayed

10. Have you/the pupils been able to apply/generalise from the GBG outside the classroom e.g. playtime?
C. Factors affecting implementation

Aims: factors influencing implementation; operation of factors as barriers or/and facilitators; responses to barriers (e.g. disregarded, re-active/pro-active adaptations)

The list below outlines the key anticipated factors that may influence the implementation of the GBG at programme, classroom and school levels (there is likely to be interaction across levels). The questions above should have addressed most of these; however, please be aware of these factors so that answers may be probed or questions revisited if necessary.

Factors potentially affecting implementation

<table>
<thead>
<tr>
<th>Level</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programme level</td>
<td>suitability of resources</td>
</tr>
<tr>
<td>Teacher level</td>
<td>self-efficacy, knowledge and skill proficiency, level of training,</td>
</tr>
<tr>
<td></td>
<td>external support, support from colleagues, curriculum time, preparation</td>
</tr>
<tr>
<td></td>
<td>time, attitude/buy-in – do not perceive need for or benefits of the GBG</td>
</tr>
<tr>
<td></td>
<td>– not compatible with teaching style</td>
</tr>
<tr>
<td>Pupil level</td>
<td>Meets needs, engaged, responsive (appropriate resources), classroom</td>
</tr>
<tr>
<td></td>
<td>climate/pupil behaviour impede implementation</td>
</tr>
<tr>
<td>School level</td>
<td>Prior positive/negative involvement with similar approaches and existing</td>
</tr>
<tr>
<td></td>
<td>climate supportive/not supportive of SEL/the GBG approach</td>
</tr>
<tr>
<td></td>
<td>the GBG integrated with other aspects of curriculum</td>
</tr>
<tr>
<td></td>
<td>Head teacher and senior management team actively supportive of the GBG</td>
</tr>
<tr>
<td></td>
<td>Head teacher and senior management state that supportive but not actively</td>
</tr>
<tr>
<td></td>
<td>demonstrating support (status of the GBG within school)</td>
</tr>
<tr>
<td></td>
<td>Sufficient resources allocated – classroom/curriculum time</td>
</tr>
<tr>
<td></td>
<td>The GBG integrated with other aspects of school-life posters just in</td>
</tr>
<tr>
<td></td>
<td>classrooms or across school? Whole-staff awareness of the GBG (other</td>
</tr>
<tr>
<td></td>
<td>than teachers directly involved in delivery).</td>
</tr>
</tbody>
</table>
11. How easy has it been to implement the GBG?
   - Is there anything about your school that has made it easier?
   - Is there anything about your class that has made it easier?

Looking for effects of classroom climate on ease of implementation.

12. Have there been any challenges to the implementation of the GBG?

Ask for specific examples (positive and negative).

D. Attitudes to the GBG specifically

Aims: clarify teacher and pupil attitudes to the GBG, including perceptions of impact; clarify fidelity and dosage, pupil responsiveness; describe and/or explain modifications or adaptations; inform interpretation of process data; inform future roll-out of the GBG in UK context

13. What do you think about the GBG?

Ask for examples; probe for explanations.

14. How familiar are the concepts, strategies?

Looking for:
   - Changes to usual practice, foundations for the GBG

15. How useful do you feel it is to have a structured approach to classroom management?
   - Do you find yourself using the ideas/strategies outside of the game, across other lessons, activities?

16. What do you think about the GBG resources (if not included above)
How appropriate/suitable are the resources?

How useful has the GBG been for meeting specific needs in your class?

Are there any aspects of the GBG that you have found particularly useful for your class?

Are there any aspects of the GBG that you have found not useful/appropriate?

What do the pupils in your class think about the GBG?

Are some groups more responsive than others (e.g., SEN, EBD, quiet/withdrawn)?

The GBG has been designed for all the children in the class; have you found that it is useful for some groups more than others? (e.g., EAL, SEN, EBD, withdrawn)

Do they look forward to doing the GBG?

Are they engaged by?

Are there any particular aspects they like?

How do they respond/feel about winning?

How do they respond/feel about losing?

Do the pupils support each other in teams?

Ask for specific examples (positive and negative).
E. Training/coaching

Aims: teacher perceptions of self-efficacy, confidence, competence, skills and/or knowledge to implement the GBG; attitudes to training – quantity/quality, timing, content, utility etc; attitudes to support/coaching model - quantity/quality, timing, frequency, type of support available, utility etc. (NB to inform future roll-out)

18. Have you had any additional training relating to the GBG specifically?

19. Have you had any other opportunities for training/professional development around classroom behaviour management?

Ask for examples.

20. On-going support (coaching model) - In addition to the initial training, the GBG programme includes ongoing support from a GBG coach who has been assigned to your school. How important do you feel it is to have access to ongoing support?

- What sort of initial support have you received so far, and how useful has this been?
- How often do you have support visits from your GBG coach?
- Has the coaching support changed over the year?

F. Perceptions of impact:

21. Has the GBG made a difference to your pupils? All pupils, or some groups of pupils particularly? The school more widely?

- General
- Reading/academic achievement
- Behaviour
- Particular groups of pupils (i.e. boys, FSM)

Ask for examples e.g. improved behaviour, relationships, social skills, understanding of emotions, self-control, confidence and participation (e.g. quiet pupils more prepared to participate), classroom climate/ethos/atmosphere, learning, motivation for learning, attendance, SEN.

22. Has the GBG made a difference to you?

- Has it affected classroom management?
- Has it made it easier to be a teacher?
• How does it compare to what you did before?

G. Sustainability

Aims: attitudes towards the GBG and change over time; sustainability

23. The GBG project runs for two years; how likely do you think it is that you will continue with the GBG after this?
   • The entire programme?
   • Particular lessons?
   • The GBG framework (structure) but with amended lessons?
   • Key aspects of the GBG? (Ask for examples)

Ask for specific examples (positive and negative).

24. If you were not doing GBG as part of a research trial would you continue to use it?

As you know, the project is examining how well the GBG works in English schools. If it is successful, then it may be rolled out to more schools. Based on your experiences of the GBG so far, what advice would you give to a teacher in another school who has just been told she/he has to implement the GBG next term?

Ask for specific examples if appropriate (positive and negative).

H. Closing the interview

Aims: unanticipated experiences, factors etc.; emergent themes

• Is there anything that you would like to add?
• Is there anything that you think I should have asked you about, or missed out?

ASK: Do you have any questions?

Thank you very much for your help and time. I will now turn off the recorder.
Appendix 4: Pupil Focus Group Schedule

The Good Behaviour Game - Pupil Focus Group Schedule

<table>
<thead>
<tr>
<th>Participants</th>
</tr>
</thead>
</table>
Pupils from each of Y3 (2015/16) and Y4 (2016/17) with focus on EAL, SEND and FSM. Mix of boys and girls as far as possible.

<table>
<thead>
<tr>
<th>Aims</th>
</tr>
</thead>
</table>
1. To explore pupil responsiveness to and engagement with the GBG (as intervention e.g. resources; as delivered by teachers e.g. quality)
2. To explore generalisation of concepts/strategies
3. To explore ‘fit’ with English educational context
4. To triangulate with data on:
   - Pupil responsiveness/engagement
   - reach
   - generalisation of concepts/activities,
   - fidelity/dosage (frequency of delivery; knowledge and understanding of key concepts and strategies)

NB Check that the school has received consent forms from parents

Preamble
I’m having a chat with you today to find out all about the GBG in your school and what you think about it. I will be talking to children in other schools too, to see what they think.

Are you okay to tell me about the GBG? I will ask you some questions, but if you don’t want to answer then that is okay. It isn’t a test and there aren’t any right or wrong answers – I am just interested in what children like you think about the GBG.

I am going to record our chat on this recorder – I am going to put it here so that it will be able to pick up what you each say. I am recording it so that I can listen to it again later and won’t have to write down your answers whilst we chat. I won’t take your names or play the recording to the school, teachers or your parents or tell them what you say. However/but if someone tells me something that makes me think that they are being hurt or are going to hurt someone else, then I will need to let someone know. Is it okay to record our chat?

Ask: Have you any questions?

It’s okay to tell other people that we have had a chat about the GBG today. But because I am asking you what you think about the GBG, it’s also important that we don’t tell others what each of you have actually said, because that isn’t really fair
If you feel okay about taking part in our chat and me recording it, then just tick the smiley face/write your name at the bottom of the sheet. If you don’t want to take part, that is OK.

(if a pupil decides at this point that he/she does not wish to participate and there is no support available to escort the him/her back to the classroom, allow the child to move their chair away from the group if they wish/ have an activity prepared e.g. paper and pencil to draw)

Ethics:
Remind pupils:
- Our chat will take about 20 minutes.
- If you don’t want to answer any questions, that is fine
- If you don’t understand any of the questions or what I am asking, then tell me and I will explain it
- If you don’t want to carry on once we have started, that is OK

Explain procedure:
I will start by saying my name, the date, time and a code number for your school.

The first question will be to do with what you know about the GBG

Ask: Have you any questions before we start?

Ask: Is it okay for me to start recording now?

Focus Group Schedule

NB as this is as a focus group, the aim is to let the pupils discuss the GBG independently. It is suggested that you start with question 1 and then refocus discussion as and when necessary to answer questions 2 – 4 and to achieve aims of FG as outlined above. Use visual aids if needed to assist with memory, language and understanding.

1. Tell me about the GBG. Prompts:
   - What do you do/play (fidelity)?
   - What have you learnt (fidelity; understanding)?
   - Have you found that the GBG helps you in anyway? How/why? (ask for examples e.g. do you use the GBG in the classroom, outside the classroom, at home)
     (Prompt further if needed e.g. indicate ‘visual aids’)
     (looking for generalisation of concepts/strategies)

2. What do you like about the GBG? (ask for examples)
   (Prompt further if needed e.g. indicate ‘visual aids’)
   (responsiveness; engagement)

3. Is there anything that you don’t like about playing the GBG?
(Prompt further if needed e.g. indicate ‘visual aids’)  
(responsiveness; engagement)  

4. When do you play the GBG? (dosage; reach e.g. are pupils with SEND withdrawn)
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.

In consenting to your child’s participation, you are also giving permission that for the purpose of the study, information provided will be linked with the National Pupil Database (held by the Department for Education), other official records, and shared with the Department for Education, Education Endowment Foundation (EEF), EEF’s data contractor FFT Education, and in an anonymised form to the UK Data Archive.
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

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:

<table>
<thead>
<tr>
<th>My name</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>My child’s name</td>
<td></td>
</tr>
<tr>
<td>Name of my child’s school</td>
<td></td>
</tr>
</tbody>
</table>

Signed: ___________________________  Date: __________
Appendix 6: Case Study Teacher Information Sheet and Opt-In Consent Form

THE GOOD BEHAVIOUR GAME

www.gbguk.org

Case study school - Teacher information sheet

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.

In addition to our annual implementation and outcomes surveys, we will also be conducting case studies in 6 GBG schools. These case studies are a very important part of the project as they will allow us to assess whether the GBG meets schools’ needs, how it is viewed by members of the school community, and whether it can be delivered successfully.

Your school has expressed an interest in participating in the case study school part of the trial.

In this letter we describe what participation as a case study school will entail. 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?

Your school is taking part in the Good Behaviour Game Project and has agreed to participate as a case study school.

What would I be asked to do?

One of our researchers will need to visit your school for one fieldwork visit per term over the two years of the trial - 6 visits in total.

During the course of these visits, we will need to speak to key people involved in the implementation of the GBG. This is likely to include the head teacher, Year 3 and 4* teachers and classroom assistants.

We will also collect information from a number of parents and children in order to obtain their views of the GBG.

We would also like to conduct informal observations of the Year 3 and Year 4 * classes, in particular observe a GBG lesson in practice.

Information will be collected using a variety of methods, such as interviews, focus groups, observations, and analysis of school documentation.

*In the second year of the trial

Interviews and focus groups will be conducted by experienced and trained members of the research team, all of whom have undertaken Disclosure and Barring Service checks at the Enhanced level. They will take place in a quiet, private room at your school, or over the telephone in the case of some parents, at a mutually convenient time. All data will be anonymised during the transcription process, with pseudonyms given to participants in
order to guarantee anonymity and confidentiality. For children selected to participate in focus groups, we will obtain opt-in consent from their parents/guardians.

What happens to the data collected?

The data will be stored on a secure server at the University and analysed by our research team. 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 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 anonymised.

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 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).

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.gbguk.org

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

If you would like more information please contact Dr Alexandra Barlow (email: alexandra.barlow@manchester.ac.uk, tel: 0161 275 3504). We will be in touch shortly to arrange a convenient time for the first research visit.

Yours sincerely,

Dr. Alexandra Barlow
Research Associate
0161 275 3504
alexandra.barlow@manchester.ac.uk

Professor Neil Humphrey
Principal Investigator
0161 275 3404
neil.humphrey@manchester.ac.uk
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 case study strand of the Good Behaviour Game project. 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.

1. I confirm that I have read the information sheet on the above study, have had the opportunity to consider the information and ask questions, and had these answered satisfactorily. 

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason.

3. I understand that the interviews will be audio-recorded

4. I agree to the use of anonymous quotes

5. I agree that any data collected may be published in anonymous form in academic books or journals.

My details are as follows:

<table>
<thead>
<tr>
<th>My name</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>School name</td>
<td></td>
</tr>
<tr>
<td>Year group</td>
<td></td>
</tr>
</tbody>
</table>

Signed: ____________________________  Date: _________
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 one of our case study schools involved in the project. As a case study school we will use interviews and focus groups drawing upon the views of a range of informants (e.g., pupils, teachers, school leaders, parents). We would like to speak to your child as part of a focus group with other children from the school about their involvement in the Good Behaviour Game.

Please take time to read the following information carefully and decide whether or not you would like them 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 the 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 and are implementing the Good Behaviour Game over a two year period as part of the research project.

What would my child be asked to do if he/she took part?

A member of the research team will conduct a focus group at your child's school, involving your child and another few pupils, on one or two occasions between March 2016 and July 2017. The focus group discussion will cover general questions around their involvement in the Good Behaviour Game, for example, do they like playing the Good Behaviour Game? If so, why? If not, why not? The focus groups will be audio recorded and each take no more than 30 minutes.

What happens to the data collected?

The audio recordings will be transcribed so that they can 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 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 name or your child’s name) will not be used.

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.
Please complete the attached consent form to let us know if you are happy for your child to take part.

If you decide your child can take part and then change your mind, you are free to withdraw them from the focus group part of the study, or the whole study, 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**

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’s participation in the focus group 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.
PARENT CONSENT FORM – FOCUS GROUPS

If you are happy for your child to participate please complete and sign the consent form below. Then return it to the address below using the freepost code given (no stamp required):
FREEPOST RLYU-KAAB-AXRC
Dr. Alexandra Barlow,
Manchester Institute of Education
The University of Manchester,
Ellen Wilkinson Building
Oxford Road,
Manchester, M13 9PL.

1. I confirm that I have read the information sheet on the above study, have had the opportunity to consider the information and ask questions, and had these answered satisfactorily.

2. I understand that my child’s participation in the focus group study is voluntary and that I am free to withdraw him/her at any time without giving a reason.

3. I understand that the interviews will be audio-recorded

4. I agree to the use of anonymous quotes

5. I agree that any data collected may be published in anonymous form in academic books or journals.

6. I agree my child can take part in the above project

Please Initial Box

My name

My child’s name

Name of my child’s school

Signed: ___________________________ Date: ___________
Appendix 8: Case Study Pupil Opt-In Consent Form

The Good Behaviour Game

PUPIL FOCUS GROUP

I’m having a chat with you today to find out all about the GBG in your school and what you think about it. I will be talking to children in other schools too, to see what they think.

☑️ I am happy to talk to you about the GBG.

☑️ I know if I don’t want to answer then that is OK.

☑️ I know this is not a test and there are no right or wrong answers.

☑️ I am happy for you to record our chat on the recorder.

☑️ I know my name won't be used.

☑️ I know the recording will not be played back to the school, my teachers or my parents.

If you feel OK about taking part in our chat and me recording it, then just tick the smiley face/write your name at the bottom of the sheet. If you don’t want to take part, that is OK.

Please tick .......................... ..........................

Name ........................................................................................................................................

Year Group ..................................................................................................................................
Appendix 9: School Cover Letter

www.gbguk.org

5th June 2015

Dear colleague,

Thank you for agreeing to participate in the Good Behaviour Game (GBG) project. The project is a collaboration between the Education Endowment Foundation, Mentor UK, and the University of Manchester. The University of Manchester are leading the evaluation aspect of the project.

Next steps:

We are writing to inform you about some important "next steps" in the evaluation.

1. Parental consent:

   Parental consent needs to be sought before we can collect any survey data about your pupils. Enclosed are parent information and consent letters. Please distribute these to all parents of your current Year 2 pupils. We strongly advise that you send this information directly to parents at their home address. We have asked parents to only return the slip if they wish to opt out of the evaluation. The letters contain a freepost code and address for them to return it to the research team. We will inform you of any parents who have opted out. However, if parents do return the slip to you, please forward to the research team immediately.

2. Teacher-pupil surveys:

   As part of our baseline assessments, teachers are required to complete a short online survey for all current Year 2 pupils at your school. Each survey should take on average 2-3 minutes per pupil and can be completed by any member of staff who knows the pupil well (i.e. teaching assistants as well as class teachers). The survey will focus on behaviour. Our online survey window will open Monday 22nd June, and can be accessed 24 hours a day, 7 days a week. Further information and instructions on how to complete these surveys will be sent to you nearer the time.
3. Teacher wellbeing and usual practice surveys

In addition to the above, each teacher who will be taking Year 3 in September 2015 will be required to complete a short survey on their usual practice in behaviour management and their wellbeing. This survey should take no more than 15 minutes to complete. As part of the Memorandum of Agreement (MoA) that you have been asked to complete by Mentor UK (see below), you have been asked to provide information about your Year 3 teacher(s) for 2015/16. If you have not done so already, please complete and return the MoA (including Year 3 teacher information) as soon as possible. Our online survey window is currently open, and can be accessed 24 hours a day, 7 days a week. Further information and instructions on how to complete these surveys will be sent to you nearer the time.

4. Application process:

As a reminder of the application requirements for the project, in order to be eligible for randomisation schools need to:

- Send the completed and signed Memorandum of Agreement to Mentor UK as soon as possible (if not done already).
- Make payment relevant to the size of their school (£750 for single form entry school, £1,500 for double form entry and £2,250 for triple form entry) to Mentor, as soon as possible (if not done already).
- Complete the collection of baseline data in the summer term May-July 2015 (at least 90% completion rate is required)

Thank you for your support with the Good Behaviour Game project. Further information can also be found at our project website: www.gbguk.org. If you have any questions please do not hesitate to contact us.

Yours sincerely,

Dr. Alexandra Barlow
Research Associate
0161 275 3504
alexandra.barlow@manchester.ac.uk

Professor Neil Humphrey
Principal Investigator
0161 275 3404
neil.humphrey@manchester.ac.uk

Andy Cawthera
Project Officer
0749 434 5831
andy.cawthera@mentoruk.org
Appendix 10: Data Assumptions

Appendix 10a: Non-Zero Variance

Table A.10.1: Variance scores for disruptive behaviour and reading attainment

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictor Variable</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial Group</td>
<td></td>
<td>0.250</td>
</tr>
<tr>
<td>Behaviour</td>
<td>Baseline behaviour</td>
<td>0.662</td>
</tr>
<tr>
<td>Risk Group</td>
<td></td>
<td>0.331</td>
</tr>
<tr>
<td>Reading</td>
<td>Baseline reading</td>
<td>14.123</td>
</tr>
<tr>
<td>Risk Group</td>
<td></td>
<td>0.607</td>
</tr>
</tbody>
</table>

Appendix 10b: Homoscedasticity

Table A.10.2: Levene statistic for disruptive behaviour and reading attainment

<table>
<thead>
<tr>
<th>Factor</th>
<th>Levene Statistic</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disruptive Behaviour</td>
<td>Risk Group</td>
<td>44.067</td>
<td>14</td>
<td>4136</td>
</tr>
<tr>
<td></td>
<td>Trial Group</td>
<td>4.159</td>
<td>1</td>
<td>2461</td>
</tr>
<tr>
<td>Reading Attainment</td>
<td>Risk Group</td>
<td>5.086</td>
<td>3</td>
<td>2345</td>
</tr>
<tr>
<td></td>
<td>Trial Group</td>
<td>0.783</td>
<td>1</td>
<td>2347</td>
</tr>
<tr>
<td></td>
<td>Baseline Reading</td>
<td>6.531</td>
<td>12</td>
<td>2335</td>
</tr>
</tbody>
</table>

Appendix 10c: Normally Distributed Error

Figure A.10.1: Histogram for disruptive behaviour model
Figure A.10.2: P-P plot for disruptive behaviour model

Figure A.10.3: Histogram for reading attainment model
Appendix 10d: Linearity

Figure A.10.4: P-P plot for reading attainment model

Figure A.10.5: Scatterplot for disruptive behaviour model
Appendix 10e: Random Intercepts

Figure A.10.6: scatterplot for reading attainment model

Figure A.10.7: histogram for disruptive behaviour model
Figure A.10.8: histogram for reading attainment model

![Histogram for reading attainment model](image)
**Appendix 11: RQ2 Complete Case Analyses**

### Behaviour

**Table A.11.1 MLM for differential effects of the GBG on pupils’ disruptive behaviour**

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School-Level (ICC = 16.9%)</strong></td>
<td>0.104</td>
<td>0.020</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>Trial Group (if GBG)</strong></td>
<td>0.019</td>
<td>0.086</td>
<td>.413</td>
</tr>
<tr>
<td><strong>Pupil-Level (ICC = 83.1%)</strong></td>
<td>0.509</td>
<td>0.015</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>T1 TOCA Disruptive Behaviour Score</strong></td>
<td>0.819</td>
<td>0.021</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>Risk Group:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Risk</td>
<td>◯</td>
<td>◯</td>
<td>◯</td>
</tr>
<tr>
<td>Medium Risk</td>
<td>0.129</td>
<td>0.050</td>
<td>.006**</td>
</tr>
<tr>
<td>High Risk</td>
<td>0.306</td>
<td>0.142</td>
<td>.017*</td>
</tr>
<tr>
<td><strong>GBG*Medium Risk</strong></td>
<td>0.089</td>
<td>0.069</td>
<td>.101</td>
</tr>
<tr>
<td><strong>GBG*High Risk</strong></td>
<td>-0.116</td>
<td>0.180</td>
<td>.261</td>
</tr>
</tbody>
</table>

-2*log likelihood = 5472.691

### Reading

**Table A.11.2 MLM for differential effects of the GBG on pupils’ reading attainment**

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School-Level (ICC = 14.3%)</strong></td>
<td>0.063</td>
<td>0.012</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>Trial Group (if GBG)</strong></td>
<td>-0.063</td>
<td>0.089</td>
<td>.241</td>
</tr>
<tr>
<td><strong>Pupil-Level (ICC = 85.7%)</strong></td>
<td>0.373</td>
<td>0.011</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>T1 KS1 Reading Point Score</strong></td>
<td>0.213</td>
<td>0.004</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td><strong>Risk Group:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Risk</td>
<td>◯</td>
<td>◯</td>
<td>◯</td>
</tr>
<tr>
<td>Low Risk</td>
<td>-0.037</td>
<td>0.054</td>
<td>.248</td>
</tr>
<tr>
<td>Medium Risk</td>
<td>-0.037</td>
<td>0.073</td>
<td>.307</td>
</tr>
<tr>
<td>High Risk</td>
<td>0.002</td>
<td>0.093</td>
<td>.491</td>
</tr>
<tr>
<td><strong>GBG*Low Risk</strong></td>
<td>0.136</td>
<td>0.075</td>
<td>.037*</td>
</tr>
<tr>
<td><strong>GBG*Medium Risk</strong></td>
<td>0.072</td>
<td>0.098</td>
<td>.232</td>
</tr>
<tr>
<td><strong>GBG*High Risk</strong></td>
<td>-0.027</td>
<td>0.124</td>
<td>.414</td>
</tr>
</tbody>
</table>

-2*log likelihood = 4484.068
Appendix 12: RQ3 Complete Case Analyses

Behaviour

Table A.12.1 year 3 MLM for implementation and behaviour

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Coefficient</th>
<th>SE</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class-Level</td>
<td>0.138</td>
<td>0.035</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Yr3 FidQual (if high)</td>
<td>-0.034</td>
<td>0.152</td>
<td>.412</td>
</tr>
<tr>
<td>Yr3 Dosage (if high)</td>
<td>0.220</td>
<td>0.152</td>
<td>.077</td>
</tr>
<tr>
<td>Pupil-Level</td>
<td>0.544</td>
<td>0.025</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Behaviour_T1</td>
<td>0.858</td>
<td>0.035</td>
<td>&lt;.001**</td>
</tr>
</tbody>
</table>

Full Model: Disruptive Behaviour Year 3

\[ \beta_{0ij} = -1.521(0.102) \]

-2*log likelihood = 2275.919

Table A.12.2 year 4 MLM for implementation and behaviour

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Coefficient</th>
<th>SE</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class-Level</td>
<td>0.206</td>
<td>0.055</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Yr4 FidQual (if high)</td>
<td>-0.147</td>
<td>0.250</td>
<td>.279</td>
</tr>
<tr>
<td>Yr4 Dosage (if high)</td>
<td>-0.097</td>
<td>0.208</td>
<td>.321</td>
</tr>
<tr>
<td>Pupil-Level</td>
<td>0.526</td>
<td>0.028</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Behaviour_T1</td>
<td>0.973</td>
<td>0.042</td>
<td>&lt;.001**</td>
</tr>
</tbody>
</table>

Full Model: Disruptive Behaviour Year 4

\[ \beta_{0ij} = -1.610(0.124) \]

-2*log likelihood = 1721.754
**Reading**

*Table A.12.3 year 3 MLM for implementation and reading*

**Full Model: Reading Attainment Year 3**

<table>
<thead>
<tr>
<th>Class-Level</th>
<th>Coefficient</th>
<th>SE</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yr3 FidQual (if high)</td>
<td>0.389</td>
<td>0.140</td>
<td>.004**</td>
</tr>
<tr>
<td>Yr3 Dosage (if high)</td>
<td>0.028</td>
<td>0.148</td>
<td>.425</td>
</tr>
<tr>
<td>Pupil-Level</td>
<td>0.339</td>
<td>0.016</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Reading Score_T1</td>
<td>0.208</td>
<td>0.006</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Risk Group:</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yr3 High FidQual*Low Risk</td>
<td>-0.191</td>
<td>0.127</td>
<td>.069</td>
</tr>
<tr>
<td>Yr3 High FidQual*Med Risk</td>
<td>-0.205</td>
<td>0.164</td>
<td>.108</td>
</tr>
<tr>
<td>Yr3 High FidQual*High Risk</td>
<td>-0.128</td>
<td>0.206</td>
<td>.269</td>
</tr>
<tr>
<td>Yr3 High Dosage*Low Risk</td>
<td>0.039</td>
<td>0.139</td>
<td>.390</td>
</tr>
<tr>
<td>Yr3 High Dosage*Med Risk</td>
<td>0.042</td>
<td>0.163</td>
<td>.399</td>
</tr>
<tr>
<td>Yr3 High Dosage*High Risk</td>
<td>-0.347</td>
<td>0.198</td>
<td>.042*</td>
</tr>
</tbody>
</table>

\[-2 \log \text{likelihood} = 1706.664\]

*Table A.12.4 year 4 MLM for implementation and reading*

**Year Four**

<table>
<thead>
<tr>
<th>Class-Level</th>
<th>Coefficient</th>
<th>SE</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yr4 FidQual (if high)</td>
<td>0.489</td>
<td>0.228</td>
<td>.018*</td>
</tr>
<tr>
<td>Yr4 Dosage (if high)</td>
<td>0.301</td>
<td>0.186</td>
<td>.056</td>
</tr>
<tr>
<td>Pupil-Level</td>
<td>0.351</td>
<td>0.019</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Reading Score_T1</td>
<td>0.210</td>
<td>0.007</td>
<td>&lt;.001**</td>
</tr>
<tr>
<td>Risk Group:</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yr4 High FidQual*Low Risk</td>
<td>-0.355</td>
<td>0.206</td>
<td>.045*</td>
</tr>
<tr>
<td>Yr4 High FidQual*Med Risk</td>
<td>-0.261</td>
<td>0.228</td>
<td>.128</td>
</tr>
<tr>
<td>Yr4 High FidQual*High Risk</td>
<td>-0.270</td>
<td>0.316</td>
<td>.196</td>
</tr>
<tr>
<td>Yr4 High Dosage*Low Risk</td>
<td>-0.324</td>
<td>0.160</td>
<td>.024*</td>
</tr>
<tr>
<td>Yr4 High Dosage*Med Risk</td>
<td>-0.205</td>
<td>0.199</td>
<td>.154</td>
</tr>
<tr>
<td>Yr4 High Dosage*High Risk</td>
<td>-0.073</td>
<td>0.236</td>
<td>.379</td>
</tr>
</tbody>
</table>

\[-2 \log \text{likelihood} = 1286.536\]