

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



LSHTM Research Online

Warren, EA; (2022) Are realist randomised controlled trials possible? INCLUSIVE as a case study of an emerging method. PhD (research paper style) thesis, London School of Hygiene & Tropical Medicine. DOI: <https://doi.org/10.17037/PUBS.04668024>

Downloaded from: <https://researchonline.lshtm.ac.uk/id/eprint/4668024/>

DOI: <https://doi.org/10.17037/PUBS.04668024>

Usage Guidelines:

Please refer to usage guidelines at <https://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license. To note, 3rd party material is not necessarily covered under this license: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

<https://researchonline.lshtm.ac.uk>

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



**Are realist randomised controlled trials possible?
INCLUSIVE as a case study of an emerging method**

Emily Ashbrook Warren

**Thesis submitted in accordance with the requirements for the
degree of**

Doctor of Philosophy of the University of London

March 2022

Department of Public Health, Environments and Society

Faculty of Public Health and Policy

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

No funding received

Research group affiliation(s): None

Volume I

I, Emily Ashbrook Warren, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Up to one third of adolescents in the UK are bullied. Bullying and aggression are associated with poorer physical and mental health, and worse social and economic wellbeing across the life-course, making the need for effective interventions imperative. Evaluations of a wide range of interventions have found promising results, but outcomes are often inconsistent across contexts. INCLUSIVE is a school-based cluster-randomised controlled trial (RCT) evaluating Learning Together, a whole-school intervention that integrated restorative practices, social and emotional skills curricula, and student/staff action groups. INCLUSIVE is also the first RCT to be underpinned explicitly by a realist approach aiming to understand how intervention mechanisms interact with context to generate outcomes. This approach has been criticised by some as philosophically incongruent and practically unfeasible. Specific criticisms include the beliefs that 1) trials are inherently positivist; 2) randomisation and control make it impossible to assess context-mechanism-outcome configurations; and 3) trials are insufficiently theorised and concerned only with estimating effect sizes. This mixed-methods study uses data from INCLUSIVE's process and outcome evaluations. Process evaluation data were analysed to understand fidelity, feasibility and acceptability of intervention resources, and how these varied by context. Further analysis of three diverse case-study schools sought to understand participant descriptions of their contexts and how changes were described as occurring as a result of using intervention resources. These were then used to inform theorisation of context-mechanism-outcome configurations which were tested using qualitative comparative analysis. Depending on both their needs and capacities, schools used resources in novel ways to decrease bullying with varying degree of success. Three key social mechanisms for reducing bullying were identified: building commitment to the school community; building healthy relationships by modelling and teaching pro-social skills; and de-escalating bullying and enabling re-integration of perpetrators back into the school community. Analysis suggests that these mechanisms were also activating in some control schools, indicating that they are plausible and potentially transferable. Based on these findings and other analyses of INCLUSIVE data, I reflect on the process of conducting the first realist RCT and conclude that they are philosophically cogent, and can produce nuanced findings about how an intervention works, for whom and under what conditions.

Dedication

This work is dedicated in very grateful memory to my parents,

Douglas and Leslie Warren.

Without your unconditional love, none of this would have been possible.

And P.S. Dad, you were right. All is well.

Acknowledgements:

This work would have been impossible without the generous contributions of many.

Thank you to the study participants who shared their time and stories with me.

Thank you to the peer-reviewers who forced me to clarify my thinking and writing as this thesis developed.

Thank you to my supervisor, Professor Chris Bonell, for teaching me to be a better researcher, guiding me through this process, and encouraging me when I thought this was all a very bad idea.

Thank you also to my supervisor, Professor G.J. Melendez-Torres for being supportive and enthusiastic while you taught me methods I had never heard of before meeting you. Thank you for your belief in, and consistent reminders of, my capabilities.

Thank you to Professor Cécile Knai for investing and believing in me since the beginning of my career. Thank you for always having my back.

To my 219 office-mates who became real-life mates, and Charlie, Sam, and Camille for being such incredible friends. To the Harrison.

Thank you to Matt, Katie, Kristy, Will, Abby, John, Anne, and Theo. I love you all tons.

Thank you to Julie and Wilson Wendt for such immeasurable care and kindness over the years. You have been such founts of love through everything and I could not have been luckier to stumble into your driveway and into your lives.

And finally, my profoundest thanks to Adam. Without your love and indefatigable support, this work would likely have been abandoned long ago. Your encouragement and enthusiasm have been everything. You are my favourite person.

Table of Contents

Abstract.....	7
Dedication	9
Acknowledgements:	11
Table of Contents.....	13
List of abbreviations:	16
INTRODUCTION.....	19
Chapter 1: Bullying as a public health problem and interventions used to prevent it	22
Bullying definitions and roles	22
Prevalence and risk factors.....	24
Person-related risk factors	24
Familial and socio-economic risk factors.....	26
School-related risk factors.....	26
Consequences.....	29
Short-term consequences.....	29
Long-term consequences.....	31
Costs associated with bullying perpetration and victimisation	32
Bullying prevention interventions	33
Evidence of effectiveness of bullying prevention interventions	33
Classroom interventions including SEL interventions.....	35
Restorative practice	37
Whole-school interventions.....	41
Limitations in evidence base and need for more useful evaluation evidence	45
Description and structure of this thesis	48
Chapter 2: Approaches to evaluating complex interventions.....	49
Complex interventions.....	49
Theories of change and logic models.....	49
Randomised controlled trials.....	51
Design.....	51
Process evaluation	52
Origins and developments	52
Realist approaches to evaluation	54
Critiques and limitations of conventional evaluation	57
Philosophical concerns:.....	58
Positivism and successionism	58
Practical concerns:.....	59
Randomisation and control stifle one’s ability to understand how change happens.....	59
RCTs are insufficiently theorised.....	60
RCTs are unable to assess mechanisms	60
RCTs are only concerned with attribution.....	61
Realist trials	61

Critiques of realist RCTs	62
Philosophical concerns:	63
Positivism, successionism, and generativism	63
Practical Concerns:	63
Randomisation and controls stifle the ability to test CMOCs:	63
Defence of RCTs and realist RCTs.....	64
Philosophical concerns:	64
Practical Concerns:	67
RCTs being insufficiently theorised is not an inherent shortcoming of the study design	68
RCTs can be used to assess mechanisms	69
RCTs are not only concerned with attribution.....	70
RCTs do not stifle one’s ability to test CMOCs.....	70
Case study to assess feasibility, intellectual coherence, and usefulness of a realist trial	71
Aims, objectives, and research questions	72
CHAPTER 3: METHODS.....	74
Mid-range theory underpinning Learning Together and INCLUSIVE.....	74
Learning Together’s theory of change and logic model	77
Description of Learning Together	80
Action groups.....	80
Training and Restorative Practices.....	81
SEL curriculum.....	81
INCLUSIVE	81
Comparator arm: Control Schools.....	84
Process evaluation	84
Summary of results from INCLUSIVE trial.....	84
My contribution to the study team as a research fellow.....	85
Thesis Methods	85
Key distinctions between the trial and this thesis.....	85
Ontological and epistemological assumptions	86
Description of the methods used in each chapter	93
Using realist concepts to assess the fidelity, feasibility, and acceptability of action groups as a participative strategy for whole-school health promotion interventions-- methods for Chapter 4 ...	93
Using dimensional analysis to understand participant accounts of contexts, mechanisms, and outcomes— methods for Chapter 5.....	99
Testing CMOCs using qualitative comparative analysis-methods for chapter 6.....	105
Brief overview of QCA	106
Building data tables.....	108
Constructing truth tables.....	109
Boolean minimisation	110
Reflection on whether realist RCTs are possible—methods for chapter 7	110
Ethics.....	111
Reflexivity and Positionality	111
FINDINGS.....	115
Introduction to the first study:.....	115

<i>Chapter 4: Action Groups as a participative strategy for leading whole-school health promotion: results on implementation from the INCLUSIVE trial in English secondary schools</i>	117
Introduction to the second study:	142
<i>Chapter 5: Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention</i>	146
Introduction to the third study:	162
<i>Chapter 6: Using fuzzy-set qualitative comparative analysis (fsQCA) to explore the causal pathways to reduced bullying victimization in a whole-school intervention: results from a cluster randomized controlled trial</i>	167
<i>DISCUSSION</i>	194
Introduction to the Discussion chapter:	194
<i>Are realist randomised controlled trials possible? A reflection on the INCLUSIVE evaluation of a whole-school, bullying-prevention intervention</i>	197
Conclusion	209
Implications for practice, policy, and research:	209
<i>References:</i>	214
<i>Complete references</i>	231

List of abbreviations:

AAYP- Aban Aya Youth Project

AG-Action group

AGM-Action group meeting

ALSPAC-Avon Longitudinal Study of Parents and Children

AOR-adjusted odds ratio

CATCH- Child and Adolescent Trial for Cardio-Vascular Health

CB-Chris Bonell

CCTV closed-circuit television

CI-Confidence interval

CMOC-context, mechanism, outcome configuration

csQCA- crisp set qualitative comparative analysis

CTU-Clinical Trials Unit

CVD-Cardiovascular disease

ES-Effect Size

ESYTC – Edinburgh Study of Youth Transitions and Crime

EW- Emily Warren

FSFF-Friendly Schools Friendly Family

FSM-Free school meals

fsQCA- fuzzy set qualitative comparative analysis

GBS – Gatethouse Bullying Scale

GJMT-G.J. Melendez-Torres

GSMS- Great Smoky Mountain Study

INCLUSIVE-Initiating Change Locally in bullying and aggression through the School Environment

ITT-Intention to treat

LSHTM-London School of Hygiene & Tropical Medicine

LT-Learning Together

MRC-Medical Research Council

NAR-Needs Assessment Report
NHS-National Health Service
NIHR- National Institute for Health Research
OBPP-Olweus Bullying Prevention Programme
Ofsted-Office for Standards in Education
OR-Odds ratio
OTA-On-treatment analysis
PE-Process evaluation
PSHE-personal, social, and health, education
QCA- qualitative comparative analysis
QoL- Quality of life
RAiS-Restorative Approaches in Schools
RCT-Randomised controlled trial
RIPPLE-Randomised Intervention trial of PuPil-Led sex Education
SAFE- Sequenced, Active, Focused, and Explicit
RP-Restorative practice
SDQ- Strengths and Difficulties Questionnaire
SES-Socio-economic status
SEL-Social and emotional learning
SLT-Senior leadership team
SWEWBS – Short Warwick-Edinburgh Well-Being Scale
TMSA-Transformational Model of Social Activity
VA-Value-added
WSI-Whole-school intervention

INTRODUCTION

Randomised controlled trials (RCTs) are considered by many to be the gold-standard study design for assessing causation. Randomisation minimises the risk of selection bias and distributes potential confounders with approximate evenness between trial arms. The control group acts as a counterfactual to support the estimation of an intervention's effect. These effect sizes can help policy-makers and other stakeholders make informed decisions about whether an intervention can potentially generate certain outcomes in a given place or population. Despite these methodological strengths, the potential of RCTs is not always maximised in practice to understand crucial questions about how interventions are being used and by whom, what mechanisms may be generating the observed changes in outcomes, and how those vary by context and population.^[1] In 2012, Bonell et al proposed realist trials as a way to maintain the rigour of trials while asking more nuanced questions about what works, for whom, under what conditions and how, which are traditionally the foci of realist evaluations.^[2] Some realists have argued against RCTs on the premise that they are inextricably linked to positivism, that they are unfit to explore questions about the role of context on the emergence of outcomes in complex, open, social settings where laboratory controls are unfeasible, and that trialists view constant conjunctions as sufficient for understanding causation.^[1, 3, 4]

The aim of this thesis is, for the first time, to analyse outcome and process evaluation (PE) data from an RCT to explore if realist trials are philosophically coherent, practically feasible and generate useful findings. This thesis uses data from the Initiating Change Locally in bullying and aggression through the school environment (INCLUSIVE) evaluation of the Learning Together (LT) intervention.^[5, 6] Building from the theory of human functioning and school organisation,^[7] and the study's associated theory of change, I used mixed-methods approaches to assess if it is possible to maintain a realist ontology and employ methods consistent with realist enquiry to assess whether and how LT resources were used, whether this usage triggered reduced rates of bullying and aggression, and how these varied by population and place.

In the introductory chapters of this thesis, I first review the literature on the prevalence, risk factors for, and short and long-term consequences of bullying perpetration and victimisation. I then assess the evidence of prevention interventions, focusing on social

and emotional learning (SEL), restorative practices (RPs) and whole-school interventions (WSIs), highlighting the significant limitations in our knowledge. In Chapter 2, I describe, compare and contrast RCTs and realist evaluation approaches, and present the concept of realist RCTs and the ensuing debate. In Chapter 3, I describe LT and the INCLSUIVE evaluation before explaining the methods used in each empirical chapter and reflect on my own positionality and how this affected data collection and analysis.

The Findings section contains three publications. Chapter 4 explores the implementation, feasibility and acceptability of LT, focusing on the action group (AG) component that was responsible for coordinating the intervention more broadly.^[8] Understanding implementation and how this varied in relation to schools' capacities and priorities was vital as this provides the first empirical indication of which mechanisms might later be activated in which contexts. In Chapter 5, I focus on the qualitative data from three purposively contrasting case-study school contexts. Using a variant of grounded theory called dimensional analysis, I distil participant descriptions of the mechanisms that appeared to activate as a result of participants engaging with LT resources, and how those varied by context.^[9] These mechanisms to reduced bullying included 1) building commitment to school, 2) building healthy relationships and modelling pro-social skills, and 3) de-escalating bullying among a core group of students. This analysis was used to inform context-mechanism-outcome configurations (CMOCs) theorising how the activation of mechanisms was contingent on the presence of key contextual features. In the final empirical analysis, I employ fuzzy-set qualitative comparative analysis (fsQCA) to test the aforementioned CMOCs to assess under what conditions the aforementioned mechanisms appeared to operate and whether evidence from control schools also suggested the activation of the same mechanisms via other, non-LT resources.^[10]

In the Discussion section^[11] (Chapter 7), I synthesise the work contained in this thesis, as well as the studies led by other team members, and reflect on when and to what extent analyses of trial data were aligned with realist enquiry and whether or not the methods we employed were useful for answering questions central to realist evaluation. I then respond to critics' concerns about philosophical compatibility between trials and

realism, and conclude with the limitations we faced and how realist trials can be improved in the future.

Chapter 1: Bullying as a public health problem and interventions used to prevent it

Bullying definitions and roles

There is no single agreed definition of bullying but key features include intentionally hurtful, repetitive, aggressive behaviours targeted at those with lesser power.[12-15] Bullying commonly manifests in mistreatment that is physical (in the form of hitting, kicking, punching, or taking or damaging people's belongings), verbal (using offensive terminology, name-calling, insults or mockery), relational (intentionally excluding or isolating someone), indirect (spreading rumours) or cyber (using the internet to bully someone).[16]

Bullying most commonly occurs at school[17, 18] but it increasingly happens in cyberspace,[12, 19] via phone calls, text message, emails, chats, or social media platforms.[16] Like in-person bullying, cyberbullying can take multiple forms including harassment, cyberstalking, outing or exclusion.[16] Distinct features of cyberbullying include the aggressor's ability to exert dominance through their anonymity,[19-21] that it can occur 24-hours a day and it may reach a larger audience.[22] In this thesis, in-person and cyber-bullying are considered together because of their shared risk factors and consequences.

Historically, bullying researchers have classified students into one of three roles: bully, victim or onlooker/bystander.[23] Bullies are understood to have a strong will to power and exert it by demeaning victims, and eroding their sense of security and safety while increasing their own sense of high social-status and self-esteem.[23] Bullies can establish dominance using both anti- and pro-social means.[24] Bullies may be aggressive to socially vulnerable peers while also being charismatic and friendly with peers who can award them higher social-standing. Because of its perceived potential to improve status, in-person bullying generally occurs when bystanders are present to observe it. They in turn can choose to defend the victim, support the bully or, most commonly, remain passive. [16, 23]

More recently, researchers have focused on how bullying is not a unidirectional phenomenon and participants' roles are not mutually exclusive: in the literature, people who are simultaneously victims and perpetrators are called 'bully/victims.' Bully/victims have been described as "the most troubled: impulsive, easily provoked, low in self-esteem, poor at understanding social cues, and unpopular with their peers." [25, pg. 1959] Bully/victims may emerge as the result of social learning in which they see others' use of aggression against them as an effective tool for dominance and therefore seek to deploy the same tactics against others.[26] Bully/victims have been found to exhibit the most aggressive behaviour and, despite being the victims of bullying, also reported the most accepting view of bullying behaviours.[27, 28] Perpetrating in-person bullying is a significant risk factor for later cyberbullying, and perpetuating cyberbullying is a risk factor for later in-person bullying.[21, 22, 29, 30]

Multiple researchers have used Bronfenbrenner's ecological model of human development[31-33] to parse which factors influence bullying as a social behaviour. [34, 35] Using the theory, researchers can examine: microsystems, such as family and the school environment; mesosystems, which include the relationships between microsystems (e.g. the connection between families and schools); the exosystem, which includes outside actors influencing the school, such as the Office for Standards in Education (Ofsted) in England; the macrosystem, which includes social beliefs and cultural values; and the chronosystem, or where we are in broader social history and how that impacts on bullying.

It is important to note early in this thesis that young people are referred to as bullies, aggressors, perpetrators, victims, bully/victims or bystanders. These terms are linguistic shortcuts to avoid the cumbersome terminology of, for example, "children who engage in bullying behaviours." I acknowledge that the children and adolescents described in this thesis are still undergoing social, emotional, behavioural, neurocognitive and physical development, and these behaviours do not encapsulate who they are or imply that their role is somehow fixed or predetermined.

Prevalence and risk factors

In the UK, prevalence estimates of bullying experienced in the previous three months vary from 10% of girls and 12% of boys[36] to one third of all young people,[37] highlighting the extent of the problem. In the US, prevalence of bullying either as a victim or perpetrator range from 20.8% for physical bullying to 53.6% for relational bullying.[24, 26] In a 2020 systematic review of longitudinal studies of in-person and cyberbullying, prevalence of cyberbullying perpetration varied widely between 5.3 and 66.2% (median 11.7%) and victimisation between 1.9 and 84% (median 14.4%). Rates may vary so widely partly due to the use of different assessment tools and recall-time periods.[22]

Person-related risk factors

One's personality, social-standing, and mental well-being affect the risk of being either a perpetrator or victim. Agreeableness has been found to be a negative predictor of in-person bullying perpetration while extraversion, Machiavellianism, psychopathy and sadism were all positive predictors.[38] In a systematic review of longitudinal studies, depression and social anxiety at baseline were highly predictive of victimisation at a later time point, and poor emotional regulation and anger were predictive of future perpetration.[22] Hyperactivity increases the risk of both perpetration and victimisation. Hyperactivity is associated with impulsiveness, difficulty processing social information and perceiving neutral social signs as hostile, and people with hyperactivity may therefore react more aggressively to social situations.[26] Hyperactive students are more likely to be judged by their peers as "annoying" which researchers postulate may increase their risk of becoming a target for bullying.[21, 39]

Longitudinal analysis suggests a U-shaped relationship between popularity and cyberbullying perpetration, with the least and most popular children having the highest risk, although authors speculate that the motivations for bullying between those two groups may differ.[40] In addition to how many friends a student has, the social skills within the peer group affect the risk of perpetration. One longitudinal study found that having fewer peers with prosocial skills was associated with a 50% increase in the risk of bullying perpetration.[21] Researchers proposed that prosocial peers may model

acceptable behaviour; their absence may mean other students are not proscribed when they begin to engage in bullying.[21]

Evidence continues to grow about the relationship between substance use and bullying. Substance use (including alcohol and marijuana) appears to increase the risk of perpetration, with the greater the variety and frequency of substances used being associated with greater risk for bullying perpetration.[41, 42] Evidence also indicates that bullies and bully/victims have the highest rate of substance use compared to victims and uninvolved peers,[43] although whether this is a risk-factor or a consequence is not yet fully disentangled. Researchers postulate that both substance use and bullying are part of a larger pattern of anti-social behaviours with shared risk factors and outcomes, and are therefore useful to study as part of a complex behavioural web.

Gender and age also affect one's risk of engagement in bullying. Evidence from systematic reviews, meta-analyses, cross-sectional and cohort studies are conclusive that boys are more likely to be involved in bullying, either as a victim or perpetrator.[21, 26, 44-46] A peer nomination study in the US found that boys were more than three times as likely as girls to be bully/victims and twice as likely to be victimised.[45] Children's descriptions of the types of bullying they experience also change with age. Younger children report higher rates of hitting and insulting while adolescents report greater social exclusion and rumours being spread about them.[47] In-person bullying rates tend to be highest in primary school and decrease through secondary school.[48-51] A four-wave panel study found that cyberbullying grows through late childhood, peaking at approximately age 14 and then decreasing throughout the remaining teen years.[52] A meta-analysis from the USA suggests that bullying connected to one's sexual relationships, sexual competition, or perceived and actual sexual orientation peak around 15 years old.[47]

Individual-level assets that protect young people from bullying include strong pro-social and problem-solving skills and high self-esteem. Personal characteristics that appear protective against perpetration include high status amongst peers, high degrees

of openness, agreeableness, empathy and conscientiousness, and low degrees of extraversion.[38, 53]

Familial and socio-economic risk factors

A child's family, home-life and broader socio-economic context affects the risk of bullying involvement. Commonly accepted risk factors for bullying perpetration include high parent-child conflict, the use of physical discipline, punitive parenting, child abuse and maltreatment and domestic violence.[54-56] Children of parents who express the acceptability of violence or aggression and negative views about school are also more likely to bully others.[55] In a literature review on the role of family factors on bullying, 77% and 75% of included studies found a positive association between poor parental mental health and increased risk of perpetration or victimisation, respectively.[55] Researchers have postulated that growing up in a high-conflict or neglectful environment either teaches children or fails to correct behaviours which are then enacted in school, where those behaviours are judged to be inappropriate or unacceptable.[24, 55, 57] Parental overprotectiveness, including low child autonomy, is associated with increased risk of victimisation.[55]

Conversely, positive parental interactions and being in a family that is supportive and has warm communication are protective against bullying victimisation.[53] Likewise, having a higher socio-economic status, attending a school and living in a neighbourhood that students view positively are protective against both in-person and cyber-bullying.[53] Children from dual-parented families are at a lower risk of involvement with bullying than children from single-parented families.[36, 46, 58, 59]

School-related risk factors

The prevalence of bullying varies dramatically between schools, even after adjusting for students' socio-demographic characteristics and prior misbehaviour, indicating that the school environment itself may affect bullying. In one review of multilevel studies, Sellström found the presence of school-wide health and anti-smoking policies, positive school climate, above average socioeconomic status and being in an urban area were all associated with decreased behavioural problems, including bullying.[60] Another multilevel study of schools drawing on data from the INCLUSIVE trial found that school

type and quality as rated by the Office for Standards in Education, Children's Services and Skills (Ofsted) both impacted upon levels of bullying perpetration and victimisation.[59] Where students feel unsafe or marginalised, they may form groups and establish a sense of safety in numbers via bonds created around shared risk behaviours, including displays of violence and aggression.[61, 62] These behaviours function as important identity-markers and status symbols when pro-school markers are perceived as unobtainable.[63, 64] Jamal et al found evidence of peer social instruction and regulation concerning behaviours that students use to facilitate feelings of safety through "tough fronts." [62] Students who signal their high social status via violent behaviours, including bullying, may feel protected from harassment in otherwise chaotic environments.[62, 64, 65]

Conversely, a strong connection to school may positively affect young people's mental and physical health, including decreased emotional distress, suicidality, violence and partner violence,[66-70] all of which may be protective against bullying perpetration and victimisation. In schools where students have supportive teacher relationships, students also report more prosocial behaviour, higher achievement, greater well-being and increased interest in school.[60, 71-77] As part of her systematic review, Johnson created a construct of teacher/student relationships and found that the more positive the relationships were, the lesser the risk of violence and bullying. In four of the six studies measuring teacher support, it was significantly associated with decreased victimisation and perpetration of bullying and violence.[78] Building a positive social climate, (commonly understood to include perceived fairness, parental involvement, and strong student-student and student-teacher relationships) has been shown to decrease cybervictimisation.[22] Moreover, two systematic reviews on the impact of the school environment on violence found that when students think rules are fair or are engaged in making them, there is less fighting.[78, 79] Schools with high value-added (VA) scores (denoting the extent to which students perform better on standardised assessments than expected as predicted on baseline attainment[80-82]) have lower rates of fighting than schools with low VA scores.[79, 82, 83]

Thus far, the evidence presented on school environment-based risk factors and bullying has been observational. However, two RCTs evaluating the Aban Aya Youth Project

(AAYP) and the Gatehouse Project, have provided experimental evidence on how the school environment may impact upon a range of health and well-being outcomes, including bullying and aggression. AAYP was a three-arm, school-based cluster-RCT comparing control schools, receiving a new general health curriculum, to schools receiving only a new social and emotional learning (SEL) curriculum and to schools teaching the new SEL curriculum and carrying out a school/community intervention.[\[67\]](#) In the third arm, AAYP implemented a community task force of school staff, students, parents and members of the community surrounding the school to improve the school climate, rules and policies, and build community support for school improvement efforts. These changes were hypothesised to improve factors influencing students' health and improve the conditions in which they live and study. AAYP was implemented in Chicago schools primarily serving African American students. The trial found no significant improvement for girls in any of their six outcomes. For boys, however, the arm that included both the curriculum and the community support significantly reduced violent and provoking behaviour by 47% and 59% respectively compared to the control arm,[\[67\]](#) providing evidence of benefits for schools and communities addressing multiple risk factors simultaneously.

The Gatehouse Project was a complex intervention that sought to improve inclusion and connection to secondary schools in Australia.[\[84\]](#) Researchers theorised that the intervention would encourage participation in school activities, foster feelings of security and trust, increase communication skills, and provide more opportunities for building healthy relationships. By changing the school environment, researchers aimed to promote feelings of safety, warm relationships and good communication between students and staff.[\[84\]](#) Researchers provided each school with staff training, a new SEL curriculum, a report with the amalgamated results from their students' surveys and the resources to recruit someone to serve as a 'critical friend.'[\[84-86\]](#) The 'critical friend' was meant to be a supportive, objective outsider providing advice, guiding reflection and encouraging student participation in developing context-specific strategies to improve implementation.[\[85\]](#) The outcome evaluation found lower odds of antisocial behaviour (AOR 0.78; CI 0.57-1.07) indicating a trend, but no significant difference.[\[87\]](#)

Consequences

Health behaviours established in adolescence can become entrenched in adulthood and have significant consequences throughout the life-course.[36, 68, 88, 89] Therefore, preventing bullying in childhood and adolescence may have profound consequences on the needs of health and social interventions in adults.[25, 69, 90, 91]

Short-term consequences

In the short-term, bullying victimisation is associated with increased somatic complaints, [20, 36, 92] substance use,[92, 93] feeling insecure and lonely,[45, 92] and psychological distress[36, 92] including depression, social phobias, post-traumatic stress disorder, self-harm and suicide.[92, 94, 95] A meta-analysis found a three-fold increase of suicide attempts amongst occasionally bullied students and a four-fold increase for students who described their victimisation as frequent.[92] Teachers are also more likely to report sadness and anxiety in the students they believe are being bullied compared to bullies or un-involved peers.[45] However, caution should be exercised when assuming victimisation causes poor mental health as poor mental health is also a risk factor for victimisation. Understanding the relationship between the two phenomena is challenging because few studies control for pre-existing mental health problems.[25, 96] However, sub-group analyses of longitudinal studies that controlled for pre-existing mental health problems[94, 97] indicate that victimisation is associated with elevated risk of poor mental health, such as non-suicidal self-injury.[92]

Cyberbullying victims are more likely than victims of in-person bullying to report general and social anxiety, loneliness, and depression.[21, 22] Longitudinal studies have reported that cyberbully victims were significantly more likely than their peers to report feelings of mistrust of others, sense of personal defectiveness, negative body image and anger.[22, 98-100] Moreover, after adjusting for experiences of in-person bullying, victims of cyberbullying reported more anxiety, depression, suicidal ideation and an externalised locus of control.[101] Over time, students who are routinely victimised by cyberbullying may experience both higher degrees of peer-rejection and lower levels of support from parents and friends,[102] increasing their isolation and distress.

The poorer health outcomes associated with bullying are not confined to the victims. Perpetration of cyberbullying is associated with hyperactivity, conduct problems and somatic complaints.[21] Perpetrators report worse school experiences,[36] increased alcohol and substance use,[21, 103] poor mental health including self-harm, and are more likely to carry a weapon and engage in violence.[22, 45, 103-105] While teachers report that bullies experience higher status than non-bullying peers, they are also reportedly avoided by their peers who fear them.[45]

Cross-sectional and repeated cross-sectional studies with middle and high-school American students found that victims, bullies, and in particular bully/victims, reported greater use of alcohol or marijuana.[106, 107] In the UK, Fletcher et al analysed interviews with 30 students from diverse backgrounds and with varying level of school engagement and proposed that substance use may be an escapist strategy to cope with unhappiness and social isolation.[108] As shown earlier, as bullies and their victims often lack sufficient social support, substance use may be a mechanism through which the consequences of bullying are handled.[43, 108]

In a cross-sectional study, one quarter of students reported being cyber bully/victims.[109] A large cross-sectional study from the USA found that bully/victims had worse psychosocial, psychological, and behavioural outcomes than peers.[27] Bully/victims are at increased risk of depression, and have lower perceptions of their scholastic potential and self-worth,[27] and feel more excluded than their non-bullying and non-victimised peers.[110, 111] Studies exploring bullying rarely include biomarkers. However, one study collected data on the secretion of the stress hormone, cortisol, over two days and found that bullies, victims, and bully/victims all had significantly greater elevation of cortisol compared to cyber-bystanders. Bully/victims and, to a lesser but still significant degree, victimised-only students had the highest increases in stress, suggesting the negative impact that bullying can have on all involved.[112]

Young people who report being victims of bullying are at greater risk of being the victims of dating violence, and those who bully others are more likely to report using physical violence and social aggression against their partners.[70] A recent meta-

analysis found that, after adjusting for covariates, bullying perpetration increased the odds of dating violence (AOR 1.29). Similarly, bullying victimisation was associated with increasing odds of dating violence victimisation (AOR=1.96).[\[113\]](#) Most papers included in this review were from North America and none was from the UK, so generalisability to the UK context may be limited.

In the UK, there is growing concern that low-level aggression and bullying are harmful even to uninvolved students through the increasing acceptance of violent norms.[\[91, 114\]](#) Bullying may also lessen academic attainment through increased classroom disruption and increased absenteeism, especially as victimisation worsens.[\[21, 115-117\]](#)

Long-term consequences

It is not yet clear whether outcomes associated with or caused by bullying are worse for younger or older children. Some researchers hypothesise that younger children are more negatively affected by bullying because they have not yet developed the requisite social skills to cope with the stress.[\[118\]](#) Conversely, some researchers have postulated that the impact of bullying may be worse for older children because either the health impacts or the negative social consequences build over time, leading to worse outcomes after prolonged periods of victimisation. Likewise, older children have an increased tendency to internalise problems compared to younger children.[\[96\]](#) Finally, a six-phase repeated cross sectional study documented that, over time, the support from peers, teachers and even parents weakens for long-term victims which may also contribute the burden of poor mental health.[\[96\]](#)

For many, the consequences of bullying can persist into adulthood.[\[119-122\]](#) Analysis of the Great Smoky Mountain Study (GSMS) data from the US and the Avon Longitudinal Study of Parents and Children (ALSPAC) in the UK showed that, after adjusting for confounders, adults who were bullied as children had worse mental health outcomes (including anxiety, depression and self-harm) than those who had suffered from childhood maltreatment (including physical, emotional or sexual abuse or severe maladaptive parenting),[\[122\]](#) highlighting the seriousness and longevity of mental health consequences of bullying. Violence and aggression, including bullying, are associated with mental ill-health and anti-social behaviour in adulthood.[\[55, 104, 105,](#)

[116](#), [123](#)] There is some conceptual uncertainty about the role that bullying plays in adult mental-illness, and whether other shared casual factors, such as disadvantaged upbringing or childhood psychiatric illness, might partly account for the long-term consequences associated with bullying perpetration. In an analysis of longitudinal data from the GSMS, researchers assessed the unique contribution of being a bully, being bullied, or being a bully/victim over and above adverse family relations and pre- and co-existing psychiatric problems in childhood.[\[25\]](#) Unadjusted analyses showed worse outcomes for bully/victims across health, wealth and social wellbeing domains, and bullies had the worst outcomes for risky/illegal behaviours. However, after adjusting for childhood family hardship and child psychiatric problems, the risk of worsened health outcomes for bullies across all domains was substantially attenuated. While adjustments attenuated the risks, being a bully/victim continued to be an independent predictor of worse outcomes across health, wealth and social domains. Finally, a dose-response relationship for outcomes relating to wealth and social well-being was identified between young adults who had experienced one instance of bullying compared to chronic victims.[\[25\]](#) This study is important because it provides evidence about the long-term effects of being bullied or being a bully/victim while also showing that bullying behaviour may be a consequence of earlier social processes or part of a larger tendency towards anti-social behaviours.

Costs associated with bullying perpetration and victimisation

Bullying is costly both to families and governments. A cross-sectional cost of illness analysis was conducted using data from Germany. Drawing on direct healthcare costs (including outpatient care for injury and illness and psychotherapy), direct non-medical costs (travel costs and time for parents,) and indirect costs of parents not being at work, costs were significantly higher for frequently bullied students compared to non-bullied controls ($p=0.008$, cost difference= €5,323.01).[\[124\]](#)

The social and economic impacts of bullying can last decades.[\[25, 125\]](#) Researchers used the UK's National Child Development Cohort survey data from children born in 1958 to assess the long-term impact of bullying. In 2008 outcomes including employment, earnings, housing tenure, savings, health service use for mental health conditions and mean annual societal employment costs were assessed. After adjusting for childhood

confounders including childhood IQ, emotional and behavioural problems, family social class in childhood and childhood adversity, researchers found that, at age 50, women who were frequently bullied as children had worse economic outcomes in every domain and women who were occasionally bullied had worse outcomes in all domains except mean weekly earnings from paid employment. At a societal level, being frequently bullied was associated with an average £717 higher mental health care cost over an eight-year mid-life time point. When extrapolated to a social level, those costs represent estimated costs of £4.5 million annually. Men who were frequently bullied were significantly more likely to be unemployed, not be home-owners and have no to low levels of savings. At a social level, employment-related social costs were £271 annually, representing a national annual costs of £17.9 million.[46] Authors hypothesised that lower self-esteem, altered physiological responses to stress and poorer mental health caused by bullying may have impacted outcomes. These economic impacts, seen four decades later, represent a meaningful prevention opportunity.[46]

Bullying prevention interventions

Evidence of effectiveness of bullying prevention interventions

Due to its high prevalence and the seriousness of its immediate and long-term effects, reducing bullying is a public health and educational priority.[126, 127] However, how schools respond to and manage bullying is inconsistently evidence-based and will vary widely between schools.[91] Systematic reviews and systematic reviews of reviews have shown that school-based interventions can reduce bullying.[80, 83, 120, 128-133] A recent meta-analysis of 100 studies by Gaffney et al found that both perpetration and victimisation were significantly reduced following the implementation of anti-bullying interventions (p values <0.001 for both).[15] This drop in bullying represents a 19-20% decrease in perpetration and a 15-16% decrease in victimisation. This is similar to the earlier meta-analytical findings by Ttofi et al who found an overall decrease of 17-23% in bullying (odds ratio (OR) 1.36, mean difference (d) 0.17) and victimisation (OR 1.29, d=0.14)).[132] While these effect sizes may be considered small,[132, 134] when extrapolated to the population, they would substantially reduce the burden of bullying-related ill-health both in the short and long-term.

One finding repeated in the literature is that interventions work better when they are implemented well. Prospectively monitoring intervention implementation is associated with lower rates of self-reported victimisation across multiple studies.[\[128, 131, 132, 135\]](#) In their meta-analysis, Gaffney et al reported that, when implementation was monitored, perpetration and victimisation effect sizes were doubled compared to when implementation was not monitored.[\[15\]](#)

Despite the many meta-analyses of anti-bullying interventions, meaningful gaps in the evidence remain. In Ng et al's 2020 meta-analysis of 17 education-based interventions, authors explain that the field of bullying research contains so many contradictions because seminal reviews and studies may now be outdated, age ranges vary widely and may not focus on adolescents, and evidence may focus on children who are more likely to be targeted for bullying (such as lesbian, gay, transgender or queer children or children with disabilities).[\[136\]](#) The various foci, inclusion and exclusion criteria, and outcome measures have created a field in which no clear picture of 'what works' emerges from the data.

There is strong evidence that effects vary between studies and study contexts. In another recent review, Gaffney et al explore how study-level variables, such as country and continent, evaluation design and unit of randomisation moderate intervention effectiveness.[\[137\]](#) While it may be useful to know that, for example, the biggest effects were from studies conducted in Greece, the review did not explore which approaches (targeted or universal, educational vs whole-school) were most effective, why effects differ so widely between contexts and through what mechanisms bullying is reduced.

Despite the contradictions across secondary analyses, a number of intervention types are potentially well placed to address and prevent bullying. These include classroom-based social and emotion learning (SEL) curricula, restorative practices (RP) and whole-school interventions (WSI). These types of interventions have clear theoretical underpinning, have plausible mechanisms of action and may build students' life-skills (such as good communication) in a way that positively impacts them beyond preventing bullying. Each of the aforementioned intervention types and the evidence of their effectiveness are described in more detail below.

Classroom interventions including SEL interventions

Historically, schools have tried to reduce bullying via consequences and punishments. Recently, however, increasing attention has been given to prevention interventions which teach social skills to build a positive school climate[34] and develop key competencies to reduce risk factors and bolster protective assets that promote success in school and life more broadly.[128] These competencies include self-awareness (recognising strengths and weakness, and being able to identify feelings), social awareness (being able to empathise), self-management (regulating emotions and behaviours), relationship skills and responsible decision-making.[138] Deficits in such competencies are associated with bullying behaviours and are therefore important to address.[139] Divecha and Brackett summarise key theories that have been used in SEL interventions.[34] These include information dissemination theory, which like the theory of reasoned action, assumes that if students know that there will be negative consequences arising from a behaviour, they will avoid engagement.[135] Other theories include social cognitive processing, which postulates that children need to learn to assess social situations to be able to respond appropriately. Critics have argued that bullies may be highly skilled socially and are thus able to manipulate those around them.[140] Incorporating this concern, SEL interventions may also be underpinned by psychodynamic theories in which bullying is viewed as a maladaptive response to the increasing agency experienced by children as they mature.[141] Divecha and Brackett argue for the increased use of a bioecological perspective in SEL education.[34] Bioecological theory explores the constant and dynamic interplay between individuals, their contexts, and time, and reflects on how these change as development occurs. By matching SEL programmes to a child's development and addressing the factors that shape their development, interventions could be more strategically placed to have the greatest impact.[32] Despite the existence of many potentially useful theories, many interventions are not underpinned by a theory, which may explain why many interventions are ineffective.[131, 132, 142]

To be effective at the school and population level, SEL programmes must offer resources that can activate various mechanisms for different people. Despite its popularity, bystander training is unlikely to be effective for people who are inhibited or shy, and so

SEL programmes would therefore need to provide students with a wider variety of training and social skills to reduce bullying.[34] SEL programmes also need to be appropriate to the age of the students. Systematic reviews and evidence from RCTs and panel studies have shown that interventions that work well with young children do not always translate to older populations where differences in brain development and social priorities affect the types of messages and skills that may be effective for preventing bullying.[34, 47, 136] For example, prescriptive programmes may encourage disengagement by infantilising older students who want to develop their own sense of agency and independence.[34]

In 2014, Public Health England recommended SEL to improve academic attainment and reduce bullying and aggression.[143] Despite this enthusiasm, the evidence remains inconclusive. Published in 2018, a meta-analysis of interventions to improve social skills to reduce bullying identified six studies, one of which was effective, three of which produced non-significant findings and two of which were beneficial in at least one subgroup analyses, although one study also identified potential harms for younger participants.[139] In a systematic review of school-based interventions to prevent bullying, Vreeman and Carroll identified ten curricular interventions with a wide range of intensity, all utilising pre-test/post-test study designs, six of which were randomised. Six studies found no statistically significant improvement, and of the four studies that found a positive effect overall, three showed worsening bullying within certain populations or by using different scales.[120] Null or non-significant results[142, 144] and iatrogenic harms have also been documented.[145] It is important to note that the quality and length of studied curricula likely affected findings. The authors did not describe the details and objectives of the curricula, making it difficult to assess the extent to which interventions were underpinned by SEL principles.

Another meta-analysis of 213 trials found decreased risk of conduct problems, including bullying (effect size (ES) (Hedges' g) = 0.22, $p = \leq 0.05$, CI 0.16-0.29) between students receiving an SEL intervention and those in control arms. Effects were moderated by implementation fidelity and whether or not the intervention was sequenced, active, focused and explicit (SAFE).[128] Sequence refers to whether the programme had a coordinated set of activities to improve students' skills. For programmes to be

considered “active”, students had to engage in participatory learning to build new skills (e.g. role play). Focus was assessed by whether at least one component of the programme focused on fostering personal or social skills, and finally, “explicit” programmes must contain specific SEL skills rather than personal or positive development more generally. SEL interventions also appear to offer the greatest benefits to students receiving free school meals (FSM) and those underperforming in maths and literacy,[143, 146] indicating these interventions’ potential to decrease health and educational disparities. Another meta-analysis of longitudinal school-based SEL programmes identified longstanding impacts on positive relationships, sexual behaviours, income, employment and mental health 3.75 years after the end of the intervention.[138] Even when the effects sizes were small, universal approaches, such as SEL, may translate into significant benefits at a broader, social level.

Restorative practice

RP is based on restorative justice, which is commonly used within the criminal justice system. It seeks to explore and resolve conflict between victims and offenders, and to reduce recidivism through promoting increased empathy. Its implementation in English schools has been growing since 2009, when the Steer Review[91] called for the use of RP to address bullying and aggression.[147-150] RP is distinct from schools’ traditional response to bullying in that it focuses on how to minimise harm and repair the relationship so that the perpetrator can successfully re-integrate back into the school community.[56, 151] RP employs reintegrative shaming,[149] in which the focus of the shame is the act that was committed and not the person. This distinction protects the perpetrator from being reduced to their behaviour and therefore provides an opportunity to improve and act differently going forward.[24] Within schools, reintegrative shame may encourage people to acknowledge and accept responsibility for their wrong doing, and make amends to either the person or community they mistreated. In this way, shame can motivate behaviour change and provide a pathway back into the school’s social network via the victim’s forgiveness.[24]

While forgiveness is still a nascent topic within public health, a systematic review suggests that it has health benefits for both the bully and the victim, as it may replace negative thought processes with positive ones, improve mental health,[152] and

decrease stress.[153] Importantly, forgiveness is a learned skill and has health-promoting applications outside of the context of bullying. Van der Valk argues that forgiveness is particularly important for children who are still growing, learning and developing to see their mistakes as poor choices and not a reflection of their value as a human being. In this way, bullying behaviour does not become a fixed feature of a person's identity.[154]

In schools, RP often takes the form of trying to nurture participatory teaching, building strong rapport and trust between students and staff, recognising and addressing the consequences of misbehaviour, re-integrating pupils after misbehaviour, and preventing future incidents.[155] To be used preventively, people may be trained to have a 'restorative mindset', meaning that they feel open and non-judgemental. It may also involve self-reflection and the use of affective language which may support or challenge behaviour in a way that preserves or strengthens a relationship. RP can be as simple as holding 'circle time' where students may check-in with the peers in groups (for example, form groups) to share feelings and build healthy relationships.

Responsively, RP can be used when a problem or miscommunication occurs, or someone feels hurt or excluded. 'Mediation' or a facilitated discussion can unpack hurtful behaviours and plan how to make amends. In more serious incidents, 'restorative conferences' in which involved parties discuss the harm felt by the victim and perpetrators are encouraged to explain what may have led to their misbehaviour. To start reconciliation, a plan is also agreed about how to repair the harm, how to avoid future conflict and, if necessary, what the perpetrator needs to do to re-integrate into the school community. This may include making apologies or performing acts of service to the community. 'Family group conferencing' will often involve parents or carers but can also involve outside agencies, such as police or social services. A final common strategy for school-based RP is called 'community conferencing.' While this is rarer, it can be used if there is an incident in which harm was done against the community as a whole.[156, 157] These conversations can happen between student or students and staff, and may involve parents.

These responsive restorative meetings employ open-ended questions that explore the harm done and how progress can be made. Unlike traditional anti-bullying meetings in which punishments are meted out, when RP is done correctly, it may also address the unmet needs of the bully, for example, their experience of being bullied at home or feelings of fear that they may not be liked. In this way, multiple harms can be addressed and the perpetrator may be humanised again both by the person they victimised and by their teachers.[24, 158] At the school level, one proposed mechanism through which RP is hypothesised to work is by giving students a voice in decision-making and justice-related procedures, making students feel the process is more transparent and fair, and increasing students' perceptions of the school's authority as legitimate.[159] In the United States and New Zealand, where RP has been used longer in school settings, it is often employed as a "add-on" to traditional disciplinary approaches, used as a diversion programme to reduce suspensions, or implemented as a whole-school approach.[160]

In the course of preparing this thesis, only one RCT of RP other than the INCLUSIVE trial was identified. The Restorative Practices Intervention, based on ecological systems theory and affect theory, was a two-year cluster-RCT conducted in middle schools in Maine, USA. The evaluation did not find evidence of greater school connectedness or less bullying in intervention schools. However, researchers found support for their a priori hypothesis that students who had been involved in RPs (regardless of trial arm) experienced less bullying, better connection to school and more positive developmental outcomes.[156, 158] One limitation of study was that it may not have followed up students and schools long enough to establish a cultural shift in the ways schools carry out their day-to-day work.[160] Developing a restorative school culture requires substantial support from teachers and effects may not be felt until the school's culture has been transformed which may not occur until some time after the intervention implementation has taken place.[131] Evidence from other observational studies shows that systemic school change required between three to five years.[161, 162]

Furthermore, the Restorative Practices Intervention PE found that some control schools implemented more RPs than intervention schools, diminishing the trial's ability to detect change.

Generally, evidence for the use of RP in schools is positive but from weak study designs and non-peer reviewed study reports.[160, 163-166] A scoping review from 2022 reported that most studies showed positive outcomes relating to student behaviour and social and emotional skills.[166] There is some concern, particularly from researchers in the USA, that the current prominence of RP in schools exceeds what the evidence justifies. [34] In New Zealand, case studies of five secondary schools and colleges found that teachers valued RP and felt it was an effective strategy for managing misbehaviour.[167] The use of RP coincided with a decrease in exclusions in elementary, middle and high schools in four school districts in Minnesota, USA[167, 168] although the specific evaluation methods were unclear. In the early 2000s, the Youth Justice Boards in England and Wales began the process of studying the impact of RP in schools.[169] In the London councils of Lambeth and Hammersmith & Fulham, teachers reported that most restorative meetings were effective at addressing bullying, gossiping and disagreements between students and teachers, and students reported that following the implementation of RP, their school was “doing a good job to stop bullying”[169, pg. 38] compared to baseline ($p < 0.05$). However, researchers struggled to recruit and retain schools, did not have a consistent intervention and used comparison schools that had also implemented interventions underpinned by a restorative ethos.[169] In Durham, England, two secondary schools were selected to receive money and resources for the implementation of RP. Researchers interviewed head teachers and purposively selected them based on their commitment to RP and perceived school-level capacity for change,[155] likely leading to an overestimation of potential effects. Students in intervention schools reported a more mutually respectful ethos following implementation when compared to control schools. Specific findings related to bullying were sparse.[155]

A middle school in Oakland, California serving primarily low-income African American and Latino students moved from a zero-tolerance policy to RP for infractions including substance possession and fighting. Following implementation, suspensions dropped by 87%, and qualitative data suggests that relationships within the school were strengthened, students took ownership of their misbehaviour and began implementing RP within their own peer groups.[164] These findings on the transition from punitive to

restorative measures were also found in studies from Minnesota,[\[168\]](#) Bristol,[\[170\]](#) and Scotland.[\[171\]](#)

In Bristol, the Restorative Approaches in Schools (RAiS) programme was implemented as a matched, quasi-experimental study in four schools (with two control schools) characterised by high levels of deprivation, exclusions and behavioural problems.[\[170\]](#) Changing the school ethos and implementing RP consistently was challenging: three of the four schools tried to implement RP incrementally, usually by year group. By the time all teachers were meant to use RPs, many had either forgotten their training or decided that it was a threat to their authority when confronting students.[\[170\]](#) These challenges were also reported in studies from Oakland,[\[164\]](#) Minnesota, US,[\[168\]](#) and Durham, UK.[\[155\]](#)

The results presented above should be interpreted cautiously as confounders may explain observed quantitative differences and many studies were insufficiently theorised. Going forward, evaluations need to consider how RPs are meant to function, how that might vary by key contextual features in each school, and for whom they are effective.

Whole-school interventions

A primary strategy used to reduce bullying is through complex, whole-school interventions (WSIs), or interventions that seek to improve the culture, ethos and social and physical environment at multiple levels within a school.¹ WSIs may incorporate collaboration with the community around the school, parents and external agencies.[\[67, 84, 87, 143\]](#) Moore et al argue that:

"While the public health literature is dominated by intervention approaches which focus on the installation of new packages of activities to address specific health topics, the social dynamics of schools and the social relationships within them may have the potential to influence a wide range of health related outcomes."[\[172, pg. 22\]](#)

Evidence from RCTs indicates that bullying prevention interventions are most effective when implemented and reinforced at multiple levels within the school community,

¹ The results of the INCLUSIVE trial, upon which this thesis is based, are reported in more detail in Chapter 3.

including individual, classroom, year-group, and whole-school levels.[67] Because anti-bullying WSIs conceptualise bullying as symptom of systemic problems,[131, 173, 174] “an intervention on only one level is unlikely to have a significant consistent impact.”[120, pg. 86] Therefore, Vreeman and Carroll recommend multi-component interventions which seek to amend the school’s organisation.[120]

Recently, the Friendly Schools Friendly Family (FSFF) programme reported positive outcomes. FSFF was evaluated in a three-year cluster randomised trial following students in three age-groups (years two, four, and six representing children age 6-7, 8-9, and 10-11 respectively). The use of a usual-practice control arm was not possible because following the successful pilot, many schools requested and received the intervention manual. Therefore, the trial compared schools who received only the manual to those in a medium and high intensity intervention. The medium intensity intervention comprised whole-school activities to improve school climate, student and student-staff relationships, revising rules and policies, implement age-specific classroom activities, and restorative approaches when indicated. The high-intensity intervention comprised the aforementioned activities as well as resources meant to engage parents via newsletter items, staff training on parent engagement, parent workshops and suggested home activities. Students aged 8-9 (grade 4) in the low intensity (quasi-control) arm were more likely to be frequently bullied compared to students receiving the high-intensity intervention (OR=1.76²) At the end of grade 6, students in the low intensity arm were more likely to be frequently bullied (OR 1.54, p=0.005). A dose-response relationship was identified between victimisation and intervention intensity with students in the high-intensity arm having the best outcomes compared to students in the medium or low-intensity intervention.[175]

While multiple WSIs have been evaluated, the field is dominated by the Olweus Bullying Prevention Programme (OBPP).[135] The intervention has four guiding principles for adults, including showing warmth and having positive relationships with students, establishing firm limits on unacceptable behaviour, consistent negative (non-violent)

² Please note, that in this study the odds ratios generated from logistic regression were converted to effect sizes by “calculating the natural logarithm of the odds ratio and dividing the standard deviation of the logistic distribution, namely 1.81”[175, pg. 400]

consequences when rules are broken, and being an authoritative and positive role model.[135] While the intervention can be tailored, OBPP interventions generally consist of 19 components across individual, classroom, school and community domains. Key components of the programme include a coordinating committee to ensure high degrees of intervention fidelity, staff training, support from a trained anti-bullying professional, annual surveys about bullying, and regular check-ins with students to reinforce anti-bullying messages. OBPP has shown consistently positive results in Norway, where it was developed, and other Scandinavian countries but has had inconsistent and sometimes null effects in the USA.[34] Smaller class sizes, longer teacher training and a cultural norm of welfare intervention may explain more positive outcomes in Norway.[120, 131] Reviews of other bullying prevention interventions have also reported that bullying interventions decrease victimisation less effectively in the USA[130, 132] although few explanations as to why these findings emerge have been offered. Subgroup analysis of some OBPP in the USA show significant reductions in bullying for White children but no benefits for children from other racial groups.[135] How this difference in effects materialised was not explored in the analysis.

The evidence on WSI from systematic reviews and meta-analyses has contradictory results ranging from iatrogenic harms to significant benefits across a range of outcomes, [120, 131, 132, 137] but a more positive picture appears to be emerging over time. In 2004 a systematic review of anti-bullying WSIs found that 93% of effect estimates relating to victimisation and 92% of effect estimates related to bullying had negligible or negative outcomes, and study authors argue that the evidence was too inconsistent to justify their adoption.[131] Published in 2004 and 2007, two further systematic reviews found evidence of WSIs decreasing bullying perpetration and victimisation, in at least some student populations, if not overall.[120, 131] By 2011, Ttofi and Farrington's meta-analysis found an overall reduction in bullying of 20-23% and a dose-response relationship between the number of programmatic components and the intensity of the programme on decreased bullying.[132] A systematic review of interventions using the Health Promoting Schools Framework concluded that setting-based, multi-level interventions reduced bullying victimisation but not perpetration,[176] although the authors did not articulate the theorised mechanisms through which altering the school environment enabled the reduction of bullying. In 2021, a systematic review examining

69 randomised trials showed that universal approaches were more effective than targeted approaches.[177] A meta-analysis of 71 effect sizes from 11 interventions found that victimisation was significantly decreased more than one-year post-baseline (OR=0.85),[178] and a separate meta-analysis by Ponsford et al identified small but significant effects of WSI relating to victimisation (OR=0.84) and perpetration (OR=0.85).[178] In systematic reviews and meta-analyses, it can be challenging to unpick the nuance that may be contained within individual studies in order to understand which intervention components appear to work best or how populations appear to be impacted differently.

Numerous systematic reviews have found that WSIs are more effective than either curricular or targeted approaches.[16, 120, 130-132, 179] Both a systematic review of 41 studies[79] and meta-analysis of school-based anti-bullying interventions[130] found that changing the school environment was among the most effective components within studied interventions. Gaffney et al reported that the most substantial drops in bullying perpetration arose from programmes that modified their disciplinary methods, encouraged more co-operative working, had more programmatic elements and were long-term interventions, all of which are key features of many WSIs.[180]

One factor contributing to the confusion around the usefulness of WSIs is the infrequent use of moderation analyses. In their systematic review and meta-analysis of WSI promoting commitment to school to decrease violence and substance use and improving attainment, Ponsford et al reported 14 studies of 11 interventions, only eight of which (all RCTs) reported any kind of subgroup analysis in relation to violence perpetration. Sex was the most common moderator explored. Age, socioeconomic position, and baseline risk were all inconsistently explored.[178] Moderator analyses show who benefited most from interventions, demonstrating whether interventions are inequality-generating or equality-generating. Knowing who benefits, is left unimpacted, or harmed by an intervention can also guide the improvement of theory to explore why populations experienced different outcomes.

Limitations in evidence base and need for more useful evaluation evidence

SEL appears to be particularly useful in reducing health inequalities[143, 146] and RP appears to promote the development of interpersonal skills.[166] There is growing and increasingly positive evidence about the role of WSIs in reducing bullying victimisation.[178, 180] Even when effect sizes are small, universal interventions like WSIs may still generate meaningful public health benefits and reduce sequelae associated with bullying. However, gaps in the evidence base remain, particularly in relation to specific types of interventions that may work, for whom interventions work, through what mechanisms intervention resources help produce change, and how this varies by context. Moreover, many of the interventions being evaluated used insufficient or inappropriate theory, making the effective transfer of those interventions to new contexts more difficult to assess.

In relation to specific types of interventions, evaluations of RP have been more positive in non-randomised and weaker study designs.[165, 169-171, 181-183] A two-year RCT faced a number of implementation problems and may not have followed students up for a sufficient period to demonstrate significant improvements.[156, 158] Aside from the INCLUSIVE trial, upon which this thesis is based, there have not been any full-scale evaluations of complex, whole-school anti-bullying trials in the UK. A substantial amount of evidence has come from the United States and there is an ongoing debate about how cautious researchers need to be when assuming transferability.[135, 142] A systematic review found that interventions are more effective when fidelity is assessed prospectively.[128, 184] The evidence base, particularly of multi-component complex interventions, could be strengthened by assessing *how* better fidelity and to which components leads to the activation of theorised mechanisms and later outcomes. The evidence above also showed that there is insufficient use of moderator analyses in primary studies and systematic reviews in the bullying literature.[120, 142] For example, using weighted regression analyses, Ttofi and Farrington have argued that certain anti-bullying interventions, such as those targeted at children younger than 11 years-old or those that involve students working with bullied peers, should be discarded or de-prioritised because of limited effectiveness in pooled studies.[132] Smith et al counter-

argued that when those effects are examined by key moderators, such as age or length of time being bullied, they are, in fact, effective.[185] Copeland et al similarly argue that insufficient use of moderator analyses masks significant heterogeneity in outcomes.[186] A study exemplifying the benefits of exploring moderators came from the AAYP. A mixture modelling study was conducted in which researchers identified three distinct groups of high-school aged African American males based on their risk of violence trajectories. The results showed that intervention effect estimates were three times as large for the young men in the high-risk group compared to non-high-risk men.[187] Conventional sub-group analysis by sex only would not have uncovered this finding which has implications on health equalities.

Generating a better understanding of differential impacts would also enable policy-makers and relevant stakeholders to select appropriate interventions for the communities they serve or ensure that interventions which will exacerbate existing inequalities are not selected. Even when subgroup analyses are used, they may only examine person-level moderators such as sex, age, or previous experience being a bully or victim.[67, 135, 178] Moderator analyses could be used to move beyond the individual to explore contextual features, such as school institutional types, student profiles or organisational capacities or cultures.

Mediational analyses could also be used more frequently to understand how interventions work and the likely impact of individual components within multi-component studies.[128, 130] For example, OBPP has approximately 19 components,[135] but no evaluation has yet sought to understand which ones are likely to be effective. A helpful example of mediation analyses within an RCT was led by Gardner et al. In their RCT of parenting intervention to reduce child conduct problems, they used Barron and Kenny's method for mediation analyses[188] and found that the development of positive parenting skills mediated the change in child conduct problems ($r=0.40$, $p=0.001$) while parental mood and self-reported confidence did not.[189]

One recent concern that has emerged within the study of WSI is a concern about what is called the "healthy context paradox." [190-192] Researchers found that despite the KiVa WSI "working" insofar as it reduced the overall prevalence of bullying, students who

began to be bullied and those who remained bullied in schools trying to improve their ethos, had worse depressive symptoms and lower self-esteem than students in schools that were not trying to decrease bullying or improve their ethos.[192] Researchers hypothesised that this deterioration of mental health may be caused by having reduced support networks of co-victimised peers, or that bullied student may internalise the belief that they deserve to be bullied more because fewer other students now share their experience.[190] The hypothesis was tested using data from the INCLUSIVE trial but neither of the two mediation models found statistically significant support for the existence of the paradox (See Appendix 1 for more detail),[193] giving further support to the need for nuanced analyses about how, for whom, and under what conditions interventions ‘work.’

The review of this evidence also showed how infrequently qualitative data are being used to explore how intervention resources reduce bullying and aggression. This lack of nuanced data on implementation of complex interventions contributes to confusion about whether and to what extent interventions are helpful, for whom, where, and how different contextual features affect intervention effects. Understanding participant narratives about the use of resources would also help explain the theoretical connection between implementation and outcomes, which may facilitate a better understanding of transferability and generalisability within the field of bullying research. (For a more detailed exploration on the use of qualitative data to explore causal mechanisms, see Appendix 2.)

Interventions are not always informed by theory and when they are, it may not be appropriate to the phenomena they hope to explore. For example, the theoretical underpinning for the Gatehouse Project intervention was attachment theory formulated by John Bowlby, a psychiatrist and psychoanalyst. Generally exploring the relationship between young infants and primary caregivers, the theory was applied to schools for its relevance to having secure adult figures (parents or teachers) and settings (home or the school).[194] Applying a psychological theory about dyadic relationships to an institutional process was not helpful in their evaluation. Gatehouse researchers were clear about the components integral to their theory of change but they did not assess whether the empirical evidence supported it.[85] Selecting an appropriate theory is

important because it can elucidate how and why change occurred. If evidence supports the hypothesised mechanisms of change, then it may be easier to select or design interventions that are more effective in different settings.

To make clearer sense of all the literature in this field, researchers must ask more nuanced questions focusing not only on what works, but for whom, under what conditions, to what extent, and how.^[1] This thesis will explore these questions within a realist framework.

Description and structure of this thesis

This thesis contains one further introductory chapter exploring approaches to evaluating complex interventions (Chapter 2) before moving on to the methods (Chapter 3).

The results section contains three empirical chapters on the following topics:

- 1) An exploration of intervention AGs as a participatory approach for coordinating whole-school health promotion interventions and how context appears to affect implementation;^[8]
- 2) Using grounded theory within a realist trial to develop hypotheses about how context and mechanisms interact to generate outcomes;^[9]
- 3) Using qualitative comparative analysis (QCA) to assess complex causal pathways to reduced bullying victimisation in a whole-school, realist, randomised controlled trial (returned to reviewers post revision at the Journal of School Violence).

The discussion section is a methodological commentary exploring:

- 4) Are realist trials possible? Is the use of randomisation and control groups philosophically and methodologically compatible with realism, can realist trials make a useful contribution to the evaluation of complex interventions, and within this first exemplar, what worked well, what did not work well, and how can realist trials be improved upon in the future?^[11]

Chapter 2: Approaches to evaluating complex interventions

Complex interventions

Complex interventions are characterised by component synergies and feedback loops, emergence, adaptability and unpredictability.[195-197] Feedback is the phenomenon in which one change creates the context for further change. For example, evaluations of the KiVa antibullying programme identified improvements in peer-defending, which diminished the social rewards that bullies received, which therefore decreased the frequency of bullying.[198] Emergence relates to changes that occur because the components behave synergistically or antagonistically in contingent ways that cannot be limited to the level (individual, dyadic, organisations, populations) the intervention sought to influence.[199] Adaptability refers to people changing their actions, behaviours or beliefs in response to the effects of the intervention. Finally, unpredictability refers to our inability to speculate on all possible outcomes emerging from a complex intervention operating in a complex environment.[200]

Theories of change and logic models

A crucial, but often underdiscussed, feature of interventions is that they implicitly express a theory of how something may produce change. For example, if an intervention seeks to educate patients, then those developing the intervention are theorising that people lack learning and that providing them with knowledge or skills will change behaviour, which would consequently improve outcomes. As evaluations of complex interventions are primarily employed to test hypotheses, the majority of evaluations are underpinned by a 'hypothetico-deductive approach.'[201] However, the extent to which hypotheses are grounded in a theory and an intervention theory of change are informed by an appropriate mid-range theory, is highly variable.

Ideally, intervention theories of change and evaluation hypotheses are informed by mid (also called middle) range theories.[202, 203] In his 1949 essay, "On sociological theories of the middle range", Robert Merton defined middle range theories as the:

"theories that lie between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the

observed uniformities of social behaviour, social organisation, and social change."[204, pg. 39]

Placing middle-range theory between "piecemeal empiricism" and "grand theory", it enables researchers to build and test social science theories about how the phenomena of interest works in the world. Relevant theories for public health interventions may come from psychology, sociology or economics but it is important that they are relevant to the intervention itself.

Building from an appropriately selected mid-range theory, an integral theory of change and logic model should inform the design and evaluation of complex interventions.

Theories of change are the explicit articulation connecting: the problem an intervention seeks to improve; intervention materials or supplies; the intended intervention processes that these resource seek to enable; the mechanisms; and short, medium, and long-term outcomes the intervention is aiming to address[205] (although not all theories of change will include each of these steps). In short, theories of change are plausible hypotheses of how or why a specific intervention may produce a change in certain contexts or with different populations.[206]

When theories of change are drawn diagrammatically, they are called logic models. Logic models can be useful throughout the research process and help intervention developers make explicit why they theorise that providing certain resources will be helpful to improving a problem. Developing a logic model may expose weakness in the intervention design, internal inconsistencies, and potentially unreasonable expectations.[195] The 2015 Medical Research Council (MRC) guidance on process evaluations emphasised the need for logic models to contain a clear description of the intervention inputs, the hypothesised mechanisms, and how context will likely impact their activation.[195] Logic models may also be easier than theories of change for a broad range of stakeholders to understand so that a broader coalition may contribute to intervention development and evaluation. This level of detailed, explicit theorisation is relatively recent and has been emphasised in the UK, in particular by the MRC.

Randomised controlled trials

Complex interventions are commonly evaluated using RCTs. Since their rise to prominence in the 1940s, RCT methods have been refined and developed.[207]³

Design

In the simplest terms, an RCT aims to test hypotheses about cause and effect by randomly allocating a sufficient number of people or groups to receive (or not) an intervention, following them up for an appropriate amount of time, and assessing differences at the end of the experiment.[207, 209, 210] The control arm serves as a contemporaneous counterfactual, allowing researchers to get a sense of how an intervention appears to alter the distribution of outcomes over and above all the other factors influencing these within the study context. When researchers design studies that employ before-after designs without control groups, one cannot distinguish how an intervention alters the distribution of outcomes over and above time-related factors, such as maturation, secular trends or regression to the mean.

The random allocation of people or groups to either the intervention or control arm offers two primary benefits. Firstly, it minimises allocation bias by disallowing researchers or clinicians from choosing which participants are allocated to either study arm.[207] Secondly, it tends (though is not guaranteed) to minimise baseline differences between groups including potential confounders. While some confounders will be known in advance, random allocation also tends to distribute unknown confounders evenly between trial arms.[211] In this way, randomisation also tends to achieve approximately equal variation in moderating factors between arms. This enables researchers to examine potential moderators within a randomised comparison, while minimising bias.

³ In the UK, the MRC regularly leads expert consultations and releases guidance documents which are used to improve future research. In 2000, the MRC published A framework for development and evaluation of RCTs for complex interventions to improve health. [208] In 2008, Developing and evaluating complex interventions [197] was released in which the model of intervention design was revised to be less linear and more accurately reflect the iterative nature of intervention development. In 2015, Process evaluation of complex interventions [195] emphasised the relation between context, implementation and mechanisms. These documents have provided needed clarity for those conducting RCTs.

Intervention effectiveness is primarily assessed through ‘intention-to-treat’ (ITT) analysis which compares participants or groups who were assigned to receive the intervention to those who were assigned to the control arm. ITT analysis minimises selection bias arising from different levels of engagement with the intervention, where the people who achieve a high dose are likely to have different baseline characteristics from those in the control group, which is not subject to such differences in engagement.

Process evaluation

RCTs increasingly employ both qualitative and quantitative data to understand intervention processes and mechanisms.[\[212, 213\]](#) Quantitative data have been used in PEs to understand factors such as fidelity, reach and acceptability of interventions. Qualitative data have primarily been used to better understand acceptability and the experience of providing or receiving interventions but are increasingly also used to understand quality of implementation, deviations from study protocols, and to describe the study’s various contexts and their effects on intervention delivery. Participant accounts of causal mechanisms can also be used to explore how intervention resources and their use by various actors may generate outcomes.[\[142, 197, 213, 214\]](#) PEs are now an integral feature of many, although not all RCTs[\[215\]](#) and they can vary widely in their scope. Some aim to provide detailed information about how interventions are implemented, how they work, and how external factors affect their implementation or functioning.[\[216\]](#) Some seek only to explain the processes through which interventions are implemented and received by stakeholders.[\[217\]](#) They do not necessarily engage with *how* mechanisms of action work or how these processes impact upon the outcomes, although they can be designed to do so.[\[216\]](#) When interventions fail to show any significant outcome, PEs can be used to distinguish ineffectiveness due to failure of intervention theory (conceptual failure) or due to failure to implement the intervention (also called type-3 error).[\[214\]](#) This becomes especially important in multi-site trials of social interventions, where it is unlikely that the intervention will be delivered in the same way in different places.[\[214\]](#)

Origins and developments

Edward Suchman first articulated the concept of PE in 1967, focusing primarily on implementation fidelity.[\[218, 219\]](#) By the early 1980s, investigators from various cardiovascular disease (CVD) prevention demonstration projects began to focus on the concept of dose,[\[219\]](#) which was later broken down into ‘dose delivered’ and ‘dose

received.' For example, health campaigners may create and distribute an informational leaflet (dose delivered) but no patient may read it (dose received).

The growing focus on implementation and dose enabled the exploration of 'on-treatment analysis' (OTA). These may give a sense of intervention effect estimates among those who actually received the intervention (as opposed to who was assigned to receive it) and can be used to establish dose-response relationships based on fidelity and receipt findings. However, these must be conducted and interpreted cautiously since they are subject to the selection biases that ITT analysis protects against.[220, 221]

Later, more qualitative questions were incorporated into PEs, focusing on how participants perceived the intervention's intensity, acceptability and impact.[222] In response to growing frustration with RCTs, specifically the limitations of ITT analyses[217, 223] and the failure to explore the role of context and mechanisms, researchers began pushing for increased integration of qualitative methods within trials.[224] It was also during this time, that a few studies began to build on the concept of mechanisms in more detail. An exemplar at the time was the Randomised Intervention trial of PuPil-Led sex Education (RIPPLE). Through their PE, researchers were able to discern that more students in the peer-led arm were exposed to skills-based activities and that high-achieving students with higher socioeconomic status were more likely to be educators than their peers,[225] indicating potential mechanisms through which resources affected outcomes.[226, 227] An evaluation with so much detail was rare at the time.

The first PE for a school-based intervention was the Child and Adolescent Trial for Cardio-Vascular Health (CATCH). It examined participation, dose, fidelity and compatibility which attempts to capture whether or not an intervention fit in with the needs and expectations of those delivering it.[228] Reporting PEs findings was difficult as there were no agreed components that should be included.[223] In the late 1990s, researchers such as Baranowski and Stables sought to systematise PEs through the development of various frameworks,[229] while others were tailored to specific evaluation designs.[230] In their 2013 review of PEs, Grant et al found a lack of explicitness about which publications were PEs, what their overall purpose was, whether they were conducted simultaneous or post hoc to the intervention and what the outcome findings of the associated trials were.[230] In 2000, when the MRC

published their Framework for the development and evaluation of RCTs for complex intervention to improve health[\[208\]](#), the authors had a clear focus on the need for theory in both developing and evaluating interventions, but did not articulate a clear conceptualisation of process or the role of context. Mechanisms were mentioned only insofar as they influence how one theorises changes in outcomes, and qualitative data were largely suggested for assessing fidelity and deviations.[\[208\]](#) It was not until the MRC's guidance on process evaluations of complex interventions, that there was an explicit and detailed focus on mechanisms and how those are impacted upon by implementation and context.[\[216\]](#)

Realist approaches to evaluation

In 1997, Ray Pawson and Nick Tilley published *Realistic Evaluation*.[\[1\]](#) They recognised that trials were describing changes in outcomes but most were not being used to understand mechanisms or unpick why so many replication studies were reporting inconsistent findings.[\[1\]](#) Without changing how evaluations were conducted, researchers were leaving the “black box problem”[\[1\]](#) of what works, for whom, under what conditions, and how, unopened. In short, trials were not providing the evidence that was needed to inform policy and practice. Therefore, they developed an alternative approach to evaluating complex social interventions.

To illustrate the need for evaluations that account for context and mechanisms, Pawson and Tilley give the example of a three-arm RCT in which police responding to domestic violence incidents were allocated to A) arrest (though not necessarily charge) the perpetrator, B) provide advice to the survivor, or C) send the perpetrator away.[\[1\]](#) 10% of victims in group A contacted the police again in six-months compared to 19% and 24% in groups B and C, respectively (p-values significant but not reported). Based on these encouraging findings, numerous US cities adopted the policy. However, in six follow-up RCTs, half reported that group A had higher rates of repeated calls to police, contradicting the original and expected findings. Because they were under-theorised, the RCTs could not be used to explore the role of context and could therefore not uncover that when arrests occurred in stable communities with high employment, an arrest shamed the perpetrator into non-offending, where in communities characterised by instability and poverty, the added burden of an arrest sparked greater rage in the perpetrator, who was thus more likely to re-offend.[\[231, 232\]](#) While these RCTs were

each methodologically sound, the analysis was too descriptive. Had the research been focused on testing a theory rather than simply evaluating an intervention, police departments would have been better placed to assess whether a change in policy would be helpful or harmful in the community they serve.

When defining the key characteristics of a programme or intervention, Pawson and Tilley first argue that “Programmes are theories incarnate.” [233, pg. 3] If a researcher or policy maker did not theorise that Intervention A could reduce Problem B, it never would have been conceived of as a possible solution, let alone designed, implemented, and evaluated. A realist evaluation therefore seeks to make explicit how the physical, social, or cognitive resources (e.g. training) given as part of a programme activate or deactivate certain mechanisms within certain contexts to produce an outcome. These mechanisms are not mechanical in nature but are the “latent powers and capacities of individuals” [4, 234] that may be activated. Articulating, testing and refining a programme theory are the primary objectives of realist evaluation. [1]

Pawson and Tilley argued that the second key feature of a programme is that it is “embedded” [1, pg. 4]: they do not occur in social vacuums. Researchers need to remain mindful of how their intervention affects behaviours or contexts at various levels, including: individual, interpersonal, institutional and infrastructural. Moreover, the outcome of an intervention is likely to disrupt the interplay between these layers. For example, an intervention targeting processes within a school will also likely change individuals and the dynamic between students and teachers. The effects of a programme are not likely to be contained within the explicitly targeted level.

Thirdly, programmes are active. People need to do something with the programme’s resources to activate the mechanism that produces a benefit.⁴ Finally, programmes are

⁴ This belief, central to realist evaluators, is a deviation from Bhaskar (the originator of critical realism, discussed in more detail later in the chapter) and a point of contention with critical realists. [234-236] In *The Possibility of Naturalism*, [237] Bhaskar clarifies that when trying to understand society, there are two, separate sources of causation: 1) structured social relationships (which behave in largely predictable ways), and 2) human agency. Within his worldview, Bhaskar argues that structure precedes agency as it is the material cause for human action: we are born into a pre-existing context full of potential causalities. Our actions are mediated through human agency and social activity: “Society is both the ever present condition (material cause) and the continually reproduced outcome of human agency.” [237, pg. 34-35]

open systems, continually influenced by and reactive to the larger context in which they occur. Realist evaluators often argue that trialists believe in successionist explanations of causation while they maintain “generativist” explanations of causality. Their call for generativism comes directly from Roy Bhaskar who, despite being the originator of critical realism (the philosophy underpinning realist evaluation, discussed more in Chapter 3), never provided readers with an actual definition, instead choosing to describe its most salient features. Chief amongst them is that “generative mechanisms...must be analysed as the way of acting of things; and their operations must be understood in terms of the exercise of tendencies and causal powers.” [241, pg. 184] In effect, a generativist view holds that what natural or laboratory-based scientists might consider laws, are actually only tendencies or demi-regularities, and in order to understand how phenomena change over time, focus must be paid to the context of their emergence.[242] As Pawson and Tilley explain, “One happening may well trigger another but only if it is in the right conditions and the right circumstances.” [1, pg. 34] Moreover, within generativist understanding of causation, researchers are more likely to be sensitised to and therefore explore causal explanation that include liability, power, agency and other forms of potential, which they argue are invisible when operating from a successionist epistemology.

Realist evaluations generally begin by hypothesising how novel resources may enable people to activate one or more mechanisms within their context to produce the desired outcome(s). Distinguishing realist evaluation from traditional experimental evaluations is the focus on context (C), mechanisms (M), and outcomes (O), and subsequent CMO configurations (CMOCs). [1] Within realist evaluations, context refers to “features and conditions in which programmes are introduced that are relevant to the operation of the programme mechanisms.” [233, pg. 7] Within realist evaluations, a mechanisms is “the choices and capacities which lead to regular patterns of social behaviour.” [1, pg 216] Described another way, how a participant uses, interprets, or acts upon the newly

He called this dynamic relationship the Transformational Model of Social Activity (TMSA) and thought it should be the focus of scientists trying to understand causality involving agents. People are unable to create society (because they are born into it) but it should be understood as an assemblage which agents can reproduce and transform. It should also be noted scholars rarely use his TMSA[238] preferring instead, the more comprehensible works of Archer[239] or Giddens.[240]

available resources is the mechanism, not the resource itself.⁵ By connecting context to mechanisms, evaluators can plan investigations to assess the plausibility of various CMOCs. In one of their case studies of realist evaluations, Pawson and Tilley's research process included analysing qualitative interviews to build hypotheses about how, for whom, and under what circumstances prisoner education had positive impacts on recidivism, which were then tested quantitatively.[\[1, pg. 107-114\]](#) It is unfortunate that the quantitative analyses they used were not described.

Pawson and Tilley describe themselves as “solid members of the modern, vociferous majority...for we are whole-heartedly pluralists when it comes to the choice of method.”[\[1, pg. 85\]](#) Yet compared to the rich, varied, and quickly growing field of qualitative realist evaluations, quantitative realist evaluations are exceedingly rare[\[246-250\]](#) (although from published study protocols,[\[251, 252\]](#) we can expect more to be published in coming years). It should be noted that Tilley in particular has been exploring the idea of “data signatures” or quantitative indicators of a mechanism's activation.[\[253\]](#) These realist studies have utilised routine or programme monitoring data to conduct various non-experimental quantitative analyses to assess CMOCs. Thus far none have used trial data, although an ongoing critical-realist RCT examining the effect of music therapy on patients receiving palliative care may involve quantitative analyses.[\[254, 255\]](#) Other studies are beginning to explore the use of realist-informed PEs[\[256\]](#) which in some cases, are connected to larger RCTs.[\[257, 258\]](#)

Critiques and limitations of conventional evaluation

Despite the aforementioned strengths of RCTs, numerous critiques, primarily although not exclusively from realists, have been raised relating to both their ontological and epistemological assumptions and the limits of what they are able to examine and explain. In the section below, I summarise their criticisms of RCTs before responding to them in the later section called “Defence of RCTs and realist RCTs.”

⁵ The conceptualisation of mechanisms has evolved since the publication of Realistic Evaluation and now includes broader interactions between agency and structure that occur as a result of intervention activities. The works of Dalkin,[\[243\]](#) Lemire,[\[244\]](#) Archer,[\[239\]](#) Porter,[\[236, 245\]](#) and Bhaskar[\[237, 241\]](#) may be particularly instructive. The debate about competing realist notions of mechanisms is outside the scope of this thesis.

Philosophical concerns:

Positivism and successionism

The most often repeated philosophical concern about trials is the belief that they are ontologically positivist and employ a successionist understanding of causality⁶.[\[1, 3, 235, 260-263\]](#) Hinds and Dickson explain that, “positivism derives explanations of cause from observing correlations [and] constant conjunctions, which suggest interactions between causes and their effects.”[\[235, pg 3\]](#) Similarly, Greenhalgh and Manzano argue:

“the [RCT] epitomises the positivist approach to the evaluation of drug effectiveness, but it is also commonly applied to social problems such as crime and health promotion. The RCT is the gold standard for establishing the effectiveness of interventions and follows a successionist understanding of causality. The logic goes that creating a closed system that controls for extraneous circumstances (context) allows us to determine that intervention X was the cause of outcome Y.”[\[263, pg. 2\]](#)

Realists[\[1, 3, 237\]](#) trace their concerns about experimentation to David Hume’s Treatise of Human Nature (1739),[\[264\]](#) and argue that for Hume, the presence of a ‘constant conjunction’ in which one event consistently follows another, is sufficient for establishing a causal relationship.[\[245\]](#) Their interpretation of Hume is that in order to differentiate a causal link from a spurious association and identify the constant conjunction, experiments must be designed so that all other factors are excluded from the study, which is only possible in ‘closed’ settings. Realists opposed to trials argue that humans, having consciousness and agency, do not behave ways that are as reliable as, for example, chemicals, and the study design is therefore inappropriate for exploring social phenomena.[\[1, 3, 4\]](#) Therefore, realists argue that empiricism and Humean exploration of causal laws “must all be totally discarded.”⁷[\[237, pg. 45\]](#)

⁶ Please note that our review of the philosophy and social science literature did not identify successionism as a key tenet of positivism but is often referred to as such by realists. For this reason, we discuss them together. For a separate study on the key tenets of positivism and whether RCTs are positivist, please see Appendix 3.[\[259\]](#)

⁷ For Bhaskar, “practically all the theories of orthodox philosophy of science, and the methodological directives they secrete, presuppose closed systems. Because of this, they are totally inapplicable in the social sciences (which is not of course to say that the attempt cannot be made to apply them - to disastrous effect). Humean theories of causality and law, deductive-nomological and statistical models of explanation, inductivist theories of scientific development and criteria of confirmation, Popperian theories of scientific rationality and criteria of falsification, together with the hermeneutical contrasts parasitic upon them, must all be totally discarded.”[\[237, pg. 45\]](#)

Practical concerns:

Randomisation and control stifle one's ability to understand how change happens

Both realist and non-realist critics of trials have argued that they are too “controlled.”

It is unclear if by “controlled” they mean conducted in settings or with populations that are so homogenous that it limits any analysis of how outcomes vary by setting or population or tightly regulated as to ensure consistent implementation. Regarding the former, some trials do aim to ensure the fidelity of interventions to ensure that the study is an assessment of an intervention’s theory of change or a product’s efficacy while other trials aim to examine effectiveness in real world conditions which include real-life levels of fidelity. Regarding the former, it is true, particularly in clinical and pharmacological studies, that some trials intentionally narrow inclusion to certain settings or populations, an approach recently criticised in terms of its equity and usefulness.^[265] If the latter, critics suggest that RCTs of social interventions are as tightly controlled as laboratory-based experiments, making it impossible to assess how or under what circumstances outcomes are changed.^[250] Greenhalgh and Manzano argue:

RCTs treat context as a source of bias to be eliminated through randomisation and standardisation of the intervention (Van Belle et al., 2016). In this type of study, context is often conceptualised in terms of ‘general circumstances’ that are described as background information or mixed-up with implementation sites and troubles. Consequently, context becomes an annoyance to be minimised, obliterated or overcome.^[263, pg. 2]⁸

Marchal et al fear that the use of randomisation only allows researchers to examine the role of key confounders and argue:

“Randomised designs will also typically define a stratified sample and include statistical correction for baseline differences in gender, ethnicity and age between groups. Indeed, much of the progress in RCT methodology in recent years has been in the refinement of such techniques. However, applying such techniques to rigorously maintain internal validity leads to a situation where it is not possible to determine in which conditions and through which configuration of factors the outcome of interest is reached.” ^[3]

⁸ The Van Belle publication cited in this quote is from "Can “realist” randomised controlled trials be genuinely realist?." *Trials* 17.1 (2016): 1-6.

RCTs are insufficiently theorised

As discussed earlier, RCTs are normally used to test hypotheses but not all hypotheses are informed by intervention theories of change and not all theories of change are informed by middle range theory. Moreover, not all middle range theory is carefully chosen so that it aligns well with the intervention and its outcomes. Indeed, some “off-the-shelf” theories may not be useful and may actually blinker researchers from “seeing” important mechanisms.[203] For example, obesity interventions may target an individual’s eating habits or motivations for weight loss but fail to recognise that social and structural drivers are not related to any specific person but to the social, political, and economic context that surrounds them.[266] As mentioned in the previous chapter, social interventions sometimes employ theories derived from psychological theories which do not adequately engage with social interactions or institutions. As described earlier, the theory underpinning the Gatehouse Project was attachment theory and was applied to schools for its relevance to having secure adult figures (parents or teachers) and settings (home or school),[194] despite the relationships being non-analogous. A sociological or organisational psychology middle range theory would have been more appropriate for exploring bullying. RCTs that do not sufficiently engage with theory are unlikely to contribute to the refinement of theory which traditionally is the way in which scientific research informs transferable learning.

Non-theoretically driven evaluations may also miss key causal processes emanating from contextual features. Moore and Evans use the example of zero-tolerance tobacco policies in schools in the early 2000, when teen smoking was at its most prevalent. These campaigns sought to demarcate smoking as outside of schools’ norms. Now, when teenage smoking is less common, the same intervention produces very small results because the contextual features which made the intervention relevant are no longer present (and in fact, may encourage smoking as it then becomes a sign of rebellion).[203]

RCTs are unable to assess mechanisms

One key criticism of experimental studies is that mechanisms are conceptualised as strings of variables that explains why a correlation exists between an independent and dependent variable.[4, 250, 267] In 1983, Chen and Rossi, developers of ‘theory-driven evaluation’ [268] explained that:

A very seductive and attractive feature of controlled experiments is that it is not necessary to understand how a social program works in order to estimate its net effects. [269, pg 284]

Van Belle et al express a similar concern that within trials, mechanisms are equated with intervention components and are not understood to explore how the resources and opportunities created by the intervention are taken up (or not) by people in different contexts. [4]

A repeated criticism of trials is that questions about ‘what works’ are too descriptive and that trials have traditionally offered little explanation about *how* resources introduced into an environment led to change occurring. [2] Similarly, Hawkins argues that, “Without knowledge of the deeper reasons for an effect, the evaluation of a program or intervention into a complex system which was effective does not usually provide adequate information about what will work in a future context.” [250, pg. 277]

RCTs are only concerned with attribution

The final concern about trials is that they are only concerned with attribution, or whether or not the intervention caused change. [3, 235, 250, 269]

The consequence of these aforementioned limitations is that much of the evidence for public health interventions is difficult to interpret and offers very limited information about what interventions are useful in different contexts or populations. With the current evidence, insufficient attention has been paid to the contextual features of an intervention’s implementation, especially in relation to whether those were necessary pre-conditions for the change that may have occurred during the trial. [250, 263] It is also difficult to unpick the current evidence to understand whether, for whom, and how interventions are effective, or hence offer little insight into potential transferability.

In the next section, I will look specifically at the proposal of realist RCTs, the critiques and concerns that some realist evaluators have expressed about them, and a brief defence of them.

Realist trials

Realist trials have been proposed as an approach that incorporates the many useful aspects of realist evaluation into the methodologically rigorous design of an RCT. [2] This approach accepts some criticisms of RCTs but sees them as contingent, not necessary features. Proponents of realist RCTs argue that many concerns about RCTs

can be attended to by building more theoretically informed interventions and by designing process and outcome evaluations which enable researchers to explore how interventions work, for whom, and under what conditions. In Bonell et al's original description, the proposed evaluation consisted of three stages: 1) developing a theory of change and a priori CMOs based on an appropriately chosen, conceptually deep mid-range theory about how the intervention may work; 2) analysing qualitative data from PEs to explore participant views about plausible mechanisms, attentive to how interactions between agency, context, and interventions resources appear to affect underlying mechanisms, and revise the original CMOs accordingly; and 3) assessing these hypotheses quantitatively using moderator and mediator analyses. The proposal was that these findings and results could then be synthesised to refine the theory of change which would inform decisions about deploying this or similar interventions elsewhere.[\[270\]](#) Soon after the concept was proposed, it was incorporated into the MRC's guidance for process evaluations of complex interventions[\[216\]](#) but was met with resistance from realist evaluators, and ongoing debates about the philosophical cogence and practical feasibility have ensued.[\[2-4, 234-236, 245, 250, 270-272\]](#)

If philosophically cogent and practically feasible, realist trials would help build theoretically-grounded interventions, enable researchers to develop CMOs by drawing on data from diverse contexts in which various mechanisms are able to operate, and test hypotheses using data that is less biased than that available from the observational designs commonly employed in realist evaluations.[\[2, 62, 270\]](#) Importantly, realism forces researchers to focus on what has too commonly been ignored in trials, particularly in relation to the contextually contingent activation of mechanisms.[\[273, 274\]](#) While posited by some realist evaluators as philosophically and methodologically incompatible and even 'oxymoronic'[\[3\]](#), it is possible that realism focuses trialists' attention on more specific and nuanced questions and trials provide realist evaluations with effect estimates that are less biased and confounded and subject to random error.

Critiques of realist RCTs

The debate about the possibility of realist trials has involved many re-statements about concerns with trials generally. In addition to the concerns described earlier, further concerns have been raised about realist RCTs, specifically. These relate to whether trials have a "flat" ontology, are able to operationalise a generativist understanding of

causality, and if they can be conducted with sufficiently diverse contexts to allow for the examination of CMOCs.

Philosophical concerns:

Positivism, successionism, and generativism

In addition to the concerns about constant conjunctions (described above) realists have argued that realist trials are 'positivist'. Positivist inquiry values direct and sensory observations, therefore realists argue that its ontological perspective is flat[275] and does not enable researchers to explore causes or mechanisms which are not directly observable.[237] To illustrate their concern, Hinds and Dickson offer the following example: "An analogy is to watch many falling objects and tracing patterns between them, but not looking at deeper unseen causes such as gravity (or natural selection, or molecular structure)."[264, pg. 3]

Bhaskar,[237] Pawson and Tilley[1] and later realists[3, 275] have argued that RCTs are dependent on successionism, which uses correlations to express causation, and that generativism is more appropriate for realist enquiries because it seeks to understand causation by identifying the necessary and sufficient conditions through which mechanisms will be able to operate in order to produce change, and is guided by theory.[199] Within a realist perspective, where causation is generativist, mechanisms will be actualised differently (left latent or activated) depending on the context, social structure, and agency of actors.

Practical Concerns:

Randomisation and controls stifle the ability to test CMOCs:

A concern repeated by numerous realist evaluators is that trials are conducted in homogenous settings without the required contextual diversity to test CMOCs.[1, 3, 4, 250, 263] In a section of his paper called, "the impossibility of a realist RCT" Hawkins argues:

"Context should be an explicit part of what is tested; it is not something to be controlled or neutralised. Understanding the impact of a mechanism free of context is nonsensical to a realist. This has important implications for the possibility of a realist RCT. While a context or mechanism may be isolated or 'controlled' by a traditional RCT, context-mechanism configurations cannot be isolated and randomised; any outcomes we observe will be the result of innumerable mechanisms firing in innumerable contexts." [250, pg. 279]

Defence of RCTs and realist RCTs

In the previous sections I have described the philosophical and practical concerns about RCTs and realist RCTs. Below, I respond to them.

Philosophical concerns:

We have already dedicated an entire publication to exploring whether or not trials are of necessity or in practice, positivist, [259] so only a summary will be provided here (see Appendix 3 for more information). The philosophical and social science literature delineates four key tenants of positivism: 1) scientific knowledge is derived from direct, sensory observation; 2) theoretical terms must equate directly with empirical measurements with no interest in deeper mechanisms of causation; 3) the objective of scientific inquiry is to generate universal laws; and 4) similar methods can be used across the natural and social sciences.

In relation to the first tenet, as I previously argued, trialists normally employ a hypothetico-deductive approach to test theory rather than an inductive approach to build theory. In regard to the second tenet of positivism, we acknowledge that some trials are theorised in terms of the hypothesised association between variables, but that is a weakness in particular studies and not an inherent limitation of the study design, which can be used to test mid-range theories and explore deep mechanisms which do not align with empirical measures. In their aforementioned critique of positivism, Hinds and Dickson use the example of something falling to show that positivist research is unable to explore mechanisms (e.g. gravity) through which the outcome (e.g. something falling) occurred. It should be noted for clarity, that pre-realist scientists were able to develop a theory about what gravity is and how it works, and did not merely measure the speed at which objects fall. Realism is not the only philosophy of science which explores how and under what circumstances phenomena are or are not actualised, and how and why they vary by context.

In regard to the third tenet, few trialists believe their results will be universally generalisable, but some trialists will try to explore generalisability at the theoretical level by assessing how using resources activated mechanisms in certain contexts. Finally, while RCTs did emerge from the natural sciences, RCTs of social interventions are unique in their inclusion of qualitative data which enables the exploration of meaning, reasons, and agency.

Our literature review did not identify successionism as a key tenant of positivism but it is so commonly mentioned that it is worth addressing here. Successionism maintains that correlation between cause and effect (or intervention and outcome) is sufficient for understanding causality. However, this is a post hoc fallacy⁹ as an association is insufficient grounds for assessing causality. Moreover, a lack of an association is not disproof of a causal link because causality may be contingent on a third factor.[276] In closed laboratory conditions, constant conjunctions may be an appropriate way to understand some causal mechanisms. However, in open, complex systems exposures and outcomes are rarely, if ever, constantly conjoined; at best, there may be a tendency towards statistical association which may be contingent on the presence of other factors. In the evaluation of a social intervention that is beneficial at the trial-arm level, there will be some in a study population for whom outcomes improved, some for whom outcomes worsened and some for whom outcomes remained stable. No serious philosopher of science has argued that constant conjunctions are the only criteria upon which causation should be judged. In public health, a much more common approach to assess if a link is likely causal is to assess the connection using the Bradford Hill criteria, which includes consideration around mechanisms, temporality, plausibility and dose-response.[277]

Realists also view successionism as insufficient for understanding causality because it does not account for how contextual factors influence the emergence of outcomes, therefore risking the dismissal of a theory as wrong because a correlation was not found. In their book, Pawson and Tilley suggest collecting data either before and after someone installs security cameras in a parking lot (in a non-experimental fashion) to assess its impact of car-theft.[1] This natural experiment is conducted to examine how, why, or to what extent an intervention (security cameras) produced an outcome, and how that was affected by context. They generated hypotheses and suggested using various data sources to explain the phenomenon via theory. Described another way, they suggest looking for patterns between the introduction of an intervention and changes in outcomes and examine how these are contingent on context, in order to assess if these patterns align with their theory. While their theories are generativist,

⁹ This is also called the “post hoc ergo propter hoc” fallacy, meaning “afterwards, therefore because of it.” If it were not fallacy, there would be no grounds for arguing that, for example, wet roads do not cause rain.

their methods are successionist, albeit detailed in their breakdown by context. The literature's repeated presentation of successionism and generativism as competing approaches is a category error: successionism is about using data to shed light on whether hypotheses match reality (accepting the need to examine the contextual conditionality of correlations) whereas generativism is about using theory, logic and empirical evidence to build or refine our theories about how mechanisms generate outcomes.

Most interestingly, in the parking lot example, Pawson and Tilley propose that the 'publicity' mechanism (displaying symbols of taking crime seriously) and the 'surveillance culture' context (the ubiquitous presence of CCTV in modern-life):

“Is probably the most difficult (and we suspect the most important) to ascertain. A start on this could be made by pursuing some (experimental-type!) variations in the publicity attendant on the arrival of CCTV cameras to see if the specifics of the messages made a discernible difference to outcomes.” [[1, pg. 81, exclamation theirs](#)]

Believing that experimentation is successionist and philosophically inconsistent with realism, they still write that an experimental-type investigation is best way to answer their research question. It is unclear why they argue that quantitative testing is useful as long as it is not of an experimental nature (except under certain, unclear circumstances when experiments are needed and justified). Moreover, they do not clarify why experiments conducted with the added epidemiological and statistical protections offered by randomisation and the use of controls are unacceptable but non-randomised quantitative comparisons are encouraged.

To improve the usefulness of trials, researchers can explore for whom their intervention works best using moderator analyses, which examine how effects varied according to the baseline characteristics of either people or environments. The fact that these analyses are planned a priori reveals that trialists assume that changes to outcomes will not be consistent across the trial. Understanding for whom the intervention worked can help refine both the mid-range theory underpinning an intervention by helping think through why the impacts varied by person, group, or location, and demonstrating areas for potential intervention refinement. For example, multiple WSIs aiming to reduce

bullying have found significant improvements for boys but not girls.[67, 278] This finding can be incorporated into the theory to understand why this may be true and may help refine the intervention to make the benefits more equitable. It also enables relevant parties to make the best decisions for their settings. For example, head teachers may want to find interventions that address the causes for bullying in their school which may be different in all-girls, all-boys, or mixed-sex schools.

Practical Concerns:

Prior to conducting analyses of realist trial data, I am only able to respond partially to their concerns about practicalities. More detailed explorations are contained in the Discussion section[11] (Chapter 7) of this thesis.

Randomisation and control do not stifle one's ability to study mechanisms nor explore contextual contingencies

The use of randomisation and control does not, of necessity, prevent exploration of how mechanisms might interact with context to generate outcomes. As Bonell et al argue,

“Randomisation is merely a practical tool to reduce confounding. It does not fundamentally change the nature of the way we view or research the social world, or affect how we will use comparative empirical data to test hypotheses about mechanisms.”[272, pg. 3]

Randomisation can, for example, be drawn on to ensure fair comparisons when assessing whether particular subgroups of people or sites in the intervention arm experience greater benefits than others by comparing these with similar subgroups in the control arm. Randomisation is not used to impose homogeneity or prevent attention to context.[271]

When baseline differences between arms are identified, statistical adjustment for them does not imply that researchers are acting as though those factors do not matter; they simply enable trialists to account for them when estimating an intervention's added-value. Statisticians can also address this concern by checking for interactions before deciding to whether to adjust. Moreover, having baseline equivalence does not mean that there is no variation on factors such as age or ethnicity within these groups, but rather that there is approximately similar variation within each arm. It is, in fact, these differences which enable researchers to conduct moderator analyses, which is crucial to understand for whom interventions work, a key realist question.

In their example of prisoner education reducing recidivism rates, they use historical “usual treatment” data as a comparator rather than a contemporaneous control arm.^[1] Why this comparison, saddled with baseline non-equivalence and no possible way to understand the context of their comparator arm, is seen as legitimate^[272] and data from an RCT is unacceptable has not yet been explained in the ongoing debate. As described earlier, realist evaluators have expressed concerns that trials are tightly controlled. If by ‘control’, they mean regulated, it is accurate to say that strict fidelity is often required for clinical or efficacy trials but it is not a necessary or even sought-after feature of pragmatic, effectiveness trials. If by control, realists are concerned about interventions being implemented in homogenous settings, researchers can purposefully recruit from a wide variety of individuals or groups and stratify their sample to enable them to examine how the intervention works in different contexts.^[278] Without a comparator arm it is impossible to estimate the intervention effect size across the two groups in order to understand if the intervention creates a net-benefit compared to usual practice.^[207] Moreover, it is possible that small quantitative differences will get drowned out in study designs that do not adequately control for bias, confounding and random error, or even more so in qualitative only designs that rely on observations and participant accounts to understand the interconnections between social phenomena.

Neither the use of randomisation nor the creation of control groups erase or ignore social processes. They are simply strategies employed to test whether - all other influences being present (not excluded or controlled) and approximately equal (via randomisation) - the mechanisms triggered by local use of intervention resources add to or interact with all the other present mechanisms to generate different outcomes in the intervention arm compared to the outcomes resulting from the social mechanisms present in control arm.^[272] Fletcher et al suggest that realist trials may draw greater attention to the control arm and focus analytic attention on the mechanisms present there which are similar to the intervention --- an angle rarely considered by trialists.^[279]

RCTs being insufficiently theorised is not an inherent shortcoming of the study design

While many evaluations value assessing outcomes over improving mid-range theory, this is not always the case. Studies within the fields of economics and psychology have

historically been more focused on testing and refining theory[280], even if they have not always been attentive to context.[281] Realist evaluators have called for more research to account for how the context of an intervention's implementation may affect how change occurs. Often called "hidden moderators"[282] the contextual differences between an intervention's original location and its replication may have profound influences on whether or not the intervention "works" or whether findings can be replicated.[1] These hidden moderators are crucial because the focus of CMOCs is generally focused on how subtle contextual differences affect outcomes.

RCTs can be used to assess mechanisms

Realist evaluators have asserted that trials seek to exclude all mechanisms except those introduced via the intervention. Even within the natural sciences, for example, biopharmacology or climate science, researchers are unable to isolate their phenomena of interest and consequently, they also work in open, complex systems, although I would recognise that social systems may be even more complex because of the added factor of human consciousness and agency.[283] Unlike laboratory experiments involving closed systems, RCTs of social interventions are not trying to remove, control, or keep constant other potential mechanisms. Trialists merely want to properly account for these. The control group exists precisely so that trialists may understand how an intervention mechanism exerts effects in interaction with other mechanisms.[272] Both trial arms contain an approximately equal diversity of contexts and people, and therefore have comparable mechanisms operating. In the case of the intervention group, these are also accompanied by mechanisms that may arise as a result of using intervention resources. Therefore, by comparing the intervention to the control arm, trialists are assessing the extent to which the intervention changes the activation or de-activation of other possible mechanisms. It is a measure of *added* or *marginal*, not *isolated* value. Of critical importance is that trialists are absolutely not trying to demonstrate that an intervention in isolation causes an outcome, contrary to what realists suggest.[3, 235, 250, 269]

The critique that RCTs will be unable to detect different causal mechanisms in different contexts is difficult to understand as trialists can use the same types of data and analytic methods to explore mechanisms as are commonly used in realist evaluations. Moreover, the same resources offered to different people or organisations as part of an intervention are hypothesised to 'work' via different mechanism in different contexts,

and RCTs provide a way in which those mechanisms can be explored. In practice, researchers may find that their intervention was insufficiently implemented to carry out the analysis but that is not a reflection of the methodological shortcomings of RCTs.

Most of the debate about realist evaluation and realist trials has focused on how researchers examine mechanisms, or the way that processes are enacted via intervention resources to generate outcomes. The discussion of other intersections of trials and realism, specifically how interventions are enacted in the first place, is less well addressed. A realist approach to understanding fidelity could help sharpen the focus of trial PEs by making sure that they explore how agents use interventions and how various structures amplify or dampen the activation of mechanisms. It might also enable theorising about how these interactions will vary by context, which includes interactions between structure (e.g. policies, resources, etc) and agency.

RCTs are not only concerned with attribution

Determining the effect size of an intervention is undoubtedly an important finding in an RCT. It is useful to know overall whether an intervention is beneficial, ineffective or harmful across contexts. However, being able to report an overall effect size does not preclude trialists from conducting more realist-aligned analyses. Trial protocols can be designed to test CMOCs, explore the contingent activation of mechanisms, and contribute to the refinement of mid-range theories and interventions.

RCTs do not stifle one's ability to test CMOCs

In response to the belief that contexts need to be uniform in RCTs, it should be noted that researchers can ensure that intervention and control arms contain similar variability of contexts, for example via stratification.^[4] The diversity of populations and contexts in each arm can be a methodological strength of the design and may actually bolster realist approaches.

Fletcher et al propose using purposive sampling (which determines the pool from which to randomise rather than replacing randomisation) within pilot studies to ensure “sufficient diversity in aspects of context that have been pre-hypothesised to affect feasibility, acceptability, and causal mechanisms.”^[279, pg. 294] In order to do this better, evaluators need to engage more actively with theory, conduct more qualitative research and focus on appropriate quantitative methods. However, it must also be recognised

that focusing RCTs on questions of how intervention mechanisms operate differently by context will require larger studies.

Case study to assess feasibility, intellectual coherence, and usefulness of a realist trial

This thesis is a case-study exploring whether or not realist RCTs are feasible, philosophically coherent and generate useful findings. This PhD uses data from the INCLUSIVE RCT of the LT intervention which assessed the effectiveness, cost-effectiveness, and the process through which LT reduced bullying and improved physical and mental health among secondary school students in the southeast of England. Both the INCLUSIVE evaluation and the LT intervention are described in greater detail in the following chapter.

Study designs and methods have well-known and commonly accepted limitations. Purely quantitative data cannot account for meaning or interpretation just as qualitative data cannot be used to ascertain incidence, prevalence or measures of association. RCTs are unsurpassed for establishing effect estimates but they have frequently been conducted with insufficient attention to refining theory. To improve RCTs, researchers should collect sufficient data to test CMOCs and improve intervention and mid-range theories. RCTs should also include integrated qualitative research to refine existing theories about mechanisms and how these appear to be influenced by structure and agency, and to explore unexpected findings. They would also be improved by making greater use of the control arm to understand and potentially test whether or not there is evidence that mechanisms hypothesised to generate outcomes might operate in similar ways in the absence of provision of intervention resources. Qualitative data can be used for developing theory but are much less useful for testing whether or not the theory is empirically sound. While hotly contested, it may be possible to bridge differing epistemologies and paradigms to open the black box of how and why mechanisms activate or remain latent in different contexts while still protecting the effect estimates from unnecessary bias, confounding, and chance.

The original publications explaining how and why INCUSIVE was a realist trial[[2](#), [195](#), [270](#)] started a vigorous debate.[[3](#), [4](#), [234](#), [259](#), [271](#), [272](#), [279](#)] While the philosophical

concerns can be debated, the empirical questions remain more difficult to address without first conducting a realist trial.

Aims, objectives, and research questions

It is in this context that this thesis aims to conduct analyses of the first realist RCT and provide evidence regarding whether or not realist trials enable deeper understanding of how an intervention works, for whom and under what circumstances, and whether or not the structure provided by the trial's design limits our ability to evaluate a complex social intervention. The thesis will also assess whether realism and RCTs are philosophically cogent, practically feasible and produce results which are theoretically grounded and practically useful.

To meet this aim, this study will evaluate fidelity and explore how and why this varied by school, use qualitative data to develop hypotheses about mechanisms including their contextual contingencies, and test them quantitatively to understand how context and mechanisms co-produce or negate effects on outcomes.[\[214\]](#) Finally, the thesis will close with a discussion about whether realist trials are possible, whether analyses of trial data elucidated findings that may have otherwise been undiscovered, and address the methodological challenges encountered when applying novel approaches.

This thesis seeks to answer empirical questions, but is ultimately methodological in its orientation.

In order to achieve this aim, the following questions will be addressed:

- 1) With what fidelity were Action Groups (AGs) implemented and what role did AGs play in coordinating the intervention so that components could interact synergistically? How did this vary by school context and to what extent were AGs supported by the provision of external facilitators and data on student needs and how was this effected by the school environment? What approach did AGs take to local adaptation? Were AGs acceptable to, and engaging and empowering for their members?[\[8\]](#) (Chapter 4)
- 2) How did intervention participants describe the school context, the processes involved in LT participation and the consequences of these? How did such accounts vary between schools, and what conditions relating to schools, staff or students seem to explain these variations? What do these findings suggest about

our a priori theory concerning the mechanisms via which the intervention might generate outcomes relating to bullying, aggression or mental and physical health, and how these are affected by context?[284] (Chapter 5)

- 3) Do analyses using qualitative comparative analysis support the plausibility of the CMOCs and is there evidence that those mechanisms operate via other resources in schools without LT resources? Is qualitative comparative analysis a useful method for understanding the complexity and the variety of causal pathways possible within an RCT?[10] (Chapter 6)
- 4) Reflecting on the above process, are realist RCTs possible? To what extent did the methods used in the analysis of trial data align with realist evaluation or answer the nuanced questions fore-fronted in realist inquiries? What challenges did we encounter, how did we respond, and were we able to generate useful findings?[11] (Chapter 7 and conclusion)

CHAPTER 3: METHODS

This chapter begins with a detailed description of the theory of human functioning and school organisation.[7] I then describe how this mid-range theory underpinned LT, its theory of change and the INCLUSIVE trial. I then briefly describe the intervention before I focus on the ontological and epistemological assumptions used throughout this thesis and their connection to realism. Each chapter's methodology, including thematic analysis, dimensional analysis and qualitative comparative analysis, is then described in more detail than was possible within peer-reviewed journal manuscripts. Finally, this chapter closes with a description of my reflexivity and positionality.

Mid-range theory underpinning Learning Together and INCLUSIVE

The sociological theory underpinning LT and INCLUSIVE is Markham and Aveyard's theory of human functioning and school organisation[7] (see Figure 1, below). Building from Nussbaum's ideas of human functioning, the theory postulates that all humans have both *fundamental needs* (e.g. clean water, food, warmth) and *essential capacities* (e.g. thinking, imagining, having empathy).[285] Of all human capacities, Markham and Aveyard argue that *practical reasoning* (one's ability to think and imagine) and *human affiliation* (one's ability to build and maintain relationships) are pre-requisites for autonomy. By learning to reason, people are able to see other perspectives and can make pro-active and adaptive decisions which support their health. Likewise, people with sufficient human affiliations have meaningful friendships and a sense of belonging, purpose and social support, which in turn promote health. Developing these capacities has been described as the key developmental task of adolescence.[286] Schools teach practical reasoning via the curriculum and develop affiliation, both in relation to students' connection to the school itself and in their relationships to peers and staff. Therefore, if students are sufficiently committed to school so that their practical reasoning and affiliation are developed, schools become sites for health promotion.

Markham and Aveyard also draw on Bernstein's theory of cultural transmission[287] to theorise whether different students will commit to a school's *instructional* and *regulatory orders* in order to benefit from schools as sites for learning and health promotion. The instructional order prepares students for work through knowledge and skills, and is taught primarily through subject-specific teaching. The regulatory order is

concerned with character and personal conduct, and seeks to ensure shared values within the school community. Bernstein argues that the values which students are meant to learn and internalise at school are the product of the “controlling” or middle class [288] but that not all children who attend school will share those values.

Students can be classified based on their commitment to the instructional and regulatory orders. When students can meet the demands and comply with both orders, they are said to be ‘committed.’ These students are often middle class. When students are not able to meet the demands of, or do not comply with, either order, they are considered ‘alienated.’ Alienated students are commonly from working class backgrounds. ‘Estranged’ pupils, commonly middle class, are committed to the regulatory but may not be able to meet the demands of, or do not want to comply with the instructional order. Finally, ‘detached’ pupils are those who follow the instructional order but do not understand the regulatory order (commonly working class) or do not share it (commonly middle class).[7] Markham and Aveyard argue that ‘alienated’ and ‘detached’ pupils may fail to develop self-esteem or a sense of belonging if there is inconsistency or incompatibility between their values and those of the school, and may consequently seek to build those bonds with ‘youth culture(s)’ or remain loyal to their community’s values which may differ from the school’s.

Schools can improve commitment, and consequently students’ health, by changing the school environment to meet students’ needs. Drawing again on Bernstein, Markham and Aveyard theorise that the potential of schools to promote commitment and hence health lies in how they are *classified* and *framed*, which impacts upon how students learn and affiliate, respectively (see figure 1, below). Classification relates to the way the school is organised and the boundaries it maintains between itself and the community it serves, between teachers and students, between different students (for example via class sets and year groups), and between different teaching subjects. When a school has strong boundaries, it can be described as ‘strongly classified.’ Eroding boundaries may improve students’ affiliation with the school and reduce the risk of becoming alienated or detached because the values of the school would more closely align to the values of the community the students are from. When a school has strong student/staff boundaries, students would not be allowed to participate in decision-making processes. Classification can therefore be reduced via increased shared decision-making (e.g.

student councils or AGs). Participating in school-wide decision-making is hypothesised to improve students' practical reasoning by learning about other people's experiences and ideas[7] and may improve affiliation by giving them the opportunity to empathise with staff who also serve on student councils or similar programmes. Reasoning and affiliation can also be bolstered by weakening classifications between students via mixed-year tutor groups, task-sharing and/or mentoring.

Framing relates to the communication and pedagogic strategies used by the school[7] and is primarily concerned with re-centring provision on student needs and preferences. Like classification, Markham and Aveyard suggest that schools will be more health promoting when framing is weakened. To weaken framing and improve health, schools can allow self-guided learning, student-led pace-setting and practice participatory seminar-style teaching over traditional didactic lectures. The weakened framing would facilitate students' practical reasoning (via independent or collaborative problem-solving) and affiliation skills (via stronger relationships with peers and staff) and hence, their health.

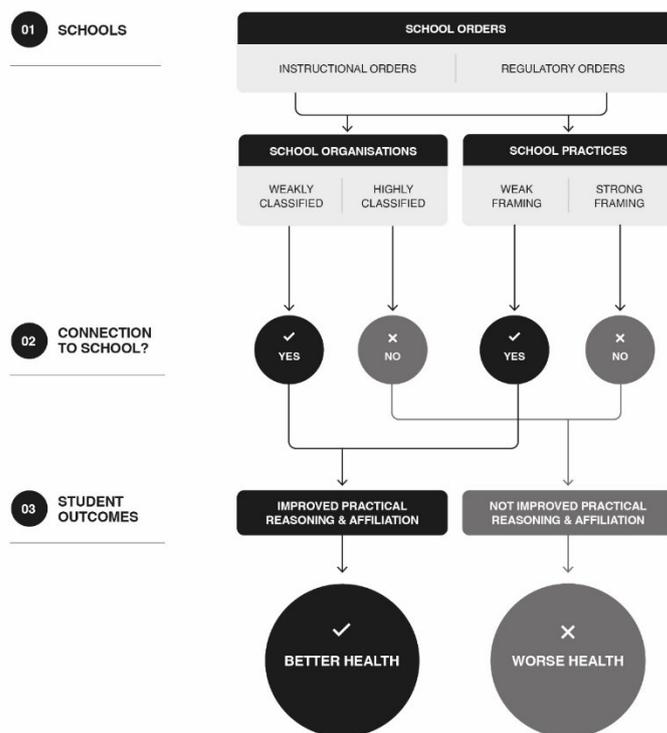


Figure 1: Theory of human functioning and school organisation, based on Markham and Aveyard[7]

Learning Together’s theory of change and logic model

Based on the theory of human functioning and school organisation, a theory of change and logic model were developed to understand *how* LT resources (described below) were intended to enable school staff and students to modify the school environment, decrease bullying and aggression, and improve mental health and well-being[5] (See Figure 2, below).

The theory of change postulates that, using the resources provided, schools will convene an AG, review rules and policies with input from students, implement RP and the SEL curriculum, and decide other locally appropriate actions based on student needs. These processes were hypothesised to reduce both the boundaries between and among staff and students, and between academic learning and students’ broader social development, as well as weaken framing of learning and behaviour management to be

more student-centred regarding teaching, discipline, school management and organisation.

These changes to boundaries and framing were postulated to improve students' commitment to the instructional and regulatory orders by making school more engaging. By increasing student and staff ownership of the rules, we hypothesised that commitment to school would increase,^[5] therefore, engaging more students in school, elevating aspirations, teaching key affiliation-based life-skills (e.g. managing emotions and effective communication), and building more trusting, empathetic and warm relationships which were thought to improve students' capacity and desire to make health-promoting decisions. Finally, these impacts were hypothesised to reduce bullying and aggression, improve quality of life (QoL), and reduce substance use, sexual risk-taking, contact with the National Health Service (NHS) and police, truancy, and exclusions. Increasing student commitment was theorised to improve practical reasoning and affiliation as well as ensure that these were oriented towards pro-social and pro-school behaviours.

The theory of human functioning provided some suggestions about how mechanisms would vary by context, for example, the greater need (and potential) for eroding boundaries in schools serving students from predominantly working-class backgrounds, but was generally under-theorised in relation to the contextual conditionality of how schools could better promote health. Therefore, the theory was augmented with additional hypotheses contained in the study's PE protocol (see Appendix 4 for more detail). For example, these included that "The intervention will be implemented with better fidelity in schools that include students with varying degrees of educational engagement in its activities (e.g. action groups), including students who have a history of, or considered likely to be involved in bullying behaviours" and "LT schools will report higher rates of student life skills and warm, trusting and empathetic relationships and lower rates of student involvement in anti-school peer groups by follow up 1 and 2."

Below, I describe the intervention components that LT included in terms of the resources provided and the activities that these aimed to enable.

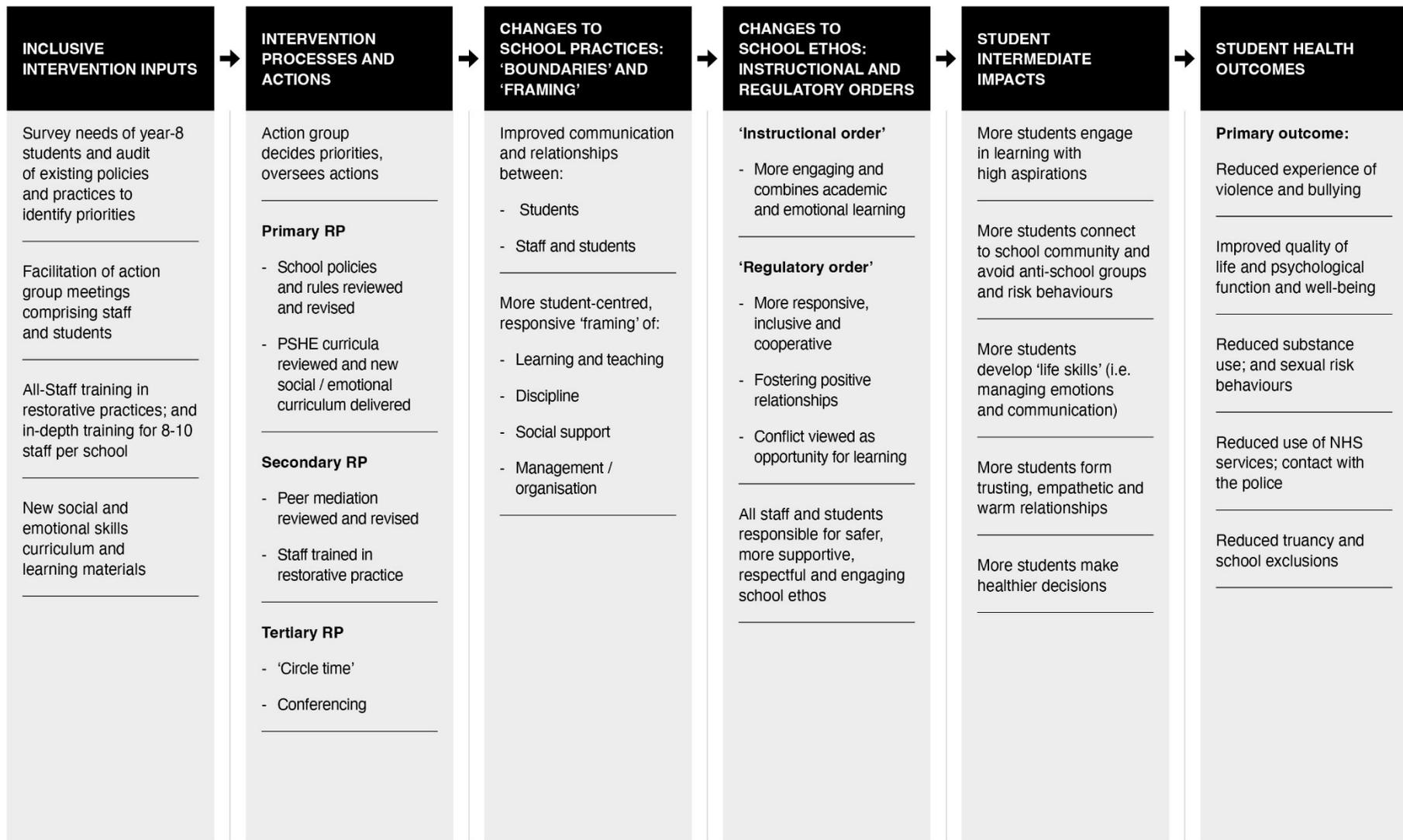


Figure 2: Logic model for Learning Together

Description of Learning Together

LT was developed to provide schools with the resources they would need to increase students' sense of commitment to schools and therefore reduce bullying and aggression.

Action groups

In the first two years of the intervention, schools were provided with a manual and an external facilitator to enable them to convene and run AGs. The facilitators worked with an intervention lead at the school (ideally someone on the school's senior leadership team (SLT)) to help lead the AG. Each facilitator had professional experience in school leadership or organisational change and was trained in the processes they were meant to guide.

AGs were intended to include a minimum of six students from the intervention cohort year-group (year 7 at the start of LT) and six staff, including one member of the SLT and at least one member of teaching, pastoral, and support staff, and where possible a specialist health-focused member of the school including a counsellor. The SLT member was hypothesised to be a critical conduit through which AG decisions could be communicated directly to the wider leadership team. The AG was intended to meet six times per year (or roughly once per half-term) and was primarily tasked with delivery of key intervention activities including:

- reviewing and revising rules and policies related to behaviour management and discipline, in order to incorporate restorative principles;
- implementing RPs within the school to prevent and respond to instances of bullying and aggression;
- implement tailored actions to address school-specific issues and priorities; and
- delivering the SEL curriculum for students in years 8-10

The facilitator worked with the AG lead to ensure that all meetings were scheduled and minuted, that students were able to participate and were listened to, and focused the group on prioritising, deciding, and implementing local actions. These actions were further facilitated by an additional resource: a needs assessment report (NAR), which contained findings from the baseline survey (conducted before randomisation), a 12-month survey at the end of year 8 (conducted in intervention schools only) and a 24-month survey. In the final year of the trial, the external facilitator was removed and the

AG lead was expected to manage the group independently. This was done so that researchers could assess whether or not the intervention was feasible without external support and could be self-sustained.

Training and Restorative Practices

Professional trainers, accredited by the UK's Restorative Justice Council, acted as the key resource to enable implementation of restorative practice activities. All staff in interventions schools participated in a half-day training on primary RP. Each school was also asked to send five to ten staff members on an intensive three-day training course in secondary or responsive RPs.

Primary RP was intended to be delivered consistently by all staff throughout the intervention, using restorative language (respectful language which may either challenge poor behaviour or support pro-social behaviour in a way that maintains and strengthens the relationship) and circle time. During circle time, groups of students (e.g. form groups) could discuss feelings and address challenges before they escalated into problems. Responsive RPs involved specially trained staff bringing together victims and perpetrators of bullying or aggression to discuss the incident(s) and the impact it had on the victim, and give the perpetrator the opportunity to learn from their behaviour, make amends and prevent future harm. When required, schools could also involve parents, police or relevant social agencies. These skills could also be applied to non-bullying problems, such as conflicts between friends.

SEL curriculum

SEL lessons were designed to either be implemented by schools in 'stand-alone' lessons, primarily envisaged to be personal, social, and health, education (PSHE) lessons, in tutor time or incorporated into other lessons (e.g. drama or English). Each year the school received resources in the form of a new curriculum (slides, lesson plans) to be used with the cohort of students being followed by the trial. Each year's curriculum contained between 5-10 hours of lessons in RP, health relationships, and social and emotional skills. The curricula was based on the curricula used in the Gatehouse Project.^[84]

INCLUSIVE

The effectiveness of LT was evaluated through the INCLUSIVE RCT, a three-year, superiority, parallel-group, cluster-RCT with 20 intervention and 20 control state-

sponsored schools from London and southeast England, conducted 2014-2017. The trial's full protocol[5] and later update[6] have both been published (and are provided in Appendices 5 and 6). Ethical approval was obtained from Institute of Education Research Ethics Committee (18/11/13 ref. FCL 566) and the University College London Research Ethics Committee (30/1/14, Project ID: 5248/001).

The INCLUSIVE trial had three primary (non-realist) research questions:

1. Is the Learning Together intervention more effective and cost-effective than standard practice in reducing bullying and aggression among 12- to 15-year olds in English secondary schools?
2. Is Learning Together more effective than standard practice in improving students' QoL, well-being, psychological functioning, and attainment, and reducing school exclusion and truancy, substance use, sexual risk, NHS use, police contacts among students, and improving staff QoL and attendance and reducing burn-out?
3. What pre-hypothesised factors moderate and mediate the effectiveness of Learning Together; including, do effects vary by socio-economic status and sex?

To be eligible for inclusion, schools had to be within the state education system (including community, academy or free schools) and could not be private (for-fee) schools, pupil referral units or schools exclusively for students with special education needs. Secondly, the school's most recent Ofsted evaluation had to be "requires improvement" or better. Schools graded as "inadequate" or "poor" were deemed unlikely to be able to prioritise an intensive intervention while also trying to address this inspection rating.

Schools were recruited between March and June 2014 and had to be within a one-hour travel time from central London.

Immediately after baseline data collection, the clinical trials unit (CTU) at the London School of Hygiene & Tropical Medicine (LSHTM) allocated schools 1:1 to either the intervention or control arm of the trial. To facilitate baseline similarities between trial arms, randomisation was stratified based on three school-level determinants of violence:

- 1) single/mixed sex;

- 2) school deprivation as measured by the percentage of students eligible for free school meals (FSM) (low/moderate 0-23% and high >23% when the English median 23%); and
- 3) school-level educational attainment ('best eight value added' (VA) General Certificate of Secondary Education GCSE exams (English mean 1000) score).¹⁰

A statistician from the CTU communicated the allocation to the PE team, who communicated it to schools and the intervention delivery team. The lead statistician (Professor Elizabeth Allen), the co-principal investigator responsible for the outcome evaluation (Professor Russell Viner), fieldworkers collecting outcome data and data-entry professionals remained blinded throughout the study. Blinding participants and PE researchers was not possible.

While LT was a whole-school, universal intervention that aimed to improve the health and well-being of children from age 11-16, data for the evaluation was only collected from the cohort of students who were in year 7 at baseline (age 11-12) who were followed for 36 months until the end of year 10 at endline (age 14-15). Teaching and support staff were followed up for the same time period. Because LT sought to improve the school environment, students and staff were included at endline even if they had not been present at baseline. Study partners at UCL's Institute for Child Health, led by Professor Viner, organised survey data collection pertaining to the outcome evaluation. School staff and students completed surveys in year 7 at baseline and 36 months later at the end of year 10 to measure primary and secondary outcomes. Additional student and staff surveys were conducted 24 months into the trial to provide the team with data on intermediate outcomes and intervention processes which would enable mediation analyses.

Within INCLUSIVE's intervention arm, six schools were purposively sampled to encompass the diversity of schools within the trial in relation to the percentage of children eligible for FSM (above or below 2012 national average of 16.3%) school type,

¹⁰ The median VA-scores for schools is publicly available. The VA score for each student was calculated by subtracting their unique output score at the end of key-stage 4 from their starting output at either the end of key stage two or three. Schools that neither increase nor decrease their outputs were given the average score of 1000. Schools that improved students outputs were given a score greater than 1000 while those whose students performed worse than expected were given a score of less than 1000.

facilitator, and the extent to which their facilitator described the school as being responsive to intervention activities 3-months post intervention commencement (described as being highly, somewhat, or poorly responsive).

Comparator arm: Control Schools

Schools that were randomised to the control arm were meant to continue with their normal practice and received no intervention resources. No intervention and control schools were geographically close to each other. Head teachers and some members of staff were aware that their school was participating in the INCLUSIVE trial but they were not told what LT entailed.

Process evaluation

The integral PE followed the MRC guidance for process evaluations of complex interventions[216] and was informed by realist evaluation.[1] It was comprised of five separate domains: 1) context; 2) fidelity; 3) participation, reach and dose; 4) reception and responsiveness; and 5) intermediate outcomes.

As this thesis draws primarily from the PE, details of the data sources used will be described within each chapter, as relevant.

Summary of results from INCLUSIVE trial

After 36-months, bullying scores were lower in the intervention than control arm (Gatehouse Bullying Scale[GBS][289] adjusted mean difference -0.03 , 95% CI -0.06 to 0.00 ; adjusted effect size -0.08). There was no significant reduction in aggression (Edinburgh Study of Youth Transitions and Crime [ESYTC] scale[290] adjusted mean difference -0.13 , 95% CI -0.43 to 0.18 ; adjusted ES -0.03). Amongst secondary outcomes, there was no difference in bullying victimisation or aggression at 24 months. Students in intervention schools had higher QoL as measured by the Paediatric Quality of Life Inventory [291] at endline (adjusted ES 1.44 , 95% CI 0.07 to 2.17) and psychological well-being scores (as measured by the Short Warwick-Edinburgh Well-Being Scale (SWEWBS)[292] 95% CI 0.00 to 0.66) and lower psychological difficulties (Strengths and Difficulties Questionnaire [SDQ][293] 95% CI -0.83 to -0.25) than students in control schools. Intervention schools also benefitted from lower levels of conduct, hyper-activity, emotional, and peer problems.[278, 294]

My contribution to the study team as a research fellow

I joined the INCLUSIVE study team as a research fellow leading the PE in April 2016. I was responsible for collecting all PE data, liaising with facilitators and teachers, and coordinating research activities with UCL staff and professional support services, including administrators and transcriptionists. The trial and PE protocols and data collection tools had already been designed and 20 months of data had been collected. I completed the data collection in the second school year as specified in the protocol. During the students' summer break, Professor Bonell and I modified the PE protocol to make data collection less arduous on staff, in order to improve the response rate and reduce respondent fatigue. Key changes included decreasing the frequency of the RP survey from termly to annually. It was also during this time that I began the qualitative analysis for the National Institute for Health Research (NIHR) funder report.

Thesis Methods

In the below sections, I show the clear distinction between my work as a research fellow on the INCLUSIVE RCT and my own unique intellectual contribution via the work contained in this thesis. I then describe ontological and epistemological assumptions underpinning this work before describing the methodologies used in each chapter.

Key distinctions between the trial and this thesis

The outcome and process evaluations for INCLUSIVE were both guided by detailed protocols.[\[5, 6\]](#) The key findings from the PE were reported in the trial's main findings paper[\[278\]](#) (see Appendix 7) and in the study report[\[294\]](#) (see Appendix 8).

When I began to analyse the qualitative data collected as part of the PE in the summer between the second and third year of the trial, I realised that the data collection tools were primarily oriented towards answering questions about context and fidelity and less well able to answer questions about mechanisms. Concerned about the implications this would have on my PhD, I modified the data collection guides in the final year to focus on how and why participants think change occurred.

According to the protocol, the PE team was tasked with examining intervention fidelity, reach and acceptability. As the research fellow on the team, I developed the fidelity tool, collated the requisite data, and analysed it to enable the exploration of reach and acceptability, and how these varied by school. The analysis contained in this thesis is closely related but focused on how groups co-ordinated the intervention's various

components differently, the processes through which schools adapted the intervention to meet their needs, the role of facilitators, and whether and to what extent members on the AG were empowered by their participation. Therefore, the analysis explores how intervention resources were used and how this varied by context. This analysis was a pre-requisite to exploring which possible mechanism may have been activated and how this varied by context. Moreover, AGs are commonly used to coordinate whole-school health promotion interventions but this had not been evaluated in detail in the literature.

The PE protocol establishes a clear process through which mechanisms could be identified and assessed. Before the trial began, hypotheses had been developed about how the LT might work and analyses including key moderators, mediators, and possible harms had been planned. The original plan involved building CMOCs based on LT's theory of change and logic model and later using PE data to modify these a priori hypotheses. The CMOCs could then be tested and the mid-range theory would be updated in light of empirical findings.[\[270\]](#) I decided early on in my PhD that refining and testing the a priori hypotheses would not give me the qualitative data analysis skills that I wanted to develop during my PhD. While they were not directly incorporated into my work, I was familiar with them and their relation to the theory of change, which was still the basis of my work. Instead, of refining the a priori hypotheses, I developed hypotheses iteratively from the qualitative data and later tested them using qualitative comparative analysis.

Ontological and epistemological assumptions

The philosophy of science underpinning this thesis is critical realism, developed by Roy Bhaskar.[\[237, 241\]](#) Bhaskar's overarching question driving the development of critical realism was, "to what extent can society be studied in the same way as nature?" [\[237, pg. 1 emphasis his\]](#) Critical realism is built on three key assertions called *ontological realism*, *epistemic relativism*, and *judgmental rationality* (see Figure 3). *Ontological realism* is the belief that things and events are real and exist independently from human perception and knowledge. While people will ascribe different meanings to things and experience them in distinct and competing ways, they remain fundamentally real. *Epistemic relativism* is the recognition that human knowledge does not necessarily correspond to the way things are. Science is a social process driven by the agency of scientists

choosing to explore certain phenomena, but is inherently limited by the available resources and capacities at any given time. Knowledge then is always finite, contextual, relative and perspectival: people ascribe meaning to events based on their perspectives and objectives. Finally, this partial and perspectival knowledge is subject to *judgmental rationality*, or one's ability to evaluate competing claims about knowing and knowledge. [295] For example, using the best knowledge and resources available in his time, Aristotle argued that the Earth was the centre of the universe. It wasn't until the early 17th century that the scientific, methodological, cultural, religious, and material culture had changed sufficiently to demonstrate a heliocentric world view as articulated by Galileo. These two competing views on the nature of the universe was subjected to a widespread exercise in judgmental rationality in which people evaluated the evidence and gradually accepted a more accurate depiction of the universe. This process of adjudication is possible because the sun and its place in the universe are real, and not merely social constructions (ontological realism). As methods and resources continue to change and people make continued advancements in science, more of what is currently accepted as knowledge generated through social processes of scientific research (epistemic relativism) will be challenged and further subjected to the continual process of judgmental rationality. Within critical realism, *knowledge is judgement* and our beliefs about knowledge have a profound impact on how we think. Therefore, critical realism is the means through which we can engage with science better, because science is an ongoing exercise in judgmental rationality.



Figure 3: Diagram of key realist ontological beliefs

A theoretically crucial demarcation between realism and other scientific epistemologies, is the idea of *stratified reality*. In brief, Bhaskar postulates that reality functions on three interconnected levels: the ‘empirical’, the ‘actual’ and the ‘real.’ The ‘empirical’ level refers to the most superficial level of one’s experience: a flawed and incomplete window into the realm of the ‘actual,’ where events and interactions happen, regardless of whether one is aware of them. The deepest level is the ‘real.’ In this realm, unobservable, invisible and ubiquitous mechanisms and processes occur which causally influence events in the realm of the ‘actual’ and our awareness and experience of them in the ‘empirical’ realm.¹¹ These realms can be summarised as the realms of

¹¹ Kant describes noumenons as posited objects or events which exist independently of human perception.[296] The noumenon was later expanded upon by realists[297] who describe them as mechanisms, or things which are not perceivable but that which exerts causal powers that may render themselves visible through their consequences. For example, we perceive through

‘experiences’, ‘events’ and ‘causes’, respectively. [298] This layered conceptualisation of reality allows for the interrogation of the “‘underlying’ mechanisms...to explain how things work by going beneath the surface (observable) appearances and delving into their inner (hidden) workings.” [1, pg. 65] The implications of this are two-fold.

Ontologically, causal mechanisms may exist but not be activated, and methodologically in complex, open, systems, scientists may not yet have the methods or knowledge to separate causation from noise. Hence, the aim of realism is to understand how mechanisms work and how the same inputs may lead to different outcomes in different contexts.

Believing that causal mechanisms themselves are unobservable because they are in the realm of the ‘real’, realist evaluators suggest that engaging with scientific theory will enable thinking through what data are needed to test hypotheses about causation. Data can be gathered which may support or refute theories, but realists are clear to caution that researchers should not confuse data with the phenomena it is meant to measure. For example, collecting trial data on levels of literacy amongst low-income children is not the same thing as the actual literacy capacities of students or the mechanisms through which literacy improved.

Through his description of ontology, Bhaskar sought to shift the locus of scientific inquiry away from empirical, observable phenomena themselves and onto the “structures that generate them.” [237, pg. 19]. His key premise is that the traditional parties to arguments about the philosophy of science have both been unhelpful. On one side are those whom he calls *the naturalists*: positivists who carry out experiments or who conduct quantitative research or analysis. On the other side, are the *anti-naturalists* or *the relativists*: hermeneuticists who try to elucidate meaning but do not study how change happens. Bhaskar demarcates realism as “diametrically opposed” to positivism and one that “departs in fundamental respects from the hermeneutical tradition.” [237, pg. 18] This shift from assessing outcomes or meaning to understanding mechanisms and the conditions needed for their emergence became the key difference in the focus of realist-informed evaluations.

our senses a pencil falling from a desk (a phenomenon) but the noumenon or mechanism behind that (gravity) is invisible and only perceivable through its’ consequence.

Bhaskar wrote in great detail about why current philosophies of science are wrong but unfortunately, he provided little detailed instruction about how science should be done better in practice. One challenge to operationalising realism is the conflicting views about whether realist studies can ever use quantitative data, and if so when and how. As just mentioned, he was explicit that realism is “diametrically opposed” [237, pg. 18] to positivism but we cannot say with certainty if, by positivism, he included the use of any quantitative methods or data from studies that include experiments.

Pawson and Tilley, who pioneered operationalising realism via realistic evaluation, declare that realist evaluations are “methods neutral” [1]. Another prominent realist, Andrew Sayer writes that quantitative data “can be appropriate for different and legitimate [research] tasks” [275pp. 19] but his writing indicates that he understands quantitative data to have very limited use within explorations of causality. [275] He argues that:

“there is more to the world...than patterns of events. It has ontological depth: events arise from the workings of mechanisms which derive from the structure of object. This contrasts with approaches which treat the world as if it were no more than patterns of events, to be registered by recording punctiform data regarding ‘variable’ and looking for regularities among them.” [275, pg. 15]

This study is based on similar beliefs: *on their own*, statistical associations are insufficient for understanding causation. However, quantitative data sources can be used in nuanced ways to understand who is experiencing change, how, why, to what effect and extent, and whether associations are contingent on one or more other factors. Data from trials does not need to be limited to examining effect sizes and should be collected to assess plausible causal pathways, which should be built from relevant and appropriate theory. Because data from an RCT have the approximately balanced distribution of potential moderators or confounders across trial arms, the comparisons have the benefit of being more fair than had they emerged from non-random comparators. Quantitative data should be used to test hypotheses arising from theory so that the theory might be refined. [212] It is in this way that scientists can make an “epistemic gain” [299] or add a brick in the wall of knowledge.

Therefore, quantitative analysis may help us understand how some change is occurring in schools, and how context affects which mechanisms activate or are over-ridden, leading to outcomes improving, worsening, or remaining stable. Schools are not treated as uniform, closed systems which will respond in predictable ways to new resources. It is for these reasons that we used data from diverse schools to assess how different environments, teachers and students with different needs and priorities describe how change occurred, and test them to see if those theories were helpful.

In many social science study reports, qualitative data are interpreted through interpretivist or constructivist epistemologies while quantitative data might be seen through a positivist lens (although it is uncommon in quantitative publications to have the author's epistemological position stated and many quantitative researchers explicitly reject positivist tenets). The maintained separation of different epistemological positions is generally represented in academic outputs through separated publications, each using data from either qualitative or quantitative sources.[213, 300] When these epistemological silos are maintained, studies may be described as multi-method because they lack *integration*, the distinguishing characteristic of mixed-method research.[212, 213] [ENREF 163](#)

Despite its controversy in realist circles, mixing methods is logically consistent with a realist approach to knowledge and may establish a dialogue between data sources where the findings from one inform the analysis of this other. In this thesis, for example, the findings from qualitative papers are then tested quantitatively. The signature benefit of mixing methods (not just using multiple methods) may be that the synergy from integration reveals a deeper or more useful window that neither would be able to offer independently.[300, 301]

David Morgan created a typology for conducting mixed methods research called the Priority Sequence Model.[302] Built on the foundational belief in the complementarity of qualitative and quantitative data, he suggested that researchers select which epistemology is most important for answering their primary question and which will maintain priority for their study (the priority, shown in capital letters). Then researchers must choose the sequence which determines whether the ancillary method will be used before (to inform) or after (to explain) the primary method (the sequence). Thus, four basic research designs are thus described: 1) a preliminary qualitative study

within a larger quantitative study (qual-QUANT), 2) a preliminary quantitative study within a larger qualitative study (quant-QUAL), 3) a qualitative study followed by a smaller quantitative study (QUAL-quant), or 4) a quantitative study followed by a smaller qualitative study (QUANT-qual).

This thesis' basic design is QUAL-quant which Morgan describes as the most appropriate for testing emerging theories,[\[302, 303\]](#) which this thesis seeks to do. However, rather than treating quantitative and qualitative as separate epistemologies these are instead regarded merely as separate methods within the integrated realist epistemology. Within this thesis, qualitative data are emphasised for two key reasons. Firstly, the qualitative data are rich and have not yet been analysed in as many ways as the quantitative data.[\[59, 278, 294, 304-306\]](#). Secondly, this thesis is a professional training programme and I wanted to focus my skill-development activities on qualitative methods in order to contribute in this domain.

The first study[\[8\]](#) (Chapter 4) is primarily qualitative, using thematic content analysis to describe how the AG coordinated the delivery of LT, how they were (or were not) supported by the external facilitator, whether or not student need data were used to inform decision making, how schools adapted the intervention, and whether or not components were acceptable and engaging. This study also explores how the delivery and acceptability of LT varied by school context and staff's capacity to implement a complex intervention in addition to their other duties. The study also employs basic quantitative description to establish fidelity and assess the extent to which AG members felt empowered. The second publication[\[9\]](#) (Chapter 5) is exclusively qualitative and used a variant of grounded theory called dimensional analysis to understand the rich context of case-study schools and to develop CMOCs, which are then tested quantitatively in the subsequent QCA[\[10\]](#) (Chapter 6). In this way, the publications build on each other to describe what happened, to explore the mechanisms that those activities may have activated and explore how they vary by context, and then test those quantitatively using QCA. Finally, this thesis seeks to integrate the findings and offer a final comment on whether or not realist trials are practically possible, philosophically cogent, and produce useful findings[\[11\]](#) (Chapter 7).

Description of the methods used in each chapter

In the following sections, the methods employed in each chapter are explained in more detail than was feasible to provide within the word limits of peer-reviewed publications.

Using realist concepts to assess the fidelity, feasibility, and acceptability of action groups as a participative strategy for whole-school health promotion interventions-- methods for Chapter 4

The focus of chapter 4 is on the AG component of LT. The primary outcome analysis found that AGs were central to building student commitment to school which was central to the study's theory of change.[278, 294] It was also the element of the intervention that varied most widely across the schools, and preliminary analysis indicated that AG were enabled and constrained by each school's context and the capabilities of involved staff. Finally, AGs are central to many school-based health promotion interventions but to my knowledge, they had not been examined in detail. [67, 84, 87] Chapter 4 is not described as being a realist study but is built on foundational realist assumptions, namely that the introduction of novel resources into a dynamic and complex context has the potential to enable changes in agent's reasoning and may change outcomes in some, but not all, contexts. Within the publication, how people interact with LT's resources is described as partially dependent on the pre-existing structures within the school and the agency of those around them. We describe schools as complex systems where other ongoing processes may facilitate or overpower mechanisms that had been activated by the introduction of LT resources. We acknowledge that the consequences of mechanisms' activation will vary by context and that feedback loops will likely occur in which changes in schools will influence future mechanisms, therefore changing the context in which LT continues to be implemented.

In order to meaningfully explore mechanisms in later work, we first had to describe the implementation of LT in each school, understand which components were feasible and acceptable, and how this varied depending on the school's culture, student body, and other contextual features. AGs were tasked with the co-ordination of other components of LT, including the implementation of the SEL curricula, the incorporation of RPs, the revision of rules and policies, and the enactment of locally relevant decisions. Therefore, by studying AGs, we were also able to assess how schools used LT resources more

broadly, and how that varied in different contexts. To assess fidelity, I analysed various data sources including: researcher observations of RP training and AG meetings; AG facilitator diaries; AG meeting minutes and attendance logs; AG surveys; SEL curriculum implementation surveys; interviews with PSHE curriculum co-ordinators; and staff surveys. Each are described in turn below.

RP training observations: Researchers aimed to conduct non-participant observations for at least one training in all schools. Observation proformas, informed by the pilot study, were used to document which topics were discussed and activities completed, and how many staff from each school participated.

AG observations: Ten AGs were randomly selected each year for researcher observation. Observation guides were similar to trainer diaries and allowed for richer emersion in the school and community environment.

AG facilitator diaries: Facilitators kept diaries for each meeting they attended. These forms were used to gather key information such as meeting date, duration, number of attendees and their role in the school (or year group for students), and gender. Diaries also asked facilitators to describe which data sources were being used to guide decisions and inform local actions, what priorities were established and what actions were being planned to address them, changes made to school rules and policies, implementation of the curriculum and how this was decided, and how engaged different AG members were. These were emailed or posted to the research team at regular intervals.

AG meeting minutes: Facilitators collected meeting minutes from the staff member leading the AG and were used to triangulate the facilitator diaries. These were emailed or posted to the research team.

SEL curriculum implementation survey: Surveys about the implementation of the SEL curriculum were sent to the teacher co-ordinating its delivery. Surveys were sent termly in the first two years and once annually (near the end of term three) in the final year. The survey asked which lessons were delivered, in which subjects they were taught, how many hours were delivered, and which resources (lesson plans, pre-made slides), if any, were used. Staff were asked to return the survey electronically to the PE lead researcher.

SEL curriculum implementation interview: Each year the research team aimed to interview the teacher with primary responsibility for implementing the SEL curriculum. If they were unavailable, interviews were conducted with a teacher who taught the curriculum. Interviews explored fidelity, acceptability, reach, perceived quality of the teaching resources, methods of delivery, student response, and facilitators and barriers to implementation.

AG member survey: In the first two years of the study, facilitators gave all AG members an anonymous survey to complete which explored issues around AG acceptability, how well they thought the AG functioned, and its composition, relating to, for example, diversity. The survey also contained an existing scale which measured if members felt empowered to make decisions.[307] In the final year, the AG lead was meant to distribute the survey to members. Surveys were posted to the study team.

Staff survey: Staff surveys were completed at the same time as student surveys but because some teachers needed to be supervising the student survey completion in classrooms, blank staff surveys were left to be completed and mailed to the study team. Staff surveys could be completed at home. The staff survey contained measures to assess staff stress and burnout as measured by the Maslach Burnout Inventory,[308] collating a description of cynicism, exhaustion, and inefficacy and staff QoL, using the Short-Form questionnaire which contains 12 items.[309]

Using the above data sources, I developed two separate fidelity scores; one measured fidelity in the first two years when facilitators were present, and a second score which was used to assess fidelity in the final, unfacilitated year. In the first score, I assessed fidelity based on whether:

- At least five staff members attend the in-depth RP training;
- Six AG meetings were held (indicated by meeting minutes or facilitator reports);
- Rules and policies were reviewed (as above);
- Locally relevant decisions were implemented (as indicated in meeting minutes, facilitator diaries or interviews);
- A 'good' or 'very good' range of members was represented on the AG (as indicated in the AG member survey);
- The AG was perceived to be well-led (as indicated in the AG member survey);

- A minimum of 5 hours or two units of the SEL curriculum was delivered (curriculum surveys and interviews); and
- At least 85% of staff reported that if there were trouble at school, staff respond by talking to those involved to get along better (as indicated in the staff survey).

In the final year of the trial when the research team had less direct contact with the schools, the fidelity measure was abridged and examined whether six meetings were held, whether locally relevant actions were implemented, whether at least 5 hours or 2 units of the curriculum were delivered, and whether or not more than 85% of the staff responded to conflict by helping the parties get along better. These indicators were selected because they reflect the key processes that LT resources were intended to initiate and because they showed sufficient differentiation amongst intervention schools.

Qualitative data use in Chapter 4 were collected using a variety of sources, including facilitator interviews, AG member interviews, and school staff interviews collected in all intervention schools as well as data from case-study only schools, including student and staff focus groups, described below. All interviews and focus groups were audio recorded with consent of participants and transcribed.

AG facilitator interviews: At the end of the first and second year of the intervention, telephone interviews, lasting 45-90 minutes were conducted with each facilitator (n=6). Speaking about each school one at a time, these interviews explored each school's culture, leadership, priorities, responsiveness to the intervention, barriers and facilitators to implementing LT, and any adaptations or deviations made.

AG member interviews: We aimed to interview two members of each school's AG, ideally with one student and one staff member during each year of the intervention. The AG lead was asked to identify appropriate interviewees. Staff were contacted by email or phone to schedule interviews which could be done over the phone or during the research team's next visit into the school, depending on their preference. Interviews with the students were arranged via staff and were always conducted in a private room (either a meeting room, spare office, or empty classroom) on school premises. Interviews were semi-structured, lasted between 30-60 minutes and explored what the school was like, what they felt the SLT's priorities were, the acceptability of the

facilitator and the intervention, AG meetings and how they could be improved, actions taken forward from the group, and any perceived outcomes or changes in the school environment since the introduction of LT.

Interviews with school staff: Semi-structured interviews were held in the first and final year of the study in all intervention and control schools. In the first year, the AG lead or another SLT member were interviewed as well as two teaching and/or pastoral care staff. All staff members were approached by phone calls and email. Control schools were interviewed in the autumn of the first year and intervention schools were interviewed in the summer term. Interviews began with a short survey exploring how authority is distributed in the school, levels of teacher-student collaboration, support available to students, use of achievement sets, disciplinary systems, and focus on academic or broader development. [305] After the survey, the interview explored the school's environment, approaches and practices used in managing student behaviour, social and emotional skills education, types of staff training, and students' ability to feed into decision-making in the school. In the final year of the study, only one SLT interview was sought in both intervention and control schools.

Data collected in case-study schools

Six schools were purposively selected as case studies to gather in-depth qualitative data on school contexts and processes associated with the implementation of LT. Schools were sampled in terms of diversity relating to the type of school, the facilitator, whether they were above or below the national average of children eligible for FSMs, and the extent to which their facilitator judged them to be highly, somewhat, or poorly responsive to intervention activities three months into the intervention. All of the following data were collected by one researcher on school premises.

Staff focus groups: During each year of the trial, we aimed to hold a focus group discussion with at least four staff members. The AG lead selected a diversity of staff members in relation to their level of seniority within the school and the level of knowledge of or involvement with LT. Discussions explored the school's culture, ethos, priorities, views about LT, and the use of RPs.

Student focus groups: Each year, two focus groups were held with students. One included students who participated in the AGs and a separate focus group was held with

students who were not directly involved with any LT activity. Students were selected by the AG leads, who were encouraged to select students who reflected the diversity of the school in terms of gender, ethnicity, school engagement, academic attainment, and social group. Discussions explored the school culture, ethos, priorities, recent school activities, student-staff relationships, how staff handle conflict, and descriptions of the school's neighbourhood. Discussions with students on the AG also explored how meetings run, their feelings about participation, and what if any changes they see in their school as a result of the group.

All qualitative data used in Chapter 4 were organised in NVIVO 11 (QSR)[310] and subject to multiple rounds of data analysis. The first round was completed prior to upgrading and was necessary for the NIHR report[294] and primary outcome paper[278] associated with the grant. This first round of data analysis served as a prolonged period of familiarisation with the key actors and schools where the intervention was being implemented.

In the first round of analysis, data sources, for example AG member interviews, were grouped together and read one at a time to become familiar with the data, and to encourage comparisons between similar groups of participants in each school and their reported experiences within the trial. After reading all the transcripts, initial codes based on the trial's research questions[5] and intervention theory of change, the theory of human functioning and school organisation,[7] and key findings from the pilot phase of the trial[174] were entered into a preliminary coding tree. Facilitator interviews were coded first as they had the broadest (and comparative) view as most worked across multiple schools. As transcripts were analysed, further codes were developed inductively to capture nuance within codes or relevant phenomena that did not connect to an a priori code.

After coding all of the qualitative data, the data classified under each code were re-read. To ensure consistency, some data were re-coded under newly developed or more clearly defined codes that emerged during the analysis process. After this process of refinement and clarification, the data contained in each relevant code were again re-read and summarised, using quotes as illustrations. These summaries were used for axial coding which helped us (EW and CB) understand the relationships between codes.

In the second round of analysis used to prepare the AG publication, data were re-analysed chronologically within each school to better understand what decisions were made and implemented (the process), how participants described the change, and to what, if anything, they attributed the change. This chronological coding process was important to create a narrative arch for each school and understand the details of implementation and its challenges. Although the analysis was primarily a thematic content analysis we employed some tactics from grounded theory, including constant comparative approaches and deviant case analysis. [260]

Further details are reported in Chapter 4.

Using dimensional analysis to understand participant accounts of contexts, mechanisms, and outcomes—- methods for Chapter 5

While coding the data from the 20 intervention schools for the aforementioned AG publication, it became clear that the six case-study schools could be typologised based on their local socio-economic context, staff capacity and student intake. Two schools were in high-poverty neighbourhoods and staff members were struggling to carry out their required duties, making additional responsibilities, such as the implementation of a complex intervention, impossible. Both of these schools dropped out of the PE in the final year of the study. Two other schools were diverse and had some staff able to take on additional responsibilities but were serving communities with unmet needs and difficult social circumstances, including complex immigration issues, community violence, and poverty. The final two case-study schools were in wealthy areas serving largely wealthy children and reported very little bullying and aggression at baseline. Rather than analyse data from all six schools, I decided to conduct a more intensive examination into the contexts, mechanisms, and reported changes in one school from each pair. Knowing the data from my previous two rounds of data analysis, I was confident that I would have enough data to reach theoretical saturation with fewer schools. Because of the volume of data collected in the study, I was also able to explore deviant cases or why outcomes may have changed in some but not all schools, which increases the value of the theory of change by engaging with contextually contingent activation of mechanisms. [202]

Chapter 5, published in *Trials*,[\[284\]](#) explores participant accounts of how the schools' contextual features, including staffs' priorities, motivations, and skills, informed how individuals used LT resources to activate (or override) mechanisms which then further interacted with aspects of school context to generate changes in bullying and aggression and/or improve wellbeing. It also explores how those mechanisms were sometimes unable to activate against the weight of a context that was unsupportive of the intervention. It is important to note that, within realist evaluations, a mechanism is commonly understood as changes in people's reasoning,[\[1\]](#) although Dalkin's influential work has led many realist evaluators to distinguish between intervention resources and changes in people's reasoning.[\[243\]](#)¹² A broader definition used in this thesis, is that a 'mechanism is causal potential.' Rather than emphasising an individuals' reasoned actions, 'mechanism as causal potential' recognises that the mechanisms may involve structure, agency, and/or the interactions between them. Within a realist view, mechanisms always exist (in the realm of the real) but will not always be activated. Moreover, mechanisms may activate and be "actualised" (in realist terms causing change in the actual realm) whether or not researchers are able to detect such change. It is also important to note here that the word "mechanism", like "context" and "outcome" are labels that we apply to a noumenon or phenomenon when we theorise. They are not categories of reality. The term 'mechanism', especially within something as pragmatic as a CMOC, is a heuristic devices which enables us to better understand reality via theorisation.[\[313\]](#) A phenomenon (such as having a peer-support network in a school) which in one CMOC is a contextual feature, may be a mechanism in another, and an outcome in a third place or time.

The analysis for chapter 5 was based on a variant of grounded theory called 'dimensional analysis'. However, the analysis was eventually presented in the published journal article as a thematic analysis. During the peer-review process, a reviewer suggested rejecting the manuscript, arguing that realist evaluation and dimensional analysis are fundamentally incompatible because the latter is based on grounded

¹² In response to growing frustration some realist evaluators were expressing in trying to distinguish context from mechanisms,[\[311, 312\]](#) Dalkin et al propose a slight refinement to Pawson and Tilley's CMOC. Arguing that "resources and reasoning are mutually constitutive of a mechanism"[\[243, pg. 4\]](#), they suggest that $M(\text{Resources}) + C \rightarrow M(\text{Reasoning}) = O$, or that resources introduced into a context activate changes in reasoning which lead to outcomes.

theory, which was developed by symbolic interactionists (and not realists). Rather than challenge the reviewer and risk being rejected by the journal, we decided to simplify the description of the methods and present it as a thematic analysis. While the analysis did identify recurring themes, it did more than this in that it developed theory about how mechanisms and contexts interacted to generate outcomes. Two important points should be noted here. Firstly, Glaser and Strauss started developing grounded theory in 1967 and Bhaskar did not begin writing about realism until 1975. Secondly, as will be described below, realism and grounded theory are widely regarded by researchers as compatible and have been used together successfully.[314, 315] Even though we used a method that a reviewer thought of as incompatible with our ontology, he did acknowledge that the findings were useful and were able to demonstrate the contextually contingent nature of mechanisms being activated (or not).

The purpose of grounded theory is to enable the systematic development of theories of causation.[315] The approach enables the inductive and deductive development of concepts through the confirmation or disconfirmation of hypotheses.[314, 316] When using grounded theory, researchers become familiar with, and begin to understand, the data in depth. This process of “open coding” facilitates that deep level of comprehension by asking researchers to identify common themes from different sources and agents.[260, 316] Data are fragmented and organised into open codes and refined and re-organised continually during what is known as “axial coding.” Axial coding shows how the codes relate to each other in higher orders, called concepts.

Our peer-reviewer was correct that grounded theory emerged from symbolic interactionism with its own distinct set of assumptions and ontological focus. However, it nonetheless provides researchers from other paradigms with useful tools for social investigation.[317-319] For realists, grounded theory offers an approach which enables the uncovering of latent concepts, (which in realist terms would lie in the realms of the actual or the real) and can develop them into a theory of causation grounded in data.[315, 320] In her publication on how critical realist grounded theory can improve social work, Oliver argues that “grounded theory can provide critical realism’s method”[315, pg. 373] and suggests three key areas of their compatibility. Firstly,

“critical realist grounded theory would address both the event itself *and* the meanings made of it, approach data with the preconceived analytic concepts of

emergence and generative mechanisms and pursue emancipatory, rather than merely descriptive, goals” [315, pg. 378] (emphasis in the original).

In other words, realists could still draw from the hermeneutical tradition of seeking to understand meaning, but would not be bound by relativism, which would limit one’s ability to make any claims about reality.[275] Secondly, realism requires a method that is open to fallibilism. As discussed earlier, all knowledge is partial and perspectival. In grounded theory, an emphasis is likewise placed on the belief that all theory is modifiable and based on a person’s experience and position.[321]¹³ Finally, Oliver argues that grounded theory and critical realism are compatible in relation to epistemic relativism. Realist evaluators endorse mixed and multi-method approaches and a central tenant of grounded theory is that ‘everything is data.’

The variant of grounded theory that I used is called dimensional analysis, developed by Leonard Schatzman[322, 323] as an alternative to the approach to grounded theory developed by Glaser and Strauss.[316] While Schatzman developed the method, he wrote only very sparsely about it. Therefore, the clearest articulation of his method was written later by his PhD students.[324, 325] The cornerstone and distinguishing feature of dimensional analysis is what users call the *explanatory matrix*. [325] The explanatory matrix provides a framework to break accounts into four key components: context (the boundaries of the inquiry), conditions (dimensions that facilitate, block or shape the phenomenon), process (mechanism), and consequence (outcome). Although the language is slightly different from that used in realist evaluation (and in the case of context, slightly confusing), this maps neatly onto the realms of a CMOC. However, what realists call context, dimensional analysts call conditions. Context in dimensional analysis can include broader social features, for example the nature of the English

¹³ It should also be noted that, while not discussing critical realist grounded theory, Sayer makes a similar argument about the need for a middle way between “foundationalism” (or the belief in an absolute truth) and the post-modernist assertions that ‘objects’ are socially and discursively constructed, and the subsequent denial of empirical testing. Sayer proposes a thought experiment in which he asks readers whether or not the following two beliefs can be simultaneously held: 1) There is no objective access to the world, knowledge is social and language, which mediates our access to the world, is opaque and not completely shared; and 2) we can develop knowledge of the world and make scientific progress.[275] Both foundationalists and post-modernists would agree and disagree with one of the two statements, while realists can simultaneously hold both as true.

education system or the impact of austerity measure in English schools, which affect the intervention but are broad enough to be shared across all schools.

According to Kools et al, the first analytic steps in dimensional analysis is called *designationalising*, where researchers code their data by both their *dimensions* and *properties*. Within the parlance of dimensional analysis, a 'dimension' simply means identifying and naming an abstract concept or component of a phenomenon, and *properties* describe the quantitative or qualitative parameters of a dimension.[325] For example, data from an interview might be coded under 'school ethos' (its dimension) and sub-coded under 'authoritarian' or 'restorative' (its property). Coding continues until sufficient designations are identified and explored, and a theory with potential explanatory power begins to emerge. As analysis continues, certain designations which seem as though they may explain the data, are elevated to being the *perspective*, or the lens through which researchers are best able to make sense of their data. If a particular perspective does not improve the intelligibility of their theory, it is removed from its hierarchical place, and another designation is elevated to being the perspective and tested for its explanatory power. It is through this process of exploration that dimensional analysis moves beyond description into explanation.

In their exemplar of using dimensional analysis, Kools et al explore how nurses distinguish acute confusion from dementia among older hospitalised patients. One of the first *designations* that seemed to explain why nurses misdiagnosed acute confusion related to whether or not nurses felt the behaviour was disruptive or a symptom of a health problem (two potential properties). When they elevated this dimension to a *perspective* which might explain the patterns they were finding in their data, this hypothesis did not hold up to further scrutiny. Later they elevated the dimension of staff training to the explanatory perspective, but that also failed when the *property* (being more or less well trained) did not explain misdiagnoses of acute confusion. Eventually, they examined the nurses' opinion on aging and whether or not older patients were assumed to be exhibiting confusion/dementia, if they were presumed healthy, or if nurses conceptualised their patients as vulnerable to ill health. When analysing data through this perspective, researchers were able to explain most (mis)diagnoses and explain deviant cases, therefore building a theory of how nurses' preconceptions affect clinical practice.[325]

Data used in this paper were from three of the six case study schools and included interviews with facilitators, SEL curriculum leads, AG members, and focus group discussions held with staff, students on the AG, and students not involved with the intervention (details of which are described above). Additionally, this analysis also included interviews with students who have participated in restorative conferences. The study team aimed to conduct semi-structured interviews with two students who had been involved in a restorative conference per year in each case-study school. During the first two years of the study additional interviews were also carried out in eight non-case-study schools. Where possible, AG leaders were asked to organise an interview with one boy and one girl who may have been either a perpetrator or victim. Interviews explored the instance that was addressed restoratively, the process of conferencing, its acceptability, resolutions, and outcomes. In some cases, interviews explored the use of RPs to address miscommunications, classroom misbehaviour, or friendship challenges.

For the analysis presented in Chapter 5, data were organised in Nvivo 11 QSR[310] and coding occurred chronologically, school by school, and in waves. The only a priori codes were “context”, “condition”, “process”, and “consequence.” As each interview or focus group was analysed, codes were created based on the phenomena participants were describing. For example, if, after being asked to describe their school at the beginning of an interview, a student reported that the school was “safe”, I coded that description under the parent node “condition” and under a child node called “feeling safe in school.” In another school or with another participant, safety might be described as an outcome and was therefore coded under “consequence.” This is important to note because there is some confusion and debate within realist evaluations about how to account for whether a phenomenon is a context, mechanism or outcome.[245, 258, 311, 326-329] For this study, it was important to code phenomena as participants described them. Therefore, what someone describes as an outcome in one place may be a contextual feature in another, and a mechanism in a third place or time.

Subsequent coding aimed to identify inter-relationships between these explanatory matrices in order to develop an overall analysis with a unifying ‘perspective’: i.e. an overarching logic of what is occurring with the social phenomena being examined. Comparison between the explanatory matrices developed for each school focused on

how processes were reported to occur in each school, how these processes were affected by the reported conditions in schools and with what reported consequences. These matrices, representing sub-processes, explain how changes occurred in schools.

Throughout the analysis process, I drew and continually updated diagrams of the context, community and school conditions, intervention processes, and outcomes as described by participants. Analytic memos were also kept throughout the duration of the analysis and writing up.[320] In the form of a research diary, the memos contain research plans, reflections on my analysis and its progress, thoughts on why certain quotes seemed important and what role they might play in generating a larger theory, or how some students' responses contradicted what I expected. Notes were also kept on emerging themes and ideas that I thought would aid in the writing process.

The 'perspective' identified in this analysis was found to be building an inclusive and cohesive school environment, which was comprised of three mechanisms: 1) improving student commitment to the school community; 2) building healthy relationships by modelling and teaching pro-social skills; and 3) de-escalating bullying and aggression, and enabling re-integration within the community. These processes, when joined together, help explain how mechanisms might generate changes in individuals and the broader school.

Testing CMOCs using qualitative comparative analysis-methods for chapter 6

Based on the original proposal for realist trials,[2] I had planned to study the impact of context and the activation of mechanisms using moderator and mediator analyses, respectively.[5] Simply put, "moderators identify on whom and under what circumstances treatments have different effects. Statistically, they are understood as a third variables modifying the relationship between two other variables. In trials, moderators modify the strength along all or part of the causal pathway between allocation and outcome.[330] Mediators identify why and how treatments have effects." [331, pg. 887] They examine variables that occur along the causal pathway [332] that might explain the causal impact of the intervention. Explained another way, mediators are third variables that account for the association between two others.

Mediator analysis may also demonstrate the presence of non-linear or synergistic effects of the intervention components and may help elucidate which inputs had effects in which contexts. In INCLUSIVE, relevant moderators might include Ofsted rating, deprivation, or baseline experience of bullying or aggression while mediators may include intervention fidelity or rates of staff burnout.

As explained above, I argue that both moderator and mediator analyses are compatible with a realist ontology,[\[294, 304\]](#) although I acknowledge that this view is not widely accepted amongst realists.[\[3, 4\]](#) While I led on the publications contained in this thesis, I also contributed to studies led by other INCLUSIVE team members using moderation,[\[278, 294\]](#) mediation,[\[304\]](#) and moderated-mediation analyses[\[306\]](#) (see Appendices 7-10). With these methods already applied to INCLUSIVE data and acknowledging that I did not actually want to learn how to conduct statistical analyses, I decided to explore whether QCA could be an additional and appropriate method to use as part of a realist inquiry.[\[333, 334\]](#) The use of QCA in trials is uncommon and therefore its application to INCLUSIVE had greater potential for innovation.

Brief overview of QCA

In brief, QCA is both a method and approach for identifying complex combinations of conditions which do or do not lead to a pre-determined outcome. Increasingly used in evaluation studies, QCA is “conjectural in its logic, examining the various ways in which specified factors interact and combine with one another to yield particular outcomes.”[\[335, pg. 1079\]](#) Befani et al argue that realist evaluation and QCA “share a complex view of causality, a generative perspective, a theory-driven approach to empirical observation, and a limited claim to generalization.”[\[333, pg. 171\]](#)

It is important to note, that despite being quantitative, QCA is not a modification of statistics to better suit small sample sizes.¹⁴ It is not variable focused and it does not examine correlations or net-effects. QCA is used to study configurational causality. QCA focuses on the “cases” which may be a country, a school, or studies within a systematic review, and examines the *configuration* of conditions which may be contextual, personal, or interventional features that, when grouped in certain ways, give rise to the

¹⁴ While INCLUSIVE collected data from nearly 6,000 students, it is considered a medium N study because the cases are schools (not students) meaning that there are only 20 cases when examining intervention schools and 40 cases when analysing both trial arms.

outcome of interest.[336] These configurations are commonly described as “pathways” or “recipes” and are grouped into these pathways using Boolean operators ‘and’, ‘or’, and ‘not’.

There are multiple forms of QCA, the two most common being crisp-set QCA (csQCA) and fuzzy-set QCA (fsQCA). Within csQCA, all conditions are binarized as 0 or 1, indicating that they either fully do or fully do not represent the condition of interest. In fsQCA, data is calibrated to fit anywhere between 0-1 (described in detail in Chapter 6) so that the degree of caseness can be explored with greater nuance.[337]

Without imputations or other statistical tools, QCA cannot cope with missing data; if there is a missing datum for a case, the whole case must be excluded. To avoid reducing sample size, researchers can quantify qualitative data or use their expert knowledge of a topic to assign approximate values between 0 and 1 for the condition of interest. When a case (e.g. a school) is given a 0 for a condition (e.g. having an authoritarian ethos), the 0 indicates that the school is fully not within the set of schools that meets that condition (e.g. they are fully not an authoritarian school), while a 1 represents that they fully represent the condition within that set (they are authoritarian).[199] Even without missing data, researchers can assign numeric values to describe the case. Before the analysis for any type of QCA can begin, all data, whether originally qualitative or quantitative needs to be calibrated so that they fit between 0 and 1, showing the extent to which they represent each condition.

QCA is not without disadvantages. QCA may be susceptible to random error in which coincidental patterns are confused for causal patterns. Small samples, which are common in QCA, make such coincidences more likely. QCA is unable to take account of confounding in which the relationship between concepts is actually a reflection of unmeasured, underlying concepts, or random error, in which the pattern of causation that emerges would be different if large samples had been included. Therefore, a degree of caution is needed when reflecting on the appearance of casual patterns to query whether the results are spurious.

Analytic methods are described in Chapter 6 and expanded upon here.

Building data tables

To begin the analysis, we extracted the CMOCs developed through qualitative research from the end of Chapter 5 and placed each in their own dummy table, marking each component part of the hypotheses as context, mechanism, or outcome. A dummy table is a table with the appropriate columns and rows labelled but has not yet been completed with any data. QCA analyses generally employ cross-sectional data but we decided to maximise the use of longitudinal data to understand the process of how change occurred over time. Therefore, to make the models resemble our hypotheses as closely as possible, we decided that contextual features should be pulled from baseline surveys completed by all students, outcomes should be taken from the final survey or represent change over time, and mechanisms should be taken from mid-line data collected at 24-months. I then reviewed all relevant data collection tools and selected the questions, scales or data points that were the most appropriate proxies for the phenomena that we wanted to explore. These were discussed, modified as needed and agreed by the study team (EW, GJMT, CB).

When the proxy measure was from the PE, I had access to the data and extracted the relevant values for each school. When the proxy was from the outcome evaluations or larger surveys (such as the AG survey), GJMT extracted the data on my behalf. Once the dummy tables were filled in with the raw data, it needed to be calibrated to explore the distribution of each variable between 0 and 1. 'Caseness' or the degree to which a case demonstrates the phenomena is assessed by establishing three key cut-off points (called anchor points) The first cut-off point corresponds to the point above which all schools clearly represent the condition of interest. The second cut-off point represents the point below which the schools clearly do not represent the condition of interest. The final cut-off point is placed at the spot of greatest ambiguity where researchers are unclear whether or not the condition is represented. With these three cut-off points, the data are divided into four groups: clear examples of a case; those that are more a case than not a case; those that are more not a case than a case; and those who are clear examples of not being a case. Cut-off points for each proxy were established based on two primary criteria. Firstly, the cut-off points needed to reflect the meaningful presence or absence of our condition of interest (i.e. the cut-off value for decreasing conduct problems needed to represent a change which would be of public health significance) and secondly, cut-offs needed to provide us with a reasonable degree of

distribution (i.e. it is not a helpful indicator if every school either meets or does not meet the conditions of interest). Once each conditions' cut-off points had been set, GJMT calibrated the data in STATA and returned them to me. Within the parlance of QCA, once the data has been calibrated, (i.e. made all the data fit between 0 and 1), those scores are referred to as "truth values" and represent the degree to which a case is or is not an example of a condition.

To make sense of and "see" the patterns of causation, I colour coded the data to represent the degree to which each cell represented a case of each condition. Cells that represented cases (truth values greater than 0.9) were dark green, cells that represented more caseness than not caseness (truth values 0.5-0.89) were light green, cells that were represented less caseness than caseness (truth values 0.05-0.5) were orange, and definite not-cases (truth values less than 0.05) were red. As someone who has never done quantitative analysis, this colour coding system, while unconventional, was vital to making sense of the data. I then explored the data by copying and pasting schools with the same recipes next to each other in Excel and drew on my in-depth knowledge of cases to understand patterns of emergence or non-emergence and explain visible contradictions.

Constructing truth tables

After I had thought through the patterns I was seeing in the data tables, I then used the Tosmana add-in[338] for Microsoft's Excel to produce the truth tables. Truth tables show the recipes or pathways that either led to change or kept outcomes stable over time and the cases (in this case schools) that followed each pathway. When truth tables are created, QCA software also calculates *coverage* and *consistency* scores. Consistency relates to the proportion of cases within each recipe that have the same outcome.[339] Consistency will be low if there is weak or contradictory evidence that this pathway consistently leads to the outcome of interest. Coverage relates to how much of the outcome is explained by the model. Low coverage indicated that the model does not contain all key explanations. When the consistency and coverage were low, additional conditions theorised from the qualitative data were added to the data table, and the same process described above was repeated.

One of the mechanisms I was assessing relating to how bullying can be de-escalated amongst a core group of offenders. When the trial was designed we did not anticipate

that feeling empathy or contrition and accepting that punishment was fair would be key mechanisms, and therefore no measurements for them were incorporated into the data collection tools. Given their prominence in the qualitative data, I did not want to disregard them from the analyses simply because I did not have suitable measures, especially when QCA allows for the quantification of qualitative data. Therefore, I re-read all of the interviews with students who had been involved in restorative conferences and extracted data on what their conflict was about, how serious it was, whether the aggressor expressed feeling empathy either during or as a result of the meeting, and whether they reported feeling that their punishment was fair and they felt contrition. The data were grouped by school and ordered by the amount of empathy and contrition expressed. 'Caseness' was then assessed and cut off points are established at the following places: 1) the line above which the person clearly felt empathy; 2) the line between feeling some and feeling no empathy; and 3) the line below which the person displayed minimal feelings of empathy. With these three lines, the data are divided into four groups: those who clearly showed empathy, those who showed some empathy, those who showed minimal empathy and those who showed no empathy. These four groups are then assigned numeric values of 1, .67, .33 and 0 and analysis carried on as before. Quantification of qualitative data is a useful way to keep the QCA sample as large as possible but it should be clearly acknowledged that this process may have led to meaningful misclassification and that social desirability bias may have encouraged students to report feeling more contrition than they actually felt.

Boolean minimisation

After each of the truth tables had satisfactory consistency and coverage or when there were no further conditions we could add to the model which emerged directly from our earlier qualitative analysis, we sought to minimise solutions to produce the clearest and most parsimonious models. As our models were not large and we had few cases, this was done by hand by EW and GJMT.

This publication has been submitted to the Journal of School Violence. Peer reviewers have asked for revisions, and those have been re-submitted for their further consideration.

Reflection on whether realist RCTs are possible—methods for chapter 7

As described so far in this chapter, each publication in this thesis builds on the one that precedes it. I began by assessing whether or not schools were able and willing to implement the intervention with fidelity and explored how this was affected by different features of the schools' context and agents' choices. I then selected three case study schools partially based on their ability or inability to deliver the intervention. These participant accounts were analysed to show how LT resources were used, how they were described as affecting subsequent changes in reasoning which may have affected outcomes, and how those changes were limited or enabled by the various contextual features in each school. The CMOCs derived from that analysis were then tested using QCA to understand the causal recipes of contextual features and mechanisms that appeared to be linked to decreases in bullying (or not). This approach also enabled us to assess whether similar mechanisms appeared to create reductions in bullying in control schools, shedding some light on the theoretical transferability of the identified mechanisms. In this final commentary, I synthesised these analyses along with other publications from the INCLUSIVE trial to answer the question about whether or not realist trials are philosophically coherent, practically feasible and generate useful findings.

The final publication reflects on the process of conducting these realist analyses within the first explicitly realist RCT. The idea of realist RCTs has been hotly contested and debated but until these empirical papers were published, many of the arguments were theoretical. In this paper, I reflect on what has been learned, what methodological challenges we faced, whether or not we were able to adequately address them, and whether the failings or short-comings which inevitably occurred were the result of our particular study design, an inherent flaw in the idea of realist trials or other factors.

Ethics

I received ethical approval for secondary data analysis for this thesis through the London School of Hygiene & Tropical Medicine's (LSHTM) Research Ethics Committee, in line with the Research Degree Handbook (ref: 16091).

Reflexivity and Positionality

When I began this PhD, I had four years' experience as a research assistant and research fellow at LSHTM, and I considered myself a novice at all of the methods that this study required. I had published systematic reviews and papers using thematic analysis, but

was completely new to grounded theory and PEs, and was unaware of realism, dimensional analysis and QCA before joining the INCLUSIVE team. While some of my peers predicted that they would be experts when they completed their PhDs, either in relation to the topic or the methods they employed, my goals were more modest. [340] By the end of my PhD I wanted a strong foundation in qualitative methods that would enable me to be a more independent researcher and I wanted to learn enough about quantitative methods to be a valuable team member who understood why certain decisions were being made while working within mixed-methods study teams. Having been on short-term contracts where there is always pressure to meet tight research deadlines and publish papers, I was also aware that my PhD was going to be one of my only opportunities to have protected time to study and explore methods in depth and think about the broader issues, such as ontology. Therefore, I planned a thesis with a strong methodological focus and one that contained methods I would not have the time to learn on a short-term contract. I also wanted to complete a thesis with strong methods to provide me with the tools to work on studies across a range of public health topics.

In relation to the daily work of being a research fellow and PhD student, my position on the team affected how research participants interacted with me. I joined the study team after the intervention was developed and the evaluations designed. This enabled me to tell staff and students honestly that I was hired to evaluate LT and that I would not be offended if they gave me their honest feedback about what they did not like and what did not work for them. This appeared to be liberating, particularly for some staff who disliked LT, knowing that they could vent their frustrations with me without risk of causing offense.

While at university, I worked for three years as a sexual assault, domestic violence and stalking crisis counsellor. I also trained five cohorts of approximately 25 volunteers to work on a hotline and accompany sexual assault survivors during forensic examinations. My instinct when hearing about experiences of violence, including bullying, is still unshakeably informed by this experience. During one day of data collection, I was meant to be interviewing a boy who had held a sawn-off BB gun to a girl's head, made her kneel and pulled the trigger making the gun click. Thankfully, it was unloaded and she was physically unharmed. Weeks later, when he came into the

room to be interviewed, the boy was aggressive in his manner with me and before I could explain why I was there or what I wanted to speak to him about, he had said that he “didn’t fuckin’ want to talk” to me and called me a “cunt.” When I excused him from the interview and told him to return to class, the AG lead brought in the girl he had victimised. Knowing in advance what had happened to her, and seeing how nervous she was and how physically small she tried to make herself when discussing what happened, I was aware that the format of my conversation with her was coming directly from my work as a crisis counsellor and not my training as a researcher. When analysing the data later, it was clear from the transcript that parts of the interview followed the crisis counselling structure of “stabilise, normalise, validate, empower.” It should be noted that throughout the interview, I reminded her that she didn’t need to speak with me about this and she could stop at any time. She wanted to tell her story since she felt RP had been helpful.

When I began to analyse interviews, I was aware that I felt more inherent trust in the interviews with victims than perpetrators. I think my experience as a crisis counsellor, where I was trained to believe and respond to the needs of survivors created an unintentional drive in me to therefore focus on their narratives. However, when trying to understand why violence and aggression happened, I needed to actively push myself to give more meaningful consideration to perpetrators because they were the ones enacting the behaviours LT sought to prevent. I had to be cognisant of how this bias might affect my analysis and I knew that, at times, despite feeling angry about the cruelty of their behaviour, the interviews were conducted with children who were still learning and who had the potential to make better choices going forward. To correct for this potential bias or under-coding of perpetrator narratives, I could engage in a simple exercise of mindfulness to clear my preconceptions and code the interviews from a place of curiosity rather than judgement.

Being American was very helpful and changed the way that participants, particularly students, interacted with me. Often, as soon as I introduced myself, students would pick up on and ask about what many perceived to be a heavy accent. Introducing myself, explaining how I came to work in England and answering the questions about the US generally or California specifically was often a time in which I could build rapport with the students. Being an outsider, not only to their school but also to the English

education system, meant that during interviews, students would take extra time and effort to explain nuances that they were unsure I would understand. This sense that they could explain what school was “really” like to someone with no inside knowledge and in a setting where I told them they had privacy to tell me what they really felt about school (within the limits of child protection) meant that we had rich conversations in interviews.

In this thesis, I explore whether realist trials are possible. I begin this by describing possible methods for assessing fidelity and how intervention resources interact with different school environments and lead to the emergence of changed (or sometimes unchanged) outcomes.

FINDINGS

Introduction to the first study:

An important vulnerability of evaluation studies is known as type III error, or the rejection of an intervention's general approach or theory of change as ineffective when the intervention was not properly implemented.[341, 342] Therefore, before knowing whether or not the intervention "worked", researchers need to assess fidelity, or whether and to what extent the intervention was actually implemented. In some interventions, particularly those that are clinical, pharmacological or measuring efficacy (not effectiveness), it may be of paramount importance that inputs are standardised and procedures are clearly delineated and adhered to.[343] While the MRC recommends that interventions being evaluated through pragmatic effectiveness trials are relatively standardised, they reject the notion that RCTs are incompatible with adaptation,[208] especially for interventions that are social in nature. Planned adaptations generally occur between evaluations in an effort to ensure the intervention's relevance or acceptability in a new setting. Conversely, unplanned adaptations generally occur within or during trials when, for example, people use intervention resources in unintended ways. Trialists of social interventions generally accept unplanned adaptations because the goal of the evaluation is to understand how people will use the resources in real-world settings. [ENREF 160](#) Howe et al argue that interventions should be standardised according to their *function*, which may take different *forms*, depending on contexts.[343] For example, for an educational intervention to remain faithful to form, all schools would deliver the same lessons using the same learning materials. Using fidelity of function, schools would be able to tailor the lessons based on local needs, including literacy levels or culture.

Planned and unplanned adaptations and fidelity of form and function were central to the INCLUSIVE PE, especially in relation to the AGs. AGs were meant to involve convening a group of diverse students and staff to review the NAR, rules and policies, and ensure implementation of LT but they were also the fora through which school-specific actions were decided. In this way, the process of holding AG meetings was intended to embody fidelity of form while the outputs of the groups were intended to embody fidelity of function.

Realist evaluations focus on how mechanisms, activated by people with access to novel resources provided through an intervention, interact with a context to generate different outcomes. Therefore, the first way that context affects mechanisms is by influencing the way that intervention resources are used. To begin assessing if realist trials are possible, we need to understand what was done in each setting, how implementation varied by context, what unplanned adaptations occurred, and whether or not those appear related to contextual or structural features or agents' choices. Although the following publication is not explicitly realist, its realist orientation is implicit in its assumptions that context will affect implementation and subsequent mechanisms, that novel resources will affect the reasoning and behaviours of agents, that participants are active agents in how they make use of resources, and that various mechanisms will activate or remain latent as a result of the actions that ensue.

Given the volume of data collected for the PE, I chose to focus the following analysis on the AGs because it is important to understand their role coordinating the implementation of other intervention components (including the SEL curriculum), and intended processes (including the implementation of restorative practices and the revision of rules and policies). AGs were also the element of LT that varied most widely across schools, not necessarily in relation to how many meetings held (although there was some variation) but in relation to the tasks that they chose implement and the way that students were able to participate and affect change. Moreover, our earlier analysis indicated that they are a crucial forum through which commitment to school appeared to have developed, a mechanism which was central to our theory of change. Finally, AGs are central in other whole-school, health promotion interventions but have not yet been evaluated in detail. Within a realist perspective, these variations in implementation would help explain how and under what conditions LT was able to affect outcomes.

Chapter 4: Action Groups as a participative strategy for leading whole-school health promotion: results on implementation from the INCLUSIVE trial in English secondary schools¹⁵

¹⁵ Please note that permission from the publisher to include the following paper has been obtained and is documented in Appendix 13.a

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	233090	Title	Ms
First Name(s)	Emily Ashbrook		
Surname/Family Name	Warren		
Thesis Title	Are realist randomised controlled trials possible? INCLUSIVE as a case study of an emerging methodology		
Primary Supervisor	Professor Chris Bonell		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	British Educational Research Journal		
When was the work published?	October 2019		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	NA		
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	For this paper, I collected some of the data, led on the analysis and writing. Charles Opondo calculated the Learner Empowerment Scale. Chris Bonell supervised the research and led on the submission process.
--	---

SECTION E

Student Signature	Emily Warren
Date	28/01/2022

Supervisor Signature	Chris Bonell
Date	28/01/2022



Action groups as a participative strategy for leading whole-school health promotion: Results on implementation from the INCLUSIVE trial in English secondary schools

Emily Warren^a, Leonardo Bevilacqua^b, Charles Opondo^a, Elizabeth Allen^a, Anne Mathiot^b, Grace West^b , Farah Jamal^c, Russell Viner^b and Chris Bonell^{a,*}

^aLondon School of Hygiene & Tropical Medicine, UK; ^bUCL Great Ormond St. Institute of Child Health, London, UK; ^cUCL Institute of Education, London, UK

Education policy increasingly promotes action groups as a key strategy for student and/or staff participation in school improvement and whole-school health promotion. Such groups can coordinate multi-component interventions, increase participation and engagement, and enable local adaptations, but few process evaluations have assessed this. We evaluated fidelity, feasibility and acceptability of action groups as part of a trial of a whole-school intervention to reduce bullying and aggression and promote health in English secondary schools, which reported multiple health and educational impacts. Action groups involved students and staff, supported by external facilitators, and drew on data on student needs. They aimed to: coordinate implementation of restorative practices and a social and emotional competencies curriculum; review policies and rules; and enact local decisions to modify school environments. Our process evaluation used interviews, focus groups, observations and questionnaires to assess action groups' fidelity, role in coordination, role in local adaptation, support from external facilitators and data on student needs, and acceptability in engaging members. Fidelity was high in the first two years but lower in the third year when external facilitators withdrew. Student needs data were perceived as useful, but views on external facilitators were mixed. Groups successfully reviewed policies and rules, planned activities and coordinated restorative practices, but were less successful in implementing the curriculum. Success was facilitated by the involvement of school leaders. Members reported high satisfaction and empowerment. Action groups are a promising strategy for leading whole-school health promotion. Implementation is supported by external facilitation, local data and involvement of senior managers.

Keywords: action groups; whole-school interventions; fidelity; implementation

Introduction

Action groups (AGs) are decision-making bodies that involve school staff and/or students. They are one approach to promoting 'student voices' in which students have a

*Corresponding author. London School of Hygiene & Tropical Medicine, 15–17 Tavistock Place, London WC1H 9SH, UK. E-mail: chris.bonell@lshtm.ac.uk

say in the running of schools, and are increasingly used to implement whole-school interventions to aid school improvement and promote student health (Department for Education, 2014). AGs provide a practical means of promoting a systems-based, ecological approach to render school environments more health-promoting. They have been theorised, for example, to enable positive ‘disruptions’ to school systems which can lead to improved staff–student relationships and school systems being more attentive to student needs and concerns, which in turn may increase student engagement with school and decrease involvement in risk behaviours (Bronfenbrenner, 1979; Markham & Aveyard, 2003; Bonell *et al.*, 2013b).

There is evidence from a number of trials that AGs are an important element in school-based health interventions. Our own ‘INCLUSIVE’ randomised controlled trial (RCT) reported that ‘Learning Together’, a whole-school intervention comprising AGs as well as restorative practice and a student social and emotional competencies curriculum (Figure 1: logic model), was effective in preventing bullying victimisation, as well as preventing smoking, use of alcohol and drugs and contact with police, and promoting mental wellbeing, psychological functioning and health-related quality of life (Bonell *et al.*, 2018). The present article reports from the process evaluation of this trial.

Previous trials have also reported positive results on whole-school interventions involving AGs. The Gatehouse intervention included organisation-change and classroom-curriculum components to promote the emotional and behavioural wellbeing of young people in Australian secondary schools. In each school, an ‘adolescent health team’ comprising staff, community representatives and an external facilitator drew on information from student surveys to identify risk and protective factors for each school and locally appropriate strategies to address these. An RCT reported

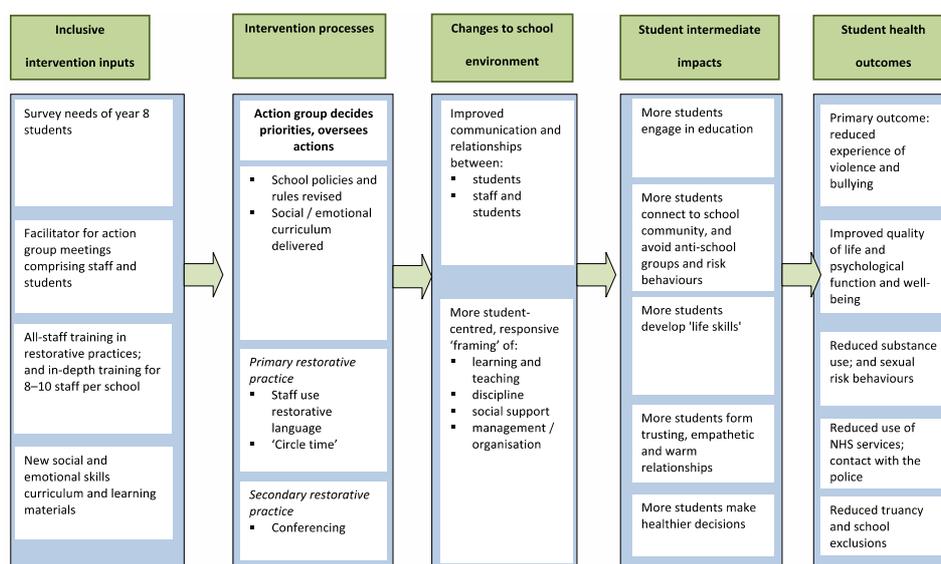


Figure 1. Logic model [Colour figure can be viewed at wileyonlinelibrary.com]

mixed effects, including no effects on depressive symptoms but significant benefits for smoking, alcohol use and a composite measure of risk behaviour (Bond *et al.*, 2004; Patton *et al.*, 2006). A similar intervention, implemented in middle schools in Chicago, involved a local school task force, including staff, students, parents and community members, which oversaw implementation, modified school policies and developed school–community collaborations. A central aim of the task force was to transform the school community to promote a sense of belonging among all members of the school community. An RCT reported that for boys, the intervention was associated with reductions in violent behaviour, drug use and recent sexual intercourse (Flay *et al.*, 2004).

The Safer Choices intervention aimed to prevent sexually transmitted infections and pregnancies among US adolescents via a multi-component intervention comprising a curriculum, student-led advocacy and parent information. Intervention activities were coordinated by a school health promotion council comprising teachers and administrators, students, parents and community representatives. An RCT reported a range of impacts reducing sexual risk behaviours (Coyle *et al.*, 1999, 2001; Basen-Engquist *et al.*, 2009). The PREPARE intervention was a multi-component, school-based prevention intervention to promote sexual health and decrease intimate partner violence (IPV) among young adolescents in South Africa. A team was convened in each school comprising the principal, teachers, the school safety officer, parents and local police. The teams undertook a safety audit to inform a local safety plan aiming to reduce the acceptability and prevalence of IPV and sexual violence. An RCT reported reduced rates of IPV in intervention schools (Mathews *et al.*, 2016).

The questions examined in this article concern the role that AGs play and how they are implemented. AGs are potentially important for three reasons. First, they enable coordination of whole-school interventions that are frequently complex and multi-component. Such interventions are feasible to implement in schools and are effective in promoting health across a range of outcomes (Langford *et al.*, 2014), with greater impacts than for single-component interventions such as health-education lessons (Greaves *et al.*, 2011). Coordination through an AG is likely to be important to ensure components interact synergistically (Craig *et al.*, 2008).

Second, AGs can enable local adaptation. There is a debate in the literature about whether implementation of health promotion should prioritise fidelity to maintain ‘active ingredients’ (Mihalic, 2004) or prioritise adaptation to operationalise an intervention’s theory of change in a locally appropriate manner (Dane & Schneider, 1998; Hawe *et al.*, 2004). The appropriate balance will likely vary between intervention types; whole-school interventions aiming to modify school environments may require more adaptation because they function as ‘interruptions to complex systems’, building on existing capacity (Hawe *et al.*, 2004; Quinn & Kim, 2017). However, adaptation should be planned and ensure fidelity to the theory of change (Hawe *et al.*, 2004), suggesting a role for AGs using standardised processes for adaptation.

Third, AGs can promote staff and student participation in decisions, ensuring interventions attend to stakeholder preferences and potentially generating health benefits because such participation can promote student commitment to school (Hulpia *et al.*, 2009), which is associated with reductions in health risk behaviours (Bonell *et al.*, 2013a).

While it is clear that AGs have been central to a number of effective whole-school health interventions, only the Gatehouse project included a process evaluation. This reported that although teams could take time to assemble and organise themselves, they were crucial in implementing whole-school change particularly in coordinating the work of disparate decision-makers across the school to build consistent policies and systems on existing work (Bond *et al.*, 2001). The team-based approach empowered members to voice their opinions, initiate and implement change, and increase the profile of health and wellbeing across the school. The use of information from student surveys in the Gatehouse project was an important element of the action teams' work, providing an impetus for action, securing the interest of senior staff and informing priorities. The external facilitators deployed to help implement action teams were viewed as important in maintaining momentum and linking schools with external sources of support (Glover & Butler, 2004). Facilitators reported that they sometimes needed to reframe public health issues, such as mental health, into more educational concepts such as school engagement, and align them with schools' institutional priorities. Facilitators reported that the intervention worked best where teams included diverse staff who shared ownership. However, the Gatehouse process evaluation did not report on the fidelity of delivery of action teams or the extent to which the teams adapted the intervention.

We will build on these findings to examine the following research questions: With what fidelity were AGs implemented? What role did AGs play in coordinating the intervention? To what extent were AGs supported by the provision of external facilitators and data on student needs? What approach did AGs take to local adaptation? Were AGs acceptable to, and engaging and empowering for their members?

Methods

Trial methods

Here we provide a summary of the methods, which are described fully in the published protocol and main trial report (Bonell *et al.*, 2014, 2018). We undertook a two-arm repeat cross-sectional RCT ('INCLUSIVE') of the Learning Together intervention in 40 secondary schools across south-east England. Our study population consisted of all students in the school at the end of Year 7 (aged 11–12 years) at baseline in March–July 2014, and at 24-month (end of Year 9; aged 13–14 years) and 36-month (end of Year 10; aged 14–15 years) follow-up. Schools were randomly allocated after baseline surveys with a 1:1 ratio to the intervention or control arm, stratified by: single-sex vs mixed-sex school; school-level student eligibility for free school meals as an indicator of family benefits entitlement (0–23%; >23%); and school-level performance in GCSE examinations accounting for student attainment on entry to the school as a measure of academic progress (above and below the median score for England of 1,000). Schools randomised to the control group continued with their normal practice and received no additional input.

Schools allocated to the intervention group were provided with resources to facilitate implementation of AGs, restorative practice and a social and emotional competencies curriculum. AGs comprised at least six staff and six students (not parents),

led by a member of the school's senior leadership team. Restorative practice is an approach to discipline aiming to improve relationships to prevent and/or resolve conflicts (Morrison, 2005) using 'circle-time', which brings staff and students together to maintain good relationships, and 'conferencing', which brings together the relevant parties to a conflict to find ways to improve relationships to avoid further harm. All school staff received basic training to implement restorative practices and 5–10 staff per school received in-depth training to deliver restorative conferences. Schools were provided with lesson plans and slides to guide delivery of 5–10 hours per year of a classroom social and emotional competencies curriculum for students in the trial cohort moving through Years 8–10 (aged 12–15 years). To support AGs, schools were provided with a manual, an external facilitator in the first two but not the third year of intervention (when schools were expected to be self-facilitating) and a report summarising the findings from student surveys on attitudes to and experiences of school. AGs aimed to coordinate intervention delivery across the school, and to revise school rules/policies and enact local decisions informed by survey data to support restorative approaches to school discipline and make the school environment more engaging and healthy. The intervention was universal primary prevention, and AGs were focused on intervention coordination and broader changes to the overall school environment. AGs were not focused on outreach, targeted support to students with particular needs, coordination with other family or health services, or wider school inspections.

The intervention's theory of change was informed by the theory of human functioning and school organisation (Markham & Aveyard, 2003) and postulated that the intervention would enable schools to reduce bullying and aggression, improve mental health and reduce substance use by increasing student commitment to school via improving relationships between staff and students and between academic education and students' broader development.

Primary outcomes for the trial were bullying victimisation measured by the Gatehouse Bullying Scale (Bond *et al.*, 2004) and perpetration of aggression measured by the Edinburgh Study of Youth Transitions and Crime Scale (Smith, 2006). Secondary outcomes included smoking and use of alcohol and drugs, mental wellbeing, psychological functioning and quality of life, and contacts with the police and NHS, which are described in our published protocol and main trial paper (Bonell *et al.*, 2014, 2018). The trial was prospectively registered as ISRCTN10751359 with the ISRCTN Registry on 30/01/2014. All amendments were approved by the independent study steering committee prior to collection of the 36-month surveys and any trial analyses.

Process evaluation of AGs

The process evaluation examined intervention implementation and receipt, guided by the trial protocol. The following aspects of the process evaluation occurred in every intervention school. External facilitators collected minutes and completed a diary for each meeting attended, which assessed duration, attendees and their roles, gender and participation, use of survey data, priorities and actions set, revisions to rules and policies, and group dynamics. Researchers aimed to conduct structured observations

of group meetings in 10 randomly selected schools each year of intervention. Observation guides focused on the same areas as diaries. An anonymous survey was distributed to all members of AGs by facilitators at the end of each year of intervention. This explored the diversity of staff and students on the AG, how well led the group was and how empowered members felt, using an adapted version of an existing scale (Frymier *et al.*, 1996). Participants placed the questionnaires in an envelope, which was sent to the researchers. Semi-structured telephone interviews with external facilitators ($n = 6$) were conducted in years 1 and 2, lasting 45–90 min and exploring school culture and priorities, interventions, adaptations and deviations, and barriers and facilitators to delivery. We aimed to interview two members (staff and student) of each school's AG per year. A member of the evaluation team contacted the AG leader and asked them to identify two individuals (staff and student) to interview. Interviews were semi-structured, lasted 30–60 min and explored views on the acceptability of facilitators and the intervention, barriers/facilitators of AG meetings and how they might be improved, the extent to which actions arising from meetings were implemented in the school, and their impact.

Fidelity in the first two years was scored out of eight points for each school, assessing whether: (i) at least five staff attended in-depth training (training records); (ii) six action-group meetings occurred per year; (iii) policies and rules were reviewed; (iv) locally decided actions were implemented (minutes); members reported that AGs had a good/very good range of members, and that groups were well/very well led (member survey); schools delivered at least five hours and/or two modules each year (curriculum survey); at least 85% of staff reported that if there was trouble at the school, staff responded by talking to those involved to help them get on better (staff survey). Fidelity in the third year was assessed using a narrower range of data, since the research team access to schools was reduced. Out of four, schools were scored on whether six AGs were convened, local decisions were implemented, schools delivered at least five hours and/or at least two modules, and at least 85% of staff reported that if there was trouble at the school, staff responded by talking to those involved to help them get on better.

Some aspects of the process evaluation occurred only in six case-study schools. These were purposively sampled in terms of diversity on percentage of free school meals, facilitator and the extent to which the school was responsive to intervention activities, as rated by the intervention facilitators three months into the intervention. In each year of intervention, we aimed to conduct one focus group with staff in each case study school, each involving 4–6 members of staff. Staff were purposively selected and invited to participate by the staff member liaising with the research team to include diversity according to degree of participation in the intervention and role within the school (including senior leaders, pastoral staff and classroom teachers). Each year, we conducted focus groups with students in each case study school, comprising 4–12 students directly involved in intervention activities (including AGs). Focus group discussions aimed to explore school culture and ethos, views about the delivery and impacts of the intervention. Focus groups were conducted in private offices on school premises facilitated by one researcher. All interviews and focus groups were audio-recorded and transcribed.

Analyses

Quantitative data were entered into CSpPro or Excel and then transferred and analysed in Stata. Quantitative analysis used descriptive statistics to assess intervention fidelity, satisfaction and empowerment. Qualitative data were managed in NVivo and analysed to explore views about intervention processes and contexts. Analysis of qualitative data was undertaken by EW and CB, who created an initial coding framework based on the research questions and theory guiding the intervention, with these codes augmented, refined and sub-categorised informed by an initial wave of coding inductively from the data. This employed in-vivo codes to identify recurring themes expressed in the data. A further wave of axial coding then explored inter-relationships with these in-vivo codes. Although primarily focused on recurrent themes in the data, coding employed some methods popularised in grounded theory approaches, such as constant comparisons and examination of deviant cases to refine the emerging analysis (Green & Thorogood, 2004). The aim of the analyses was to establish a hierarchical set of themes which informed the structure of the narrative presented below. We used an adapted version of the Learner Empowerment Scale (LES) to assess the extent to which staff and student members felt empowered on the AG (Frymier *et al.*, 1996). We report the mean percentage of positive responses across all items.

Ethics

The trial was approved by the UCL research ethics committee (ref. 5248/001). Written, informed consent was obtained from head-teachers for random allocation and for intervention, and from participating individuals for data collection. Participants received written information about their rights, consent, confidentiality, safeguarding and data management, and verbal information on the day of the research. Participants were informed that, were they to disclose that they or others had been involved in or were at risk of sexual or physical abuse, the research team would liaise with the safeguarding lead for the school, breaking confidentiality. No such reports were made. Participants were then asked to indicate their consent and reminded that they could cease involvement in the research at any time. Parents of students participating in data collection also received prior information and could withdraw their children from research if they wished. All data collected were stored on password-protected drives with access limited to those in the research team analysing the data.

Results

Response rates

Most elements of the process evaluation achieved good response rates (Table 1) but some, such as the survey of AG members in year 3, had lower response rates, reflecting that in this year not all schools delivered all aspects of the intervention.

Table 1. Response rates

			Year 1	Year 2	Year 3	Total
All schools	Interviews	Facilitators	8/8	8/8	N.A.	16/16
		Action group members	28/40	41/40	28/40	97/120
	Observations	Action group	10/10	9/10	7/10	26/30
		Action group members	228/240 ¹	184/240 ¹	32/240 ¹	444/720
	Diaries	Facilitators	115/120 ²	101/120 ²	N.A.	118/240
	Minutes	Action groups	115/120	101/120	N.A.	216/240
Case study schools	Survey	Action group members	228/240 ¹	184/240 ¹	32/240 ¹	444/720
	FGDs	Students	6/6	5/6	4/6	15/18
		Staff	6/6	5/6	4/6	15/18

¹Denominator assumes 12 AGM members per intervention school.

²Denominator assumes 6 meetings per year for 20 schools.

Intervention fidelity

AGs were delivered with good fidelity in the first two years but lower fidelity in the third. In the first year, 19 schools completed all six AG meetings as intended, but one convened only four meetings. In the second year, 11 schools completed six meetings, four held five, three schools held four and one held three. Eleven schools achieved the target of holding at least six meetings per year in both years 1 and 2. Of these schools, only two maintained this level of meetings into the third year of intervention. In year 3, four schools did not hold any meetings.

Needs assessment surveys were completed in all schools with good response rates. Responses fell below 70% in only four schools, in year 3 only. According to facilitator diaries and meeting minutes, all schools in year 1 and 14 in year 2 reported using the needs data to inform priorities and actions. In year 3, four schools reported using the needs report to inform decisions. AGs at 17 schools succeeded in reviewing school policies and rules during years 1 or 2.

All schools enabled their staff to receive introductory training in restorative practice and all but five schools sent at least five staff on the in-depth training. In all but eight schools, staff reported that restorative practice was used extensively. The curriculum was less well delivered, with nine schools delivering the minimum content in the first two years and only five doing so in the third year.

Coordinating action

As the data above indicate, most AGs successfully ensured that their school staff were trained in and delivered restorative practice, but they were less successful in ensuring delivery of the curriculum. Most AG members in the majority of schools agreed that the group made good decisions (Table 2). Members were less likely to agree that the group made sure these decisions were implemented, and this varied between schools.

This was also a common theme in interviews with facilitators and staff. According to its facilitator, one school's group revisited minor issues but was largely unable to confront bigger issues relating to student behaviour, health or wellbeing. In another school, the AG proposed numerous actions that were rejected by the head-teacher.

Table 2. Member satisfaction with AGs

Measure (response option which indicates satisfaction)	Responses											
	Year 1			Year 2			Year 3					
	Overall % members responding as indicated	N schools where >70% of members responded as indicated	N schools with no data	Overall % members responding as indicated	N schools where >70% of members responded as indicated	N schools with no data	Overall % members responding as indicated	N schools where >70% of members responded as indicated	N schools with no data	Overall % members responding as indicated	N schools where >70% of members responded as indicated	N schools with no data
Was the needs assessment report useful in helping the AG decide what actions to take? (somewhat or very)	91%	17	1	88%	14	3	76%	2	17			
Was the external facilitator useful in ensuring that all AG members could have their say? (somewhat or very)*	89%	18	1	90%	15	3	N.A.	N.A.	N.A.			
Was the external facilitator useful in helping the AG decide what actions to take? (somewhat or very)*	86%	16	1	88%	17	3	N.A.	N.A.	N.A.			
Was the external facilitator useful in helping to ensure that actions were actually implemented? (somewhat or very)*	80%	13	1	79%	13	3	N.A.	N.A.	N.A.			
Do you think the Learning Together project was a good way to ensure students contribute to decision-making at this school? (very or quite)	95%	19	1	94%	16	3	100%	3	17			

Table 2. (Continued)

	Responses					
	Year 1		Year 2		Year 3	
Measure (response option which indicates satisfaction)	Overall % members responding as indicated	N schools where >70% of members responded as indicated	Overall % members responding as indicated	N schools where >70% of members responded as indicated	Overall % members responding as indicated	N schools where >70% of members responded as indicated
Overall, do you think the AG made good decisions about what actions to take? (% very or quite)	94%	19	93%	16	98%	3
Do you think the AG made sure that these actions were implemented? (% yes)	70%	10	72%	10	69%	1
		1		3		17

*No external facilitator in year 3.

The amount of work already facing teachers, within the pressurised culture of the English education system, made implementation more difficult. According to one facilitator, the teacher leading the AG in one school had:

been given so much to do I think in his day-to-day role, it's just another thing for him to do.
(Facilitator, school BE, interview, year 2)

There was a clear trend that schools were better able to implement actions when members of schools' senior management sat on the AG or otherwise supported it. Head-teachers rarely sat on action groups, but in many schools deputy head-teachers led the groups and their presence enabled decisions to be implemented because these individuals possessed the authority to institute change. In other schools, the chair received little support from the head-teacher or other school leaders:

The head teacher there was completely uninterested when I came to give a talk to the senior leadership team, and it's the only school where the head has not said anything to me or taken any notice or made any effort to be slightly friendly. It was awful. (Facilitator, school AH, interview, year 2)

In another school, despite there being no senior leaders on the group, the lead had worked for a long time at the school and was well respected and liked by both students and staff. Thus, it was possible to galvanise action without the formal involvement of senior leaders in some cases.

Using local information

Our survey of AG members found that most individuals across most schools rated the needs assessment report as useful in informing decisions (Table 2). Interviews with students and staff on AGs, as well as with facilitators, found that most valued the needs reports and used them to identify priorities:

They've been drawing on the needs assessment report [...] It's feeding into everything they do. So they... we focused on it quite a lot in meetings, pored over the data in the first few meetings. And then we still keep coming back to it when we're thinking about what next, what actions should we be doing? (Facilitator, school BM, interview, year 1)

Use of data on student progress is normal within English schools, and staff were used to employing data in decisions. Staff commonly suggested that they used the reports as evidence to advocate to school senior leadership teams for changes. In one school, for example, the needs report was useful to advocate on behalf of students whose concerns, though previously voiced, had not been acted upon. This was also the case in another school:

It certainly gave me some ammunition that I can say, 'Well this is... this is proven because the kids have actually said this, so we need to move forward with it'. (Assistant head-teacher, school AF, interview, year 2)

In a few schools, staff expressed shock or surprise at the results, particularly about the data on students feeling unsafe at school or uncomfortable talking to staff about problems. The reports often highlighted student concerns that staff had previously never

considered, particularly relating to feelings of safety at school. Another school received a report that presented evidence of considerable student unmet needs and this information was accepted by school leaders:

Well the positive thing is, the only way is up. The year group is a challenging year group. And I remember when [facilitator] came to present to [senior leadership team] and said how terrible our data was. . . it was like a tumbleweed moment; it was so funny. I mean. . . it wasn't funny in a good way, but. . . but it was a realistic. . . realisation for everyone if you know what I mean. . . Because we all knew it was like that, but we didn't realise how much the children didn't actually like us. (Member of pastoral team, school AM, interview, year 1)

Staff at two schools felt the report was not useful or valid. In particular, they felt the questions used vocabulary that was too advanced for many students with special educational needs or those with English as an additional language. They reported that the phrasing, particularly when students were asked to respond to negative statements, was inaccessible and that many of the questions were leading. In one school, the assistant head reported that when the school conducted its own surveys, their findings diverged from the intervention's needs report:

The data from the [AG report] is meaningless; completely meaningless. . . It's the worst. . . it's a pointless exercise for our students. (Assistant head-teacher, school AE, interview, year 3)

Staff in three schools did not share some or all of the findings with students because it was seen as too long and complicated or inappropriate for students to see. Where year-on-year trends in such factors did not improve, staff members sometimes reported feeling dispirited. Such disappointments may have contributed to these schools not continuing intervention activities in the final year of the study.

External facilitation

Most AG members rated the facilitator as useful for ensuring all members could have their say (Table 2), helping groups make decisions and ensuring that the decisions were implemented, though this varied across schools. A key theme across interview data was many staff's mixed feelings about the facilitation. Some schools reported that the facilitator was relatively passive, listening and taking notes, while others felt that they simply did not contribute much or were generally dissatisfied with the facilitator's work:

They don't really do anything. . . Whether they bring anything to the meeting. I think once or twice they might have asked a couple of questions, but that's about it. They sit there looking to us. (Role unknown, school AF, interview, year 1)

Other schools were not clear on what the role of the facilitator was meant to be and therefore had trouble setting reasonable expectations for what they should contribute to the group:

I think. . . [facilitator] has been really useful as a. . . So making sure we're doing what we're supposed to do in one respect is great. But then it's sometimes flipped into, well are we doing it right? And having [facilitator]'s presence there has always. . . are we being watched, judged or supported? (Senior leadership team member, school AW, interview, year 2)

Staff in other schools reported that the facilitator added gravitas to the meetings and reinforced the notion that they would be held accountable for progress:

So working with a facilitator from outside, that has been quite good at making it... just more... focused and more effective in that way. Because I think people in schools have really, really good ideas but again, because of the time, you tend to let things slip. Whereas this has kind of imposed a formality to it which means you have to keep to deadlines and move things forward. (Staff member, school AE, interview, year 1)

Most interviewees suggested that external facilitation was not necessary in the final year, but a few suggested this was a significant loss:

The absence of [facilitator] has been incredibly significant because she... was able to tie it in all the time to the agenda. And was a touchstone I suppose really for that. And then... so that... I think that was a loss. (Senior leadership team member, school AD, interview, year 3)

Local decision-making and adaptation

AGs were effective as a means of making local decisions about planned adaptations to how the intervention was delivered in each school. This approach was well aligned with the autonomy over management decisions granted to schools in England. As recommended in the intervention manual, most schools used the survey data on local needs to decide local actions. Sixteen schools enacted local decisions in years 1 and 2. In year 3, 10 schools completed locally decided actions.

Common actions included: cascading training in restorative practice to staff who had not attended the in-depth, or to students so that they could work as peer mentors or buddies; delivering assemblies on restorative practice; staff patrolling hallways between classes to discourage aggressive behaviours; and instituting safe spaces on the school site. These actions appeared highly consistent with the intervention theory of change in their direct focus on student safety. Other actions appeared consistent with the theory of change in terms of aiming to increase student engagement in school. Some schools offered more after-school clubs. One offered drop-in services to improve engagement and various schools focused on girls' engagement. Also in line with the theory of change were actions to promote mental health. One school, for example, funded new, external specialist staff to work with students to improve mental health and wellbeing, including a counsellor and a boxing coach. Other actions were less obviously linked to the theory of change but might be construed as contributing to increasing engagement. Some schools made improvements to the physical environment, including decorating the schools with informational or motivational posters and displaying student work.

A few AGs deviated from the processes of working set out in the manual. Two schools broke the AG into multiple sub-committees. This was done in one school so that students were less intimidated speaking in front of a large group and, in the other school, so that multiple groups could get more done. Some schools used AG meetings for purposes other than to coordinate the intervention. A few schools used AGs to train students or staff in restorative practice or mentoring. One academically selective school did not view itself as having significant

problems with bullying or aggression, and therefore used the AG primarily as a way to revise the homework policy.

Participation

We drew on facilitator diaries and minutes to estimate the numbers of staff and students participating in groups. On average in year 1, 7.4 staff per school (range 3–12) participated in groups in schools for which data were available, with this falling to 5.4 (range 2–9) staff per school in year 2. On average in year 1, 7.8 students per school (range 5–14) participated in AGs in schools for which we have data, with this increasing to 8.8 (range 4–23) students per school in year 2. We did not calculate figures on participation in year 3 because schools implemented AGs in such heterogeneous ways. Schools were encouraged to, and most did, select a diversity of students to serve on the group, including by gender and ethnicity, those with a history of misbehaviour and those who struggled academically or were at risk of disengagement. The schools involved served ethnically diverse populations and AG membership reflected this. A few focused recruitment on students who were high achievers or dropped lower-attaining students from the AGs. Most schools selected students across school years (aged 11–16 years) to participate, but some invited only those in the study cohort (i.e. aged 12–13 years at the outset).

For the LES, the mean percentage of positive responses across all items among group members in intervention schools for which data were available was high (Table 3). Over two-thirds of members completing the survey reported that they worked hard on the group because they wanted to, not because they had to (Table 2). Over 80% reported looking forward to meetings and almost 70% reported that meetings were exciting and energising. Most members agreed that the group was a good way to ensure students contributed to decision-making, with the results consistent across schools and years (Table 2).

In qualitative research, students said they valued the opportunity to express their views and highlight where they were unhappy, and AGs provided a calm, structured environment to do this:

I was able to speak from my own point of view, not just like statistics and what was on a piece of paper. Because I was able to like... I started... talking about my own experiences, the teachers were acknowledging it and... It kind of felt good because it's like... it's not me having to shout at a teacher or like I want to do this... It's difficult for a teacher with a class of thirty or more students. (Male, Year 9 student, school BE, interview, year 2)

Students often reported that they felt teachers in the group listened to students' views respectfully even when they disagreed with them:

That was good because we didn't feel intimidated by the teachers. Not that you felt intimidated before, but you didn't feel like... you could say something, all the teachers would listen... everyone put across their own views with no arguments or anything. (Student of unknown sex and year, school BE, focus group, year 2)

Students' confidence grew as a result of their involvement in the group:

Table 3. LES for AG members

LES item (response option which indicates empowerment)	Responses											
	Year 1				Year 2				Year 3			
	Overall % members responding as indicated	N schools where >70% of members responded as indicated	N schools with no data	Overall % members responding as indicated	N schools where >70% of members responded as indicated	N schools with no data	Overall % members responding as indicated	N schools where >70% of members responded as indicated	N schools with no data	Overall % members responding as indicated	N schools where >70% of members responded as indicated	N schools with no data
During AG meetings, I felt confident that I could do what was expected (yes)	90%	18	1	91%	14	3	98%	3	17	17	17	
I had the power to make a difference to how things were done at AG meetings (yes)	78%	14	1	82%	10	3	89%	3	17	17	17	
The AG's work fit with what I believe in (yes)	88%	17	1	92%	17	3	89%	3	17	17	17	
My participation was important to the success of the AG (yes)	83%	16	1	81%	14	3	86%	3	17	17	17	
Other people in the AG made me feel like I was not good enough (no)	92%	19	1	79%	15	3	89%	2	17	17	17	
I actively took part in the tasks undertaken by the AG (yes)	90%	19	1	89%	16	3	91%	3	17	17	17	
I usually did more work than I had to do in the AG (yes)	23%	0	1	34%	0	3	33%	0	17	17	17	
I was overwhelmed by all of the work required of the AG (no)	75%	14	1	74%	15	3	86%	3	17	17	17	
I worked hard in the AG because I wanted to, not because I had to (yes)	83%	17	1	83%	14	3	84%	2	17	17	17	
I had a choice in the way I went about doing work in the AG (yes)	86%	17	1	90%	17	3	86%	3	17	17	17	
The things I had to do in the AG meant a lot to me personally (yes)	61%	5	1	74%	10	3	69%	1	17	17	17	
I liked to talk about what I was doing in the AG with friends or family (yes)	55%	2	1	61%	11	3	66%	1	17	17	17	
I felt nervous about what was expected of me in the AG (no)	76%	15	1	73%	12	3	80%	2	17	17	17	
I was able to affect the way things were done in the AG (yes)	76%	12	1	84%	14	3	75%	1	17	17	17	
I looked forward to the AG meetings (yes)	73%	11	1	84%	13	3	91%	3	17	17	17	

Table 3. (Continued)

LES item (response option which indicates empowerment)	Responses								
	Year 1			Year 2			Year 3		
	Overall % members responding as indicated	N schools where >70% of members responded as indicated	N schools with no data	Overall % members responding as indicated	N schools where >70% of members responded as indicated	N schools with no data	Overall % members responding as indicated	N schools where >70% of members responded as indicated	N schools with no data
Those leading the AG believed that they must control how I contributed to the AG (no)	63%	9	1	61%	4	3	79%	2	17
I got positive responses when I expressed my own attitudes and ideas in the AG (yes)	85%	15	1	92%	17	3	84%	2	17
I agree with the things that were done in the AG (% yes)	90%	17	1	94%	16	3	93%	3	17
I had the skills needed to contribute to the AG (yes)	89%	18	1	91%	17	3	80%	2	17
My ability to contribute to the AG was under my control (yes)	85%	17	1	88%	13	3	82%	2	17
Those leading the AG felt that they were always right (no)	71%	11	1	69%	9	3	72%	2	17
I found the AG to be exciting and energising (yes)	62%	17	1	71%	8	3	84%	2	17
I found the AG to be interesting (yes)	87%	17	1	90%	15	3	86%	1	17
The things I did in the AG were valuable to me (yes)	75%	14	1	80%	13	3	78%	2	17
The things I learned and did in the AG will be helpful for my future (yes)	77%	13	1	86%	15	3	78%	1	17
This AG was one of the first times I felt like I was able to contribute to something important (yes)	43%	0	1	45%	3	3	27%	0	17
This AG taught me how to work well together with others (yes)	56%	3	1	64%	3	3	75%	1	17
This AG helped me better understand some of the problems in my school (yes)	84%	14	1	88%	15	3	91%	3	17
This AG gave me the chance to do something about the problems in my school (yes)	82%	15	1	92%	16	3	88%	2	17
Mean across all items	75%	13	1	79%	12	3	80%	2	17

Well at the beginning I was very conscious of not... sounding like an idiot, making a mistake, stuttering, something stupid like that. Now... it's completely different. I will... sometimes I will interrupt people. Obviously it's not a good thing but it just shows how I feel I can say what I want in this room. (Student of unknown sex and year, school BM, focus group, year 1)

But a theme across some staff accounts was staff's struggle to create positive interactions with students. Facilitator and interviews suggested that while many students participated meaningfully, others remained disengaged or, at times, combative:

They went down the road of all choosing students who were all, I would say, challenging. And therefore not necessarily always making an appropriate contribution to the group. I mean when I think back on it, out of the six students two of them contributed really well, and the other four were really a challenge. (Facilitator, school BD, interview, year 2)

Interviews and focus groups suggested that while student views were described as being taken seriously or very seriously in many schools, in other schools, staff only took note of student views when these were deemed 'realistic'. During the interview with a staff member at one school, these tensions became apparent:

They [students] want some rules to change; so they want to be able to bring their mobiles... There were a lot of conversations about, 'no... that's not [going to happen] ...' And I think that they think they had a bigger input than they perhaps did. (Assistant head-teacher, school AH, interview, year 2)

But other schools were enthusiastic about student participation in decisions. In one school, staff in the group expanded the process of revising school rules so that this involved students beyond the AG alone:

It came about through the Learning Together project. Very nice indeed, because the way it works... so the needs analysis said 'we don't know what the rules are'. [...] She got all the forms in form times with their form tutors to come up with rules about living in a community. How should we treat each other at [school AU]? And each form came up with you know lots of suggestions... I then took it to the staff and they had their say, and there were a lot of common themes, and we picked out all the favourites. I then took it to leadership and we narrowed it down even more. (Assistant head-teacher, school AU, interview, year 2)

A recurring theme concerned the impact of the work on mutual understanding between staff and students:

It's also nice to be able to have a conversation with a teacher, because very often they're just the people at the front of the class... And they're just giving you the work; but when you actually have a conversation and you get to know them, you understand that they're real people and they're a lot more relatable. (Female, Year 10 student, school AU, interview, year 2)

This could lead to better relationships between staff and students:

I think that's really good for me to be in something that's quite positive, because I do deal with quite a lot of negative in my role. So for the students to actually see me in a positive light within things like this I think is good for my role. Because then they'll stop and talk to me in the corridor: 'Miss, are you going to the meeting tonight? Yeah, yeah, are you going?' So it's nice just to have that. And other people see them talking to me and go 'Oh OK, she's not telling everyone off or talking about problems'. (Role unknown, school BM, focus group, year 1)

Discussion

Summary of key findings

Regarding our first question about the fidelity of implementation, AGs were implemented well when an external facilitator supported them in the first two years of intervention but fidelity was reduced in the final year when this was withdrawn. However, some schools carried out the functions of the AG in the third year through pre-existing structures, such as student councils, representing a 'mainstreaming' of AG functions. Regarding the second question about the role of AGs in intervention coordination, AGs were successful in ensuring staff were trained in, and implemented, restorative practice but were less able to ensure implementation of the social and emotional competencies curriculum. This may have reflected the absence of curriculum coordinators from the groups or a lack of school support for the curriculum element. Groups were commonly led by deputy head-teachers (whose authority ensured decisions could be rapidly implemented) rather than by head-teachers.

Regarding the third question about the importance of external facilitation and data on student needs, schools were generally satisfied with external facilitators and recognised their importance in galvanising action but felt that their roles could have been better defined. Some schools felt that facilitation could have been conducted internally without hampering the intervention. Data on student needs were an important resource for many schools, building commitment for the need for action and informing priorities and local actions. Several schools initially found the results shocking and two rejected the information and saw it as not useful.

Regarding the fourth question about AGs' approach to intervention adaptation, AGs succeeded in deciding locally appropriate actions in each school based on needs data and guided by a clear theory of change. Regarding the fifth question about acceptability, AGs were regarded as appropriately diverse, well led and able to make good decisions but sometimes as less able to ensure implementation. This was particularly so where groups did not include or were not supported by senior leadership teams. The AGs appeared to be an effective way to engage and empower staff and students in decision-making. Although some staff expressed doubts about their success in working with students, most students reported that they felt their views were taken seriously, and that working collaboratively with staff had transformed their relationships with staff and their experiences of school. We found no evidence that the work of AGs was more challenging or less acceptable because of their focus on mental health and bullying rather than merely physical health.

Limitations

We undertook a broad, deep and longitudinal process evaluation of the Learning Together intervention, which enabled examination of the workings of AGs. Although generally high, some aspects of the research had lower response rates. Despite this, our multi-source approach meant that we were still able to assess what was happening in most schools, with the exception of a few schools in year

3 where it is reasonable to assume that few if any intervention activities were being implemented.

Implications for research and policy

In line with previous literature, our study suggests that AGs are a feasible and acceptable means of implementing whole-school health interventions in secondary schools, providing a mechanism for coordinating and adapting complex whole-school interventions (Bond *et al.*, 2001; Langford *et al.*, 2014) via a process that enables broad participation and ‘student voices’ that is not tokenistic (Department for Education, 2014). Even though the work of AGs was taken forward in some schools via existing structures, our results suggest that use of a novel group was important to catalysing action. The AGs ensured that local decisions and adaptations occurred through a defined process so that locally appropriate adaptation nonetheless maintained the integrity of the intervention’s theory of change, offering a practical means of achieving something previously discussed as a concept in the literature (Hawe *et al.*, 2004). Like the Gatehouse project, our findings suggest that AGs can be effective in empowering diverse members to contribute to decisions (Bond *et al.*, 2001) although for some schools, the inclusion of students regarded as challenging was perceived to make the AGs harder to run. However, in most schools the fact that these groups had a clear focus, drew on rigorous evidence of student need and were managed in a business-like manner appeared to encourage students, including those regarded as challenging, to treat them seriously.

Our findings also suggest that AGs may be an effective enabler of ‘disruption’ to school systems, bringing about benefits both through formal channels (for example, changes in school policies and practices) and via more informal routes (for example, students and staff developing more empathetic perspectives on each other and thereby improving relationships), as suggested in previous literature (Markham & Aveyard, 2003; Bonell *et al.*, 2013b). Our sense is that the AGs could encourage schools to implement more inclusive and restorative (rather than authoritarian) approaches (Morrison, 2005), but that this was most feasible in those schools which already had a critical mass of staff who were supportive of such approaches. The intervention on its own would be unlikely to challenge school cultures where these were more authoritarian.

Participating schools were representative of those invited for participation and included a good range in terms of attainment, deprivation and inspectorate ratings, suggesting that similar benefits might be achieved were the intervention to be implemented in other secondary schools in England and beyond. We found some evidence that cultural specificities of the English education system enabled (for example, via the managerial autonomy schools are granted, and via teachers’ regular use of student data to inform decisions) but also constrained (for example, via staff time pressures and work loads) the success of AGs. Other contextual enablers and barriers might operate in other systems and cultures and this should be explored in international studies.

This study focused on young people aged 11–15 in secondary schools, and further research is required to explore whether AGs can be implemented in primary schools

with younger students. It might be that some aspects of AGs' work, such as the encouragement of staff and students to empathise with each other's perspective, are less feasible in primary schools because younger children will generally have less well-developed perspective-taking capacity (Dumontheil *et al.*, 2010). While our trial found evidence that the intervention was more effective in reducing health-related risk behaviours among boys than girls (Bonell *et al.*, 2018), our process evaluation found no evidence that membership of—or participation during—AGs was greater for boys. Future whole-school interventions involving AGs should aim to promote gender equity both in AGs and other intervention elements, and evaluations should assess impacts by gender.

To optimise AGs in future intervention studies, more thought should be put into defining the role of facilitators and supporting groups to interpret and use data on needs, given that both our study and the Gatehouse project found that external facilitation was likely to be important (Bond *et al.*, 2001). It may be that schools could use mutual-aid models in which one school's AG was facilitated by a staff member from another school in the same area, chain or network. Our evaluation also provides further evidence that implementation of school-based interventions is facilitated by a local 'product champion', preferably a deputy head or other senior manager with the authority to drive action and an existing culture sympathetic with the aims of the intervention (Johnson *et al.*, 2004; Bonell *et al.*, 2009, 2013a). While our study concluded that the direct involvement of head-teachers was not required, it is important that AGs operate with support from and communication with the school leadership teams.

Our evaluation examined AGs that focused on universal, primary prevention of bullying and other health-related risk behaviours by modifying the school environment. Our AGs did not aim to link with broader internal or external school inspection or improvement programmes, target the most disadvantaged students or promote outreach to the community, for example linking marginalised families with specialist support (Downes & Cefai, 2019). There is a need for further empirical evaluations to assess whether interventions involving AGs are feasible and effective in address these broader issues.

Acknowledgements

We mourn the loss and celebrate the life of Dr Farah Jamal who co-designed the process evaluation on which this article is based, and died aged 30 during the study. We are grateful to the staff and students of participating schools for their dedication to the intervention and completion of the outcome surveys and process evaluation surveys and interviews. We are very grateful for the advice and support of our trial steering committee and data monitoring committee. Miranda Perry played a key role in developing the intervention and coordinated its implementation in the pilot and phase III trials. This article presents independent research funded by the National Institute for Health Research (NIHR) in England under its Public Health Research programme (12/153/60) and the Education Endowment Foundation. The views expressed in this publication are those of the authors and do not necessarily reflect those of the National Health Service (NHS), the NIHR or the Department of Health for England.

References

- Basen-Engquist, K., Coyle, K., Parcel, G. S., Kirby, D., Banspach, S. W., Carvajal, S. C. *et al.* (2009) School wide effects of a multicomponent HIV, STD and pregnancy prevention program for high school students, *Health Education and Behavior*, 28(2), 166–185.
- Bond, L., Glover, S., Godfrey, C., Butler, H. & Patton, G. (2001) Building capacity for system-level change in schools: Lessons from the Gatehouse project, *Health Education and Behavior*, 28(3), 368–383.
- Bond, L., Patton, G., Glover, S., Carlin, J. B., Butler, H., Thomas, L. *et al.* (2004) The Gatehouse Project: Can a multilevel school intervention affect emotional wellbeing and health risk behaviours?, *Journal of Epidemiology and Community Health*, 58(12), 997–1003. <https://doi.org/10.1136/jech.2003.009449>
- Bonell, C., Sorhaindo, A., Strange, V., Wiggins, M., Allen, E., Fletcher, A. *et al.* (2009) A pilot whole-school intervention to improve school ethos and reduce substance use, *Health Education*, 110(4), 252–272.
- Bonell, C., Jamal, F., Harden, A., Wells, H., Parry, W., Fletcher, A. *et al.* (2013a) Systematic review of the effects of schools and school environment interventions on health: Evidence mapping and synthesis, *Public Health Research*, 1(1), 1–320.
- Bonell, C. P., Fletcher, A., Jamal, F., Wells, H., Harden, A., Murphy, S. *et al.* (2013b) Theories of how the school environment impacts on student health: Systematic review and synthesis, *Health & Place*, 24, 242–249.
- Bonell, C., Allen, E., Christie, D., Elbourne, D., Fletcher, A., Grieve, R. *et al.* (2014) Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): Study protocol for a cluster randomised controlled trial, *Trials*, 15, 381.
- Bonell, C., Allen, E., Warren, E., McGowan, J., Bevilacqua, L., Jamal, F. *et al.* (2018) Initiating change in the school environment to reduce bullying and aggression: A cluster randomised controlled trial of the Learning Together (LT) intervention in English secondary schools, *The Lancet*, 392(10163), 2452–2464.
- Bronfenbrenner, U. (1979) *The ecology of human development: Experiments by nature and design* (Cambridge, MA, Harvard University Press).
- Coyle, K., Basen-Engquist, K., Kirby, D., Parcel, G., Banspach, S., Harrist, R. *et al.* (1999) Short-term impact of safer choices: A multicomponent, school-based HIV, other STD, and pregnancy prevention program, *Journal of School Health*, 69(5), 181–188.
- Coyle, K., Basen-Engquist, K., Kirby, D., Parcel, G., Banspach, S., Collins, J. *et al.* (2001) Safer choices: Reducing teen pregnancy, HIV, and STDs, *Public Health Reports*, 116(Suppl. 1), 82–93.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. & Petticrew, M. (2008) Developing and evaluating complex interventions: The new Medical Research Council guidance, *British Medical Journal*, 337, a1655.
- Dane, U. A. & Schneider, B. H. (1998) Program integrity in primary and early secondary prevention: Are implementation effects out of control?, *Clinical Psychology Review*, 18(1), 23–45.
- Department for Education (2014) *Listening to and involving children and young people* (London, Department for Education).
- Downes, P. & Cefai, C. (2019) Strategic clarity on different prevention levels of school bullying and violence: Rethinking peer defenders and selected prevention, *Journal of School Violence*. <https://doi.org/10.1080/15388220.2019.1566915>
- Dumontheil, I., Apperly, I. A. & Blakemore, S. J. (2010) Online usage of theory of mind continues to develop in late adolescence, *Developmental Science*, 13, 331–338.
- Flay, B. R., Graumlich, S., Segawa, E., Burns, J. L. & Holliday, M. Y. (2004) Effects of 2 prevention programs on high-risk behaviors among African American youth: A randomized trial, *Archives of Pediatrics & Adolescent Medicine*, 158(4), 377–384.
- Frymier, A. B., Shulman, G. M. & Houser, M. (1996) The development of a learner empowerment measure, *Communication Education*, 45(3), 181–199.

- Glover, S. & Butler, H. (2004) Facilitating health promotion within school communities, in: R. Moodie & A. Hulme (Eds) *Hands-on health promotion* (Melbourne, IP Communications).
- Greaves, C., Sheppard, K. & Abraham, C. (2011) Systematic review of intervention components associated with increased effectiveness in dietary and physical activity interventions, *BMC Public Health*, 11, 119.
- Green, J. & Thorogood, N. (2004) *Qualitative methods for health research* (London, Sage).
- Hawe, P., Shiell, A. & Riley, T. (2004) Complex interventions: How 'out of control' can a randomised controlled trial be?, *British Medical Journal*, 328, 1561–1563.
- Hulpia, H., Devos, G. & Rosseel, Y. (2009) The relationship between the perception of distributed leadership in secondary schools and teachers' and teacher leaders' job satisfaction and organizational commitment, *School Effectiveness and School Improvement*, 20(3), 291.
- Johnson, K., Hays, C., Center, H. & Daley, C. (2004) Building capacity and sustainable prevention innovations: A sustainability planning model, *Evaluation Program Planning*, 27, 135–149.
- Langford, R., Bonell, C. P., Jones, H. E., Poulou, T., Murphy, S. M., Waters, E. et al. (2014) The WHO Health Promoting School framework for improving the health and well-being of students and staff, *Cochrane Database of Systematic Reviews*, CD008958.
- Markham, W. A. & Aveyard, P. (2003) A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice, *Social Science & Medicine*, 56(6), 1209–1220.
- Mathews, C., Eggers, S. M., Townsend, T., Aarø, L. E., de Vries, P. J., Mason-Jones, A. J. et al. (2016) Effects of PREPARE, a multi-component, school-based HIV and intimate partner violence (IPV) prevention programme on adolescent sexual risk behaviour and IPV: Cluster randomised controlled trial, *AIDS and Behavior*, 20(9), 1821–1840.
- Mihalic, S. (2004) The importance of implementation fidelity, *Emotional and Behavioral Disorders in Youth*, 4(4), 83–105.
- Morrison, B. (2005) Restorative justice in schools, in: E. Elliot & R. M. Gordon (Eds) *New directions in restorative justice: Issues, practices, evaluation* (Cullompton, Willan Publishing).
- Patton, G., Bond, L., Carlin, J. B., Thomas, L., Butler, H., Glover, S. et al. (2006) Promoting social inclusion in schools: Group-randomized trial of effects on student health risk behaviour and well-being, *American Journal of Public Health*, 96(9), 1582–1587.
- Quinn, D. M. & Kim, J. S. (2017) Scaffolding fidelity and adaptation in educational program implementation: Experimental evidence from a literacy intervention, *American Educational Research Journal*, 54(6), 1187–1220.
- Smith, D. J. (2006) *School experience and delinquency at ages 13 to 16* (Edinburgh, Centre for Law and Society, University of Edinburgh).

Introduction to the second study:

The previous publication explored how AGs were used to coordinate a complex, whole-school, health promoting intervention and increase student and staff participation in school improvement efforts. It also showed that the school's environment, ethos and organisation affected the groups' ability to implement other intervention components and processes central to LT. As our a priori CMOCs suggested, commitment to the intervention was noticeably lower in schools with a disciplinarian ethos (schools AH, AT, and BD). This was particularly clear in School AT, where the head teacher rejected every change suggested by the students. The data also suggested how very different schools often used resources in the same way, albeit for occasionally different purposes. Schools AU, BM, and BD all used RPs primarily to improve classroom behaviour. However AU and BM had low rates of bullying and aggression and thought RP would be an effective mechanisms through which Ofsted inspection ratings would remain 'outstanding', while BD urgently needed a school-wide strategy to reduce the significant levels of aggression which was making community members unsafe and costing meaningful amounts of teaching time.¹⁶ While not explicitly realist, the analysis showed that despite receiving the same resources, agents used them differently and in ways that were enabled and constrained by local structures. It also showed that implementation affects which mechanisms may be activated, deactivated or remain latent, and this varied hugely by the context.

The analysis also showed the degree to which some schools engaged more meaningfully with the explicit goals of LT, for example, by using the students' NAR to advocate for more active responses to students' concerns while, in other schools, leaders were less able or willing to engage, and acknowledged that the students' participation was at least partially tokenistic. It would be anticipated that schools that either did not or could not act upon student concerns (AH, AM, BD) would be unlikely to activate some of the key hypothesised mechanisms of change in relation to eroding boundaries or decreasing framing.

¹⁶ More detailed findings about fidelity are described in the study's final report (provided in Appendix 8 [294] starting on page 45. Tables describing fidelity by school and year and provided on pages 114 and 115 of Appendix 8.

After examining both fidelity of form and function,[\[343\]](#) we can now examine participant accounts to develop theory and hypotheses about mechanisms to explore how and why bullying decreased in some schools and not others, how LT resources were deployed in ways that interacted with the school environment to activate different mechanisms, and how those mechanisms appear to generate different outcomes in different contexts. It is in this way that we begin to open the “black box” which is central to realist evaluation. Pawson and Tilley have argued, “that experimental evaluation only produces descriptions of outcomes, rather than explanations of why programs work (or fail).”[\[1, pg. 30\]](#) As argued in Chapter 2 and will be demonstrated in the following chapter, this may be true of some trials but is not a necessary or inherent feature of them.

Typically, the qualitative component of an RCT’s PE focus on participants’ experience of their involvement in the intervention and sometimes the trial. Quantitative components within a PE are often used to describe fidelity, reach or acceptability.[\[344, 345\]](#) In 2004, Oakley et al argued that PEs need to move beyond describing implementation and the meaning participants ascribe to the outcomes to diving into the:

“Real-world complexity...[examining] how interventions work or do not work, the extent to which these are implemented as intended, and how the people exposed to them (or not exposed to them) react...[and] go some way towards meeting the needs of ‘realist’ or theory-driven evaluations.”[\[217, pg. 442\]](#)

The qualitative topic guides for INCLUSIVE dedicated substantial time asking participants to describe their school’s environment. These data were vital for developing CMOCs and understanding the role that context and individual’s choices and values impacted on the emergence of outcomes. Early iterative analysis of the first two years of data also made it clear the topic guides were too focused on implementation and perceived outcomes, and less well suited to understand mechanisms. Therefore, in the final year of the study, I incorporated techniques from realist interviewing [\[303\]](#) into the data collection procedures and asked participants *how* they understood LT to have “worked” (or didn’t work) and what enabled or limited those changes.

During realist interviews, researchers are encouraged to show participants the study’s logic model or theory of change and participants may be asked to confirm, refine, or refute the theory, offering their own view on how the intervention worked in their particular context. I did not show research participants the logic model for LT nor did I

explain the theory of human functioning and school organisation.^[7] However, in the final year of the study, I shifted the topic guides to focus more about why certain activities were done (what they hoped to accomplish in relation to their self-defined goals) and how they predicted these activities would lead to desired changes. During these interviews, I was surprised how frequently staff would use the word “mechanism” of their own volition, often describing what Dalkin refers to as “reason mechanisms”^[243] relating to changes in reasoning that led to new behaviours caused by motivations that had been highlighted through intervention components, such as the NAR or AG meetings. This slight change in interview technique enabled discussions that were richer in description of peoples’ experiences while potentially less useful for examining fidelity.

Using qualitative data to explore mechanisms is an important but currently under-investigated area of research.^[346] Even with an appropriately selected mid-range theory, those implementing or receiving the intervention may perceive previously unconsidered mechanisms. For example, as will be discussed in the following chapter, neither our mid-range theory nor our theory of change considered the role that shame or empathy would play in decreasing bullying and aggressive behaviour.

In a separate study, we used INCLUSIVE as a case-study to show how qualitative data can be used to explore causal mechanisms within complex health promotion interventions.¹⁷^[346] In the first instance, we can ask participants to explain how interventions work. One way to achieve this, especially relevant in a study underpinned by realism, is realist interviewing.^[347] One concern with realist interviewing is that individual accounts are too partial and perspectival to comment on how mechanisms functioned generally. However, by drawing multiple perspectives together, a broader more comprehensive picture should begin to emerge about which mechanisms may be plausible. This theory, built from participant accounts, can then be synthesised into hypotheses that can later be tested using quantitative data.

The second, less direct way to use qualitative data to understand mechanisms, is also used in the following chapter. Researchers are able to examine participants accounts to

¹⁷ See Appendix 1 for the manuscript.

draw together theories about what mechanisms were activated, how those varied by context, and what the resultant outcomes were. As described above, no participant is expected to have a broad enough view to articulate all the ways in which change emerged, but synthesis and comparison allows us to understand how and why change may or may not have occurred. Even though LT was based on a mid-range theory and had a theory of change, these were intentionally set aside to develop hypotheses and theories directly from participant accounts. This is not to say that I somehow forgot our original theories but that this analysis did not centre on them. As Ian Dey famously wrote about grounded theory, “there is a difference between an open mind and an empty head.”[\[348\]](#)

The following study, published in *Trials*, examines participants’ accounts of LT to develop testable hypotheses relating to the relationship between school structure, actors’ agency, mechanisms, and outcomes. I explored participants’ accounts of being in the school, the surrounding neighbourhood, the management structure, staff and student reports of problems, the use of intervention resources, and how the feelings, views, relationships and behaviours of participants changed over time. I also examined participants’ descriptions of how the school environment changed throughout the intervention period. This process was crucial because a key aim of WSIs is to change the school environment.

Chapter 5: Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention



London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646

F: +44 (0)20 7299 4656

www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	233090	Title	Ms
First Name(s)	Emily Ashbrook		
Surname/Family Name	Warren		
Thesis Title	Are realist randomised controlled trials possible? INCLUSIVE as a case study of an emerging methodology		
Primary Supervisor	Professor Chris Bonell		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Trials		
When was the work published?	September 2020		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

<p>For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)</p>	<p>I designed the analysis plan and led on the analysis and interpretation of data as well as drafting the manuscript. Professor G.J. Melendez-Torres helped design the analysis and interpret the findings. Professor Russell Viner deisgned the trial and approved the manuscript. Professor Chris Bonell deisgned the trial, helped collect, analyse, and intrepret the data, and draft the manuscript. All authors approved the manuscript prior to publication.</p>
---	--

SECTION E

Student Signature	Emily Warren
Date	28/01/2022

Supervisor Signature	Chris Bonell
Date	28/01/2022

RESEARCH

Open Access

Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention



Emily Warren^{1*} , G. J. Melendez-Torres², Russell Viner³ and Chris Bonell¹

Abstract

Background: This study reports on qualitative research conducted within a randomised controlled trial to explore possible intervention mechanisms. It focuses on the 'Learning Together' whole-school intervention delivered in secondary schools in England from 2014 to 2017 aiming to prevent bullying and aggression and improve student health. Intervention schools received staff training in restorative practice, a social and emotional learning curriculum, and an external facilitator and manual to convene and run a student/staff action group tasked with coordinating the intervention, focusing this on local needs.

Methods: Informed by realist approaches to evaluation, we analysed qualitative data to explore intervention mechanisms and how these might interact with school contexts to generate outcomes. Qualitative analysis drew on 45 interviews and 21 focus groups across three case-study schools and employed thematic content analysis to explore how intervention resources were taken up and used by local actors, how participants described the intervention mechanisms that then ensued, and how these might have generated beneficial outcomes.

Results: The thematic content analysis identified three social mechanisms that recurred in participant accounts: (1) building student commitment to the school community, (2) building healthy relationships by modelling and teaching pro-social skills, and (3) de-escalating bullying and aggression and enabling re-integration within the school community.

Conclusions: Our analysis provides in-depth exploration of possible mechanisms and the contextual contingencies associated with these, allowing refinement of the initial intervention theory of change.

Trial registration: ISRCTN registry 10751359. Registered on 11 March 2014

Keywords: Randomised controlled trials, Realism, Whole-school interventions, Mechanisms, Restorative practice, Action groups

* Correspondence: emily.warren@lshtm.ac.uk

¹Faculty of Public Health and Policy, London School of Hygiene & Tropical Medicine, London, UK

Full list of author information is available at the end of the article



© The Author(s). 2020, corrected publication 2020. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

This paper draws on qualitative data collected as part of the process evaluation within a randomised controlled trial (RCT) of the 'Learning Together' intervention to provide an in-depth description of participant accounts of the processes occurring, how these varied with local conditions in schools, and with what consequences. The paper then draws on this analysis to develop realist-informed hypotheses about how intervention mechanisms might interact with context to generate outcomes.

Whole-school health interventions to prevent bullying and promote health

Bullying and aggression are common among secondary school students [1, 2] with important consequences for educational attainment and adolescent and adult physical and mental health [3–6]. Whole-school interventions are a promising approach to promoting student health across a range of outcomes, including bullying and aggression [7–10]. One approach within such interventions for which existing trials report evidence of effectiveness is for schools to convene an action group of students and staff who identify local actions that will encourage a more inclusive and engaging social and learning environment, with positive consequences for student health [11–13]. In some interventions, group decisions are informed by local data on students' health needs and views about school [14, 15]. These have been found to be more effective if a member of the senior leadership team is on the action group and where students' participation is taken seriously by staff and not seen as a tokenistic exercise [15].

A second promising approach within whole-school interventions is restorative practice, which involves responding to conflict not merely by punishing perpetrators but by understanding the causes of conflict, improving relationships, and re-integrating offenders back into the school community. This may take the form of a staff member leading a facilitated meeting between a bully and their victim, the victim being given the opportunity to describe the impact of the bullying, the bully being encouraged to acknowledge this harm and their responsibility for it, and the facilitator working with the two parties to enable healing in their relationship and the prevention of further problems. Although observational studies of restorative practice have generally been positive, before the current trial, there had been no experimental evaluations of this approach in schools [16–19].

A third promising approach is classroom interventions promoting social and emotional learning, for which there is strong evidence from RCTs that these promote student mental well-being, reduce conflict, and improve academic engagement and attainment, especially among students receiving free school meals and those

underperforming in math and literacy, indicating a potential avenue for improving health equity [20–22]. Whole-school interventions are complex interventions involving multiple components that interact with each other and with context to generate emergent, socially contingent effects [23] theorised to be greater than had the components been introduced individually [24].

Drawing on this evidence, the Learning Together intervention was developed. Drawing on the work of Bernstein [25], the intervention was theorised to work by using action groups, restorative practice, and social and emotional learning to increase student commitment to the school's 'instructional' and 'regulatory' orders [26, 27]. The instructional order consists of processes of academic learning, and the regulatory order consists of school norms of behaviour and community [27]. Increasing student commitment to these orders occurs via a process of 'reframing': eroding 'boundaries' between and among students and staff, and between academic learning and broader student development. Existing theory suggests that, particularly among socio-economically disadvantaged students, such boundaries can persistently hinder student commitment to the instructional and regulatory school orders because relationships with teachers are insufficiently strong to engender a sense that school is something for them.

This can then contribute to students' increased involvement in risk behaviours. This may be partly through students not being educated on how to avoid such behaviours [27] and partly through students committing to anti-school peer groups and risk behaviours, such as violence and substance use, which function as markers of belonging and status among such groups. The latter might occur particularly when 'official' markers of success in school seem unattainable to students who do not feel committed to school [28]. Eroding boundaries is theorised to promote commitment to the instructional and regulatory orders and therefore reduce bullying and aggression and promote student mental and physical well-being, but is something that requires a significant whole-school change in order to address these persistent structural influences on adverse outcomes.

Learning Together was evaluated through an RCT. The trial's primary analyses indicated that the intervention was effective not only in reducing bullying victimisation but also in reducing smoking, drinking alcohol, using drugs, and in promoting mental well-being, psychological functioning, and health-related quality of life among students. Moderator analyses found effects were no greater for more socio-economically disadvantaged students but were greater for boys and for those reporting victimisation at baseline. There was also evidence that the intervention increased student commitment to school, providing some indirect evidence that the above theory of change might be plausible [29]. However, these

trial analyses offer limited insights into the mechanisms of complex interventions and how these might play out across contexts to generate different outcomes in different schools. As suggested above, whole-school interventions can be thought of as contextually contingent with a consequent need to understand mechanisms and how these vary across the context of different students, schools, and school systems [30, 31].

Randomised trials and realist evaluation

RCTs aim to estimate intervention effects while minimising confounding and other biases. However, RCTs have been criticised for being too narrowly focused on estimating the overall effects and failing to examine the mechanisms or how outcomes vary across contexts [32–34]. This view is supported by some reviews identifying but not explaining heterogeneity in effects among RCTs of similar interventions across different settings or populations [35, 36].

Realist evaluation aims to address this limitation by formulating and testing hypotheses about how contexts and mechanisms interact to generate outcomes, with such hypotheses being worded as context-mechanism-outcome (CMO) configurations [32]. Interventions are viewed as producing outcomes not directly but only via introducing resources into a setting which local actors may then use and in doing so may trigger ‘mechanisms’. Mechanisms are the consequences of people engaging with the resources of a programme or intervention in a certain context that bring about a change or effect. Mechanisms exist as tendencies which may or may not generate outcomes, contingent on ‘context’, or the conditions in which interventions are introduced. Context can influence whether intervention resources are taken up and used and thus trigger mechanisms (for example, because of social norms or structural constraints), as well as whether the mechanisms thus triggered will be sufficiently effective to generate ‘outcomes’ or instead be swamped by other mechanisms operating in that context. Outcomes represent the observable consequences of mechanisms [32]. Contexts, mechanisms, and outcomes are analytical categories that social enquiry applies to make sense of the world. Realist evaluators generally assess CMO configurations using observational (non-randomised) comparisons of interventions deployed across and within differing contexts. A realist approach might therefore be helpful in moving beyond mere effect sizes to explore the contextually contingent mechanisms and impacts of whole-school interventions such as Learning Together. Within a realist-informed approach, qualitative research should be useful in exploring participants’ accounts of intervention mechanisms to assess whether these align with those theorised.

Rationale for the current analysis

The present paper aims to draw on qualitative data collected within the trial’s process evaluation to explore participant accounts of intervention processes and how these might have played out with different consequences in different schools. The paper aims to develop a rich description of school conditions, intervention-related processes, and consequences from the perspective of the staff and students involved in implementing and receiving the intervention. Our ‘Discussion’ section then aims to reflect on these findings to consider their implications for the intervention’s underlying theory of change and to propose some CMO configurations in the light of our findings. Our qualitative research aimed to answer the following questions:

- (i.) How did intervention participants describe the school context, the processes involved in participation, and the consequences of these?
- (ii.) How did such accounts vary between schools, and what conditions relating to schools, staff, or students seem to explain these variations?

Methods

Trial design and methods

Trial methods are described in detail elsewhere [26, 29]. In summary, the intervention was evaluated from 2014 to 2017 using a superiority parallel-group cluster RCT randomly allocating (via computer-generated sequence) 40 state secondary schools (stratified by mixed/single-sex and rate of free school meal eligibility) in South East England to intervention or control (comprising usual practice) arms. Schools were recruited using email, phone calls, and a recruitment event targeting all eligible schools (excluding private schools, pupil referral units, schools exclusively for students with learning disabilities, and schools with ‘inadequate’ government inspection reports). Baseline surveys preceded allocation and consisted of self-completion questionnaires completed in privacy in classrooms by students nearing the end of year 7 (age 11–12) with similarly conducted follow-up surveys at 24 and 36 months. Head teachers consented to allocation and intervention. Surveys required students’ informed (written and oral) opt-in consent. Parents were informed and had the right to withdraw their children. The main trial analyses were intention-to-treat focused on primary and secondary outcomes at 36 months. Ethical approval for the trial was obtained through the Institute of Education Research Ethics Committee (18/11/13 ref. FCL 566) and the University College London Research Ethics Committee (30/1/14, Project ID: 5248/001).

The Learning Together intervention

The intervention was delivered over 3 years and offered the following resources to schools: an intervention manual, a yearly report on student health needs and views about the school from annual student surveys, an external facilitator trained by the lead facilitator and research team to facilitate action group (for the first 2 years with the third year being facilitated internally by school staff), and a yearly social and emotional learning curriculum (dose 5–10 h of lessons per year). Restorative practice training was provided through a specialist organisation. In the first year, all staff received 2–3 h of training to understand restorative terminology and practices and integrate these into their everyday interactions with students. A further 3-day training was provided for 5–10 staff selected by schools to deliver restorative conferences convened to address instances of bullying, aggression, or wrong-doing and, depending on the seriousness of the incident, also involving parents or other external participants such as the police. The training involved interactive discussions, group work, and role plays.

Using these resources, school staff were encouraged to enact the following: form action groups comprising at least six students and six staff who meet six times per year to collaborate on and review the needs report, and formulate and implement local decisions to address the needs identified; oversee implementation of the social and emotional learning curriculum; and review school policies and rules to ensure that these support a restorative, engaging, and inclusive school environment. Schools were asked to recruit diverse students, including those less engaged with school, onto action groups and ensure their full participation. External facilitators were trained to ensure student contributions with the aim of ensuring local decision-making, and implementation was co-produced between staff and students [24]. Schools were also tasked with implementing restorative practice including preventative interventions in classrooms to avoid conflict and more intensive restorative conferences to address conflict. All delivery was face-to-face on the school site. Fidelity was generally good although more variable for the curriculum and in the third year [37].

Qualitative data and analysis

The research took a qualitative approach to explore participant accounts of the way in which the intervention was viewed as being implemented and operating. Sampling was purposive in order to examine how accounts varied by factors thought likely to be important factors in the diversity of views expressed. The present analysis draws on qualitative data from three case study schools purposively sampled in the first year of the trial based on diversity regarding student

free school meal eligibility, school type, facilitator, and facilitator-rated progress with implementation.

In each case study school, we aimed to conduct annual interviews with the staff member leading the intervention, the staff member leading the SEL curriculum, a student and staff member of the action group, and two student participants in restorative conversations. Three annual focus groups were also held: one with students on the action group, one with other students, and one with staff. Interviews were held in the first 2 years with the external facilitator. Telephone interviews with two other school staff plus one senior leader were held at the beginning of the intervention. Interview guides and prompts were tested during the intervention pilot. Interviews and focus groups were arranged in consultation with the intervention lead in each school, who were asked to select students diverse by engagement and students/staff diverse by involvement with the intervention. As with surveys, qualitative research proceeded on the basis of participants' informed opt-in consent supplemented, in the case of students, with parental right of withdrawal. Except for phone interviews, all research occurred in private school rooms. Students were told that researchers wanted to know about their school generally, their opinions on bullying and aggression, and school-wide changes. All interviews were audio-recorded and transcribed in full. Interviews and focus groups lasted between 15 and 90 min. Field notes were also made after each data collection session (Table 1). Demographic information on participants was not collected.

The analysis involved thematic content analysis. The analysis sought to identify themes pertinent to our research questions inductively from the data. Data were organised in Nvivo QSR. Coding occurred transcript by transcript ordered chronologically, school by school, and in waves, with initial open codes identifying recurring themes in participant accounts, following by further axial coding which drew on open codes to develop a deeper analysis of the social phenomena under investigation: participant accounts of mechanisms and how these interacted with context to generate outcomes.

Although we used thematic content analysis, identification and interpretation of themes were nonetheless sensitised by ideas derived from a variant of a grounded theory known as dimensional analysis. The dimensional analysis aims to draw on qualitative data to delineate social phenomena in terms of their context (the boundaries of a phenomenon), conditions (factors that facilitate, block, or shape phenomena), process (actions or interactions involved in phenomena), and consequences (effects or outcomes of phenomena) [38, 39]. Although the terminology used within dimensional analysis differs from that of context, mechanism, and outcome used within realist evaluation, we felt that this framework of context,

Table 1 Data collection and response rate per school and year

Data source	Harper's School				Meadowood School				St Anselm's School				Totals by data collection type
	Y8	Y9	Y10	School total	Y8	Y9	Y10	School total	Y8	Y9	Y10	School total	
Action group interviews	–	2/2	0/2	2/4	–	4/2	0/2	4/4	–	2/2	2/2	4/4	10
Staff focus groups	1/1	2/1	0/1	3/3	1/1	1/1	0/1	2/3	1/1	0/1	1/1	2/3	7
Action group focus groups	1/1	1/1	0/1	2/3	1/1	1/1	1/1	3/3	1/1	0/1	1/1	2/3	7
General student focus groups	1/1	1/1	0/1	2/3	1/1	1/1	1/1	3/3	1/1	0/1	1/1	2/3	7
Student participant in restorative practice interviews	1/2	3/2	0/2	4/6	2/2	2/2	2/2	6/6	2/2	3/2	2/2	5/6	15
Curriculum coordinator interviews	0/1	0/1	0/1	0/3	0/1	1/1	1/1	2/3	0/1	1/1	0/1	1/3	3
Staff Interviews	3/3	–	0/1	3/4	3/3	–	1/1	4/4	3/3	–	1/1	4/4	11
Facilitator interviews	1/1	1/1	–	2/2	1/1	1/1	–	2/2	1/1	1/1	–	2/2	6
School totals				18				26				22	
Number of interviews and focus groups completed													66

conditions, processes, and consequences provided by dimensional analysis could be used as a heuristic device within our thematic content analysis, allowing us to dissect participant accounts of the mechanisms triggered by their engagement with Learning Together resources and how these were described as interacting with school context to generate outcomes. With this in mind, axial coding was built on open coding to, where possible, identify and group together open codes describing the context, conditions, processes, or outcomes of mechanisms as they were described in participant accounts. Our results are thus presented in terms of participant accounts of conditions, processes, and consequences. In the 'Discussion' section, we reflect on these findings to consider their implications using the realist language of context, mechanisms, and outcomes.

Results

Thematic content analysis identified a central theme of the intervention resources being used by staff and students to build an inclusive and cohesive school environment. This overarching theme in turn comprised three themes describing three mechanisms: (1) improving student commitment to the school community, (2) building healthy relationships by modelling and teaching pro-social skills, and (3) de-escalating bullying and aggression and enabling re-integration within the community. All three of these themes were apparent in participants' accounts as processes which had the potential to reduce conflict in schools and each comprised smaller processes. These themes, and their constituent sub-themes describing different aspects of these processes, are described below, in each case with a description of the conditions within which processes were reported to arise and their consequences. Table 2 depicts these conditions, processes, and consequences in more detail. First, however, we present a brief description of the overall

context of our case study schools, which define the boundaries of our empirical research. Pseudonyms are used throughout. We indicate student year groups 8–10 which involve students age 11/12 through to 13/14 years.

School contexts

Harper's School

Harper's School is located in a deprived area of inner London, with a high rate of free school meal eligibility and of English as an additional language among students. Teachers described Harper's as a school in which students often felt compelled to 'act tough' both in school and the surrounding community. Students described violence as being common. When asked how he felt after seeing one student break a boy's nose and a boy threatening a girl with a knife, the student responded:

I just thought...this is school. It's not supposed to happen, but it does. (Student, Harper's, RP interview, year 9)

Staff reported feeling overwhelmed and under-supported, and the school had volunteered for the trial in order to address this. While some teachers had warm relationships with students, others were described as violent. Students reported that good teachers were leaving the school:

[Teachers] that really cared, they already left the school. Because, like, the school is, like, getting worse. The teachers... the good teachers are now, like, ...they're all going. So, the bad ones, they're now staying. (Student, Harper's, RP interview, year 9)

Both students and staff commented that 'good' students did not want to be involved in anything beyond

Table 2 Conditions, processes, and consequences of Learning Together in case study schools

	HARPER'S	MEADOWOOD	ST. ANSELM'S
CONDITIONS:			
FACILITATORS	<ul style="list-style-type: none"> Committed AG Lead High deprivation 	<ul style="list-style-type: none"> Committed AG Lead High deprivation High staff engagement Moderate student commitment to school Dedicated pastoral care team 	<ul style="list-style-type: none"> Moderately committed AG Lead Students feel safe at school and in the community Very low violence in school Moderate engagement from staff Parents very engaged with school No violence reported in the community
BARRIERS	<ul style="list-style-type: none"> Students feel unsafe in community and at school High violence in schools Low student commitment to school High staff turnover 	<ul style="list-style-type: none"> Students feel relatively unsafe in community and at school Moderate violence in school Inconsistent school leadership 	<ul style="list-style-type: none"> Highly focussed on academic achievement Inconsistent AG leadership Low deprivation
↓ ↓ ↓			
LEARNING TOGETHER INPUTS:			
ALL SCHOOLS RECEIVED:	Needs assessment report (years 1-3)	Facilitation of AG meetings (years 1-2)	RP training (year 1) SEL curriculum (years 1-3)
↓			
PROCESSES 1: INCREASED COMMITMENT			
1. Learning other's perspectives	✓	✓	✓
2. Improving student/staff relationships	✗ Insufficient and inconsistent staff attendance	✓	✓
3. Participation creates new roles	✗ Lack of support for AG activities	✓	✓
4. AG participants initiate change in students' attitudes to school	✗ AG's ideas unsupported by most staff	✓	* Pro-school attitudes normalised before trial
5. Consult with student body on whole-school changes	✗ SLT rejected suggested changes	✓	✓
↓ ↓ ↓			
PROCESSES 2: BUILDING HEALTHY RELATIONSHIPS			
1. Teachers use RP to prevent misbehaviour	✗ Teachers felt RP threatened their authority and made them unsafe	✓	✗ Low levels of misbehaviour
2. Students learn social and emotional skills	✗ Neither curriculum nor RP normalised	✓	* Social and emotional skills already present before trial
↓ ↓ ↓			
PROCESSES 3: DE-ESCALATION IN BULLYING & AGGRESSION			
1. Perpetrators feel empathy	✓	✓	✓
2. Perpetrators accept responsibility and except punishment	✓	✓	✓
↓ ↓ ↓			
CONSEQUENCES			
1. Decreased bullying and aggression	✗ Poor intervention fidelity	✓	✗ Low rates of bullying and aggression before trial
2. Improved mental health and wellbeing	✗ Poor intervention fidelity	✓	✓

the minimum requirements. When asked why, a staff member reported:

Because the good students want to be invisible ... They don't want to be thought of as being geeks or being part of the establishment. They just want to get through school, get their qualifications and move on. (Staff, Harper's, interview)

Meadowood

Meadowood is located in outer London, with more than 50% of students eligible for free school meals and with English as an additional language. Meadowood had recently transitioned through several head teachers. The school had a large body of students disengaged from education and involved in anti-social behaviour, whom one staff member described as having parents who give their children 'a very long rein at home'. Some teachers reported their struggles to enforce boundaries:

They come here and when someone says "No, actually, you know, it's no." And that's quite a new concept for a lot of them. (Staff, Meadowood, staff FGD)

Some students and teachers reported feeling unsafe at times. Students reported needing, or at least seeming, to be aggressive to get by:

When we're in school, it's like we do have this shield where we just try and protect ourselves. And teachers don't really understand that: why we get angry or why we do this. (Student, Meadowood, RP interview, year 8)

Many staff members reported feeling safer when they could assert their authority over students, and hence, some were hesitant to use the more inclusive language of restorative practice. However, the school maintained a strong pastoral care team to support students with complex needs, and this group had been instrumental in the school volunteering for the trial. Despite its problems with some aggressive students, the external facilitator described Meadowood as 'quite nice, a happy place to be' and students described receiving 'a brilliant education' from teachers that cared not only about their academic progression but also about who they are and how they feel.

St. Anselm's

St. Anselm's was located in an affluent area outside London with lower rates of free school meals eligibility and English as an additional language. One staff member commented:

The vast majority of students that come to us are very well off And, typically, each year I would say we get about a thousand people apply for two hundred spaces....They want to come here. We're not a second-choice school. (Staff, St. Anselm's, SLT interview)

No participants reported aggression as common in St. Anselm's, and when students did behave anti-socially, the school was able to arrange tailored support. Students reported positive relationships with staff and feeling cared for and supported. Parents, the school, and the students themselves all expected academic excellence.

Increasing student commitment to the school community

This theme describes a process of increasing student commitment to the school community. It was composed of a number of sub-themes describing different parts of this process.

One sub-theme described a process involving the creation of new roles whereby staff and students on the action group came to share views and experiences. This could generate consequences of increased empathy and collegiality between staff and students, and consequently improved student relationships with staff and a desire to avoid anti-school behaviours.

Action groups functioned as a safe space within which students and staff could share their experiences and views and listen to those of others. As one staff member described:

I think that the students will certainly enjoy the fact that we're doing something like this so they can be involved in it and that they can actually have their voice heard, that they can feel safe at school, that they can feel engaged with the teachers, that they can feel they're listened to. (Staff, Harper's, staff interview)

Interviews with students suggested that they valued the conversations that occurred in action group meetings and the insights provided into other members' views and feelings. For example, when asked what was the best thing about being on the group was, one student replied:

I think mainly just having other people's, seeing other people's views and seeing how... if we had the same views or... hearing someone else's point of view and thinking, "Oh yeah." (Student, Meadowood, AG interview, year 9)

Action groups provided an opportunity for staff to enact new roles discussing the challenges that they faced.

In Meadowood's group, for example, staff spoke about how just one misbehaving student could disrupt an entire lesson. By being exposed to teachers' reports of the impact of student misbehaviour in a non-confrontational setting, students began to see their teachers as people with feelings and not merely authority figures. As one teacher put it:

I suppose in some respects the students need to know that, ultimately, you're a human being. Because I think they forget that we're a human being and we have feelings. (Staff, Meadowood, staff FGD)

Another sub-theme described how the process of students working with teachers on the action group helped build relationships between students and staff by working collectively in co-producing decisions. This was especially important as some students on the action group were selected because of their perceived disengagement with school or previous experience with bullying. The relationship between students and staff was said by students in one focus group to feel:

much more respectful ... yeah, they treat you with the same amount of respect as they would do their colleagues (Students, St. Anselm's, AG student FGD, year 10)

Students reported that the sense of collegiality developed through collaborative processes of co-production in action group meetings had the consequence of giving them a more personal view of the teachers and motivating them to work harder for the staff they had learned to respect:

If you have a bond with your teacher... you want to do well for the teacher because you feel like she's paid attention to you and gave her respect. And the way you can respect her back is by working hard. (Student, St. Anselm's, AG student FGD, year 8)

In terms of the conditions required for such processes, this appeared to be contingent on a critical mass of staff and students participating in the group and the group being well-facilitated, such as in Meadowood and St. Anselm's. In contrast, only one staff member regularly attended meetings in Harper's, therefore limiting students' opportunities to understand staff perspectives.

A further theme was student participation in the action group enabled them to enact new roles of contributing to the school community. This process could generate consequences of students experiencing a sense of agency and belonging at school but was contingent on the extent to which action groups were well attended

and effective bodies [40]. For example, students participating in St Anselm's action group were asked by the staff to help explain findings in the student needs report. Thus, students were accorded the role of an expert in contrast to the more conventional role of novice learner. Furthermore, all students on the action group possessed pertinent knowledge from their own experiences of school and from friends' experiences. This was in contrast to the normal student role, for which many less-academic students felt they lacked the knowledge to render this performable. As one staff member described:

[Analysing the needs report] has then really highlighted to us as staff in school where we need to be focusing some particular work with the students... And then we've obviously taken a lot of advice and input from the students as to how they would like things to change in the school. So ... students [are] feeling like they're actually having an input. (Staff, St Anselm's, staff interview)

Building relationships and being given a concrete task were especially important in Meadowood and Harper's where students generally did not feel that they were listened to. In these schools, where students reported feeling intimidated at the beginning of the intervention, staff described how student confidence could grow:

What I've tried to do is tried to be able to break away into smaller groups as often as we can in those meetings so that they can be able to have a one-on-one with the members of staff. And as soon as that happens ... that's when really the conversations start in there. (Staff, Meadowood, AG interview)

Participation in Meadowood's action group was seized on by some students as an opportunity to transform their experience of school:

Interviewer: Were you happy to be part of the action group?

Student: Yeah. When Miss Baker told me about it and I was... sad that I had to miss some of my favourite lessons. But I wanted to because I was getting bullied as well and I wanted to be able to stop it. (Student, Harper's, AG student FGD, year 8)

The above processes would have had only limited consequences if they were restricted to students on the action group. However, further analysis revealed several processes via which the action group appeared to have consequences for other students. Many students who sat on action groups had previously been involved in anti-

school peer groups and could share their experiences and thus initiate broader changes in student attitudes:

We have a boy ... A proper naughty boy. But he has shown such maturity in the final part of his year-11 and he's been, I think, an outstanding student in year-12. But all the kids know who he is or they know of the local family. And so, if he's on board [with the action group], that sends a really important message. And I think that is critical. (Staff, Meadowood, staff FGD)

Another sub-theme described how the action group could have broader consequences via consulting with other students and/or generating actions which affected other students. For example, in St. Anselm's, teachers reported that students on the action group developed new rules for student behaviour in a remodelled common room. Students consulted with their peers and developed new rules. Staff reported that, because a broad group of students had contributed to the process, students felt consulted and were more likely to respect the rules.

In terms of the conditions required for such processes, these appeared to be contingent on meetings being well-attended, well-led, and achieving results. In Meadowood, student participation was encouraged by the pastoral care team sitting on the action group as well as the commitment of the senior staff member leading the group. Similarly, in St. Anselm's, the staff member leading the action group encouraged students to contribute ideas and ensured that their suggestions were implemented. Students' sense of agency and contribution was less apparent in the narratives of students in Harper's. While the only staff member consistently attending the Harper's action group meetings tried to nurture her students, the lack of support from other staff and the dislocation of the action group from broader school structures made this challenging. In schools like Harper's, where action groups were less effective in achieving action, students' efforts to advocate for change could result in disappointment.

Building healthy relationships by modelling and teaching pro-social skills

The second key theme described a process of building healthy relationships by modelling and teaching pro-social skills. This theme could be understood in terms of a number of sub-themes, each describing smaller, more subtle processes.

One such sub-theme was that when teachers were empowered and able to use a restorative practice to prevent misbehaviour in classrooms, students appeared to develop increased empathy, more respectful relationships, and reduced conflict. Meadowood staff integrated

restorative practice into their normal classroom management practice to prevent and respond to consistent, low-level disruption and improve student/staff relationships. When a student was acting inappropriately in class, teachers would try to resolve the issue using restorative language. If the student did not change their behaviour, they were sent out of the room. Then, either the teacher or a member of the school's pastoral team would meet with the student to have a restorative meeting, aiming to explore why they were misbehaving followed by a brief re-introduction meeting. According to staff, this process reduced frustration and animosity between students and teachers:

Because we do have a great number of students who are children in need. We have a lot of students who obviously have issues outside of school and they bring those issues within school. We wanted to be able to implement that restorative nature in the work that every member of staff was doing with kids within the classroom. (Staff, Meadowood, AG interviews)

Like action groups, such processes engendered a sharing of perspectives and increased empathy between students and staff. One teacher commented:

There's a very easy way [to talk to challenging students] in the sense of being respectful and ... considering what those students' feelings are all the way through that because they are humans too, you know....That, you know, when members of staff get to that point where they actually look over that and they kind of go, "OK, well if I talk to you and I say to you, you know this isn't what I expect in my lesson. I am giving you a half-an-hour detention, but in that half-an-hour detention you and I can be able to speak about what the problems are." (Staff, Meadowood, AG interview)

In terms of necessary conditions, this sub-process was contingent on conditions of a critical mass of classroom teachers agreeing that student behaviour was a problem and that preventative restorative practice was a plausible response. Meadowood had high levels of classroom disruption and a pastoral care team who led the implementation of this approach and was able to support teachers. Sufficient numbers of teachers implemented the new practice informed by a recognition that previous approaches had not worked:

There were those moments where I would scream and shout at kids you know. And have a go at them and try to be able to make them see my way in a

forceful way. It has no impact. (Staff, Meadowood, AG interview)

There was less evidence of such processes in the other schools. Data from Harper's do not indicate sufficient engagement from staff to normalise the restorative practice. In St. Anselm's, the limited need for restorative practice diminished staff's need to use it consistently.

Another sub-theme described a sub-process of students learning social and emotional skills, which appeared to generate consequences of students being better able to avoid or resolve conflict. Students at Meadowood who had previously been involved in bullying and aggression described how they learned to manage emotions and social relationships constructively:

I like the fact that we get...that someone's actually teaching us how to control our emotions, so if there's an argument we know how to stop it ... Instead of kicking off at your friends, just talk with a normal tone and just apologise and see how it goes from there. (Student, Meadowood, AG student FGD, year 9)

Students learned new tactics such as asking someone to stop doing something upsetting or pausing between feeling angry and responding. This process appeared to be contingent on teachers delivering the curriculum with fidelity and students having unmet needs in this area, as was the case in Meadowood. In St Anselm's, staff reported that students already had well-developed coping skills. In Harper's, only one unit of the curriculum was taught.

De-escalation of bullying and aggression among a core group of students

The third key theme identified in our analysis describes a process of de-escalating bullying and aggression among a core group of students heavily involved in such activities. One sub-theme described a process whereby perpetrators of bullying or aggression began to feel empathy for victims through restorative practice conferences.

In Harper's, one student had encouraged his friend to take a photograph of a boy on the toilet. The two then shared the image via social media. At first, the student was unrepentant:

[During the meeting,] I was like, "I didn't really do anything" ... Then they start staring at me; I'm like, "Don't look at me, I didn't do anything."

However, seeing how devastated the photographed boy was made the perpetrator feel empathy, shame, and contrition:

And when we came in, it's just, like... at first I was laughing, because I just felt it was hilarious for him... someone to be taking the pictures of him in the toilet. But then when I just saw him there sitting down at this table and his eyes were all red from the tears... I just don't... it just came to me and just shocked me. That that could have happened to me really, it wouldn't be nice.

A second sub-theme was that such processes could generate perpetrators' recognition of their responsibility and acceptance of punishment. The same boy who encouraged the photo taken in the toilet reported that:

I normally would have been moaning [about being punished], saying "No" ... But this time I actually felt what I had done was really wrong. It just makes me realise... I mean it's ... just when I saw him sitting there in that state. (Student, Harper's, RP interview, year 8)

Restorative conversations appeared to be more consequential when they removed bullies from their peers and were forced to speak directly with the person who had been hurt by their behaviour.

[I would have wanted a one-on-one meeting] Because I think maybe because all of us were in one room – you know, reputation, you don't want to look, you know, smaller than one person and look you know weaker or more emotional than the others who were involved in the kind of like, oh I don't really care. So I think maybe having that...you know, one-on-one rather than having everybody together... (Student, Meadowood, RP interview, year 9)

Contrasting the three schools, the data suggest that for restorative practice to be widely and effectively enacted was contingent on conditions of broad support from staff as well as significant levels of bullying or aggression. The conditions for Learning Together's implementation at Harper's were characterised by students in terms of substantial problems with bullying and aggression (so that the processes triggered by restorative practice might have been effective) but low staff engagement with the intervention (so restorative practices were rarely used). At Meadowood, the reported conditions also involved high rates of aggression and misbehaviour but also included a pastoral support team committed to the intervention so that restorative conversations were routinely used. The conditions at St. Anselm's included staff committed to restorative practice but also low rates of conflict so that the processes triggered by its use were unlikely to produce significant consequences in terms of reduced bullying or aggression.

Discussion

Summary of key findings

This study aimed to develop a rich description of participant accounts of, drawing on concepts from dimensional analysis, the processes involved in a whole-school health intervention in secondary schools in South East England and how these might interact with local conditions to generate consequences, drawing on qualitative data from a process evaluation embedded within the RCT. Within a thematic content analysis, the concepts of conditions, processes, and consequences were used heuristically to interpret qualitative data.

In terms of our first research question, regarding how participants described the school context, the processes involved in their participation with the intervention, and the consequences of these, our analysis suggested several possible social processes involved in the intervention. These involved interactions between participants' agency, the school's structural context, and the possibilities introduced by the intervention's resources [32, 41]. Firstly, participants described a process by which the intervention helped the student to develop a commitment to the school community via staff and students on the action group coming to share views and experiences through collaborative co-production activities [24], relationships improving, students participating on the action group empowering them to enact new roles of contributing to the school community, and influential students on the action group bringing about changes to broader student attitudes. The second process involved building healthy relationships by modelling and teaching pro-social skills. This involved teachers using a restorative practice to prevent misbehaviour in classrooms and students learning social and emotional skills. The third process involved de-escalating bullying and aggression and enabling reintegration by giving offenders the opportunity to learn empathy and take responsibility for their actions. All three processes were presented as helping to create more inclusive and cohesive school environments. It was rare that a single process was described by itself as being sufficient for creating change. Participants described inter-related chains of processes that could be assembled and work together to reduce bullying and aggression. For example, students sharing their experiences with staff would likely be insufficient to decrease bullying, but some schools may be able to reduce bullying by improving commitment to the school community.

Action groups and restorative practice were novel firstly in terms of privileging personal perspectives, which meant that students could draw on their experiences and views, and knowledge about their friends' experiences and views, i.e. their cultural and social capital [42], to contribute. Secondly, action groups and restorative

practice sessions involved a small number of individuals. This meant that both student and staff performances felt less precarious than most classroom interactions. Thus, a new role, that of a community member, was created which both students and staff could enact, transcending their previously separate (and often oppositional) roles. In enacting this role, students and staff had more opportunities to see each other's perspectives and feel empathy for and collegiality with each other. This new role also created opportunities for new forms of pro-school identity and status, especially for students who had previously felt disengaged from school. In inhabiting this new role, students reported feeling more confident in working with adults and empowered to contribute to discussions and decisions. Such processes could affect a considerable number of students in schools, for example, when the restorative practice was used widely within classrooms to prevent conflict or when action group activities cascaded out to affect a broader group of students, such as where a broader group of students were involved in re-writing of school rules.

In terms of our second question, regarding how accounts varied between schools and what conditions relating to schools, staff, or students seemed to explain these variations, our analysis described numerous ways in which participant accounts varied between schools and what conditions relating to schools, staff, or students seemed to explain these variations. For example, the process of developing empathy between staff and students on action groups depended on consistent participation on actions groups and these being well run. Students' growing confidence depended on groups being valued and achieving results. For bullying to be effectively addressed, the school needed to identify it as an issue and gain commitment from staff to use restorative practice widely to address it. In schools where a critical mass of students regularly displayed aggressive behaviour, schools benefitted from implementing the social and emotional learning curriculum with fidelity and/or using restorative practice widely as a preventative approach, which in turn required the involvement of senior staff who supported the intervention, as the divergent experiences at Harper's and Meadowood show.

Limitations

The analysis presented here draws on data from only three schools but also, within these, from multiple interviews and focus groups conducted over 3 years. Sampling relied on the school intervention lead selecting interview participants. Although the evaluation team did ask to speak with students from a broad range of experiences and levels of school engagement, it is possible that staff chose students who might reflect more positively on the school. However, given how freely some students

expressed negative feelings towards the schools, this concern is not overwhelming.

Although there were some data on how intervention-related processes might have reduced bullying and aggression, there were fewer data on how increasing student commitment to school might have reduced other risk behaviours, such as smoking, or promoted mental or physical health. This gap was partly because of the process evaluation's inevitable focus on intervention delivery and participation and the more proximal impacts. It might also have arisen because participants had less insight into more distal impacts. Nonetheless, there is broader evidence that increasing student commitment to school can reduce risk behaviours such as smoking and drug use and promote mental and physical health [30, 43]. For example, previous qualitative research describes mechanisms whereby student commitment to the school is associated with disengagement from risk behaviours. This appears to occur via commitment to school providing a sense of identity and status, so that young people feel less need to engage in behaviours such as smoking and drug use as alternatives source of identity and status within anti-school peer groups [28]. Furthermore, there is quantitative evidence from the trial that students did experience both increased commitment to school and reductions in smoking, drinking, and drug use as well as improvements in mental well-being and psychological functioning [29].

Implications for research and policy

First, let us consider the implications of our results in terms of the intervention's original theory of change before moving on to propose realist-informed CMO configurations in line with our findings.

Our results are broadly in line with the intervention's original theory of change in terms of the intervention plausibly eroding boundaries between staff and students [27]. Our findings provide extra details about how action groups and restorative practice might erode boundaries via the enactment of shared roles and the sharing of experiences generating a sense of empathy and collegiality. Our analysis suggests that action groups could engender a new role of a community member, hence building student commitment to a school's instructional and regulatory orders, in a way that, for many less academic students, classroom interactions probably could not. Our results also confirm our assumption that such erosion would only occur in contexts where there were supportive staff cultures and enabling management structures. However, our results provide the additional insights that a key enabler of action groups eroding boundaries was sufficient numbers of staff on action groups to enable sharing and that action groups achieve broader impacts

through initiating broader student involvement or making decisions that affect students more broadly.

Our analysis also broadly supports the mechanism proposed in the intervention theory of change for eroding boundaries among students, providing the additional insight that this occurs via providing students with the social and emotional skills as well as the empathy to develop more caring relationships. Our analysis suggests that such mechanisms were contingent on context: restorative practice only contributed to such mechanisms when this was perceived as necessary and where supporting cultures and structures enabled its delivery, and the classroom curriculum only contributed where it was delivered with fidelity to students in need of it. In contrast, our results provided little evidence of the intervention eroding boundaries between academic learning and students' broader development. There was no evidence that the intervention achieved a fundamental transformation of schools' instructional order.

Moreover, our findings suggest that legitimating student expertise and enabling student agency was critical to both the action group and restorative practice [24]. Whereas normal classroom interactions and punitive disciplinary processes cast the student in an essentially passive role, with a body of educational literature portraying student agency in subversive or corrosive terms [44], the action group and restorative sessions gave students more leeway in expressing their perspectives and contributing to solving problems, giving students the role of active community member and not a merely passive learner. Such roles could be said to go some way towards reconstructing students as communitarian participants rather than merely utilitarian clients within the school.

Now, let us turn to drawing on our findings to propose realist-informed CMO configurations. Our findings are consistent with a realist approach to understanding how interventions work. Our results present interventions as merely introducing resources into a setting. Any outcomes arise not directly from these resources but as a result of their being used by local actors, which can then trigger mechanisms which may or may not interact with context to generate outcomes. Our findings present examples where processes started and stalled due to the agency of students or staff or the structural features of the school environment. Our findings suggest it is the mechanisms which are important rather than the specifics of the intervention resources. In some schools, where these particular resources may not be available, our focus on mechanisms may indicate how other resources could activate similar mechanisms in similar contexts [45]. For example, in schools where some staff have additional capacity and recognise the need for better student/staff

relationships and where holding an action group is not feasible, students and staff could share experiences, improve bonds, and create new roles via interactions through extra-curricular clubs or by having more staff on already existing student councils.

Our findings enable us to define three CMO configurations, informed by realist evaluation literature [32, 46]. The first CMO configuration relates to increasing students' commitment to the school. In schools that possess both the management capacity to run action groups successfully and a supportive pre-existing ethos of wishing to involve students in decision-making (context), students will increase their commitment to school via sharing of experiences and perspectives between staff and students in a safe space, improving staff-student relationships, creation of new roles of community members for both students and staff, and transformation of broader student attitudes (mechanisms), and will decrease bullying (outcome). These mechanisms will only remain active when students can see that they are being listened to by teachers and making a tangible difference to the school.

The second CMO configuration relates to the development of pro-social skills and their application to interpersonal conflict. In schools where a critical mass of staff identify improving students social and emotional skills is necessary and they are willing to deliver the curriculum and/or restorative practice approaches to behaviour management (context), formal teaching of such skills in classrooms or in restorative practice sessions as well as by informally modelling such skills in their everyday interactions with students will increase students' ability to use pro-social skills to address interpersonal conflict (mechanisms) and consequently decrease bullying and aggression (outcome).

The third CMO configuration related to the use of a restorative practice to de-escalate instances of conflict and reduce perpetration. When sufficient staff are willing and trained to deliver restorative practice in a way that enables students to feel safe sharing their experiences and considering the consequences of their behaviour (context), providing conflicting parties with the chance to share perspectives, learn empathy, take responsibility for their actions, and reintegrate themselves into the school community, they will defuse conflict (mechanisms) and subsequently reduce bullying and aggression (outcome).

We will explore the plausibility of these CMO configurations further, but we believe that these configurations could help inform better design of whole-school interventions to address these mechanisms. Realist approaches more generally offer a more nuanced approach to the design of interventions across settings by being clearer about the distinctions between intervention

resources and the mechanisms these can trigger when used, as well as by theorising how mechanisms are contextually contingent and outcomes are emergent [45]. In the case of whole-school interventions, this suggests the need for intervention development and evaluation to focus on initiating contextually relevant means of building student commitment, modelling healthy relationships and teaching pro-social skills, and de-escalating conflict.

Acknowledgements

The study team would like to acknowledge all participating schools and the students and staff who gave their time to take part in the intervention evaluation. We would also like to acknowledge the funders—the UK National Institute for Health Research (NIHR) Public Health Research (12/153/60) and the Educational Endowment Fund. We would also like to acknowledge Adam Walden for designing Table 2.

Disclaimer

This work was supported by the National Institute for Health Research (NIHR) in England under its Public Health Research Board (12/153/60) and the Education Endowment Foundation (no grant number). The views expressed in this publication are those of the authors and do not necessarily reflect those of the National Health Service (NHS), the NIHR, or the Department of Health for England. The study funders (National Institute for Health Research and Education Endowment Foundation) played no role in the study; collection, analysis, and interpretation of the data; writing of the report; or decision to submit the manuscript for publication.

Authors' contributions

EW designed the analysis plan, conducted the analysis, interpreted the data, drafted the manuscript, and approved it. GJM helped design the analysis, interpret the data, and approved the final version. RV designed the trial and approved the manuscript. CB designed the trial; collected, analysed, and interpreted the data; revised the drafts; and approved this version.

Funding

This work was supported by the National Institute for Health Research (NIHR) in England under its Public Health Research Board (12/153/60) and the Education Endowment Foundation (no grant number).

Availability of data and materials

The data analysed in this study are not available due to challenges in removing all identifiable information and descriptions.

Ethics approval and consent to participate

Ethical approval was obtained by the University College London ethics committee (ref5248/001). Written, informed consent was obtained from head teachers for random allocation. Students, staff, and facilitators were given information sheets and consent forms and had the opportunity to opt-out.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Faculty of Public Health and Policy, London School of Hygiene & Tropical Medicine, London, UK. ²University of Exeter, Exeter, UK. ³UCL Institute of Child Health, London WC1N 1EH, UK.

Received: 19 March 2020 Accepted: 14 August 2020

Published online: 10 September 2020

References

1. Organisation for Economic Co-operation and Development. Country Note: Programme for International Student Assessment (PISA) Results from PISA.

2018. United Kingdom. Available from: https://www.oecd.org/pisa/publications/PISA2018_CN_GBR.pdf.
2. Brooks F, Magnusson J, Klemera E, Chester K, Spencer N, Smeeton N. HBSC England national report 2014. Hatfield: BMJ Publishing Group; 2015.
 3. World Health Organisation. Global status report on violence prevention. Geneva: World Health Organisation; 2014.
 4. Davies S. Public mental health priorities: investing in the evidence. Annual report of the chief medical officer 2014. London: Department of Health; 2014.
 5. Nansel TR, Craig W, Overpeck MD, Saluja G, Ruan W. Cross-national consistency in the relationship between bullying behaviors and psychosocial adjustment. *Arch Pediatr Adolesc Med*. 2004;158(8):730–6.
 6. Lereya ST, Copeland WE, Costello EJ, Wolke D. Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries. *Lancet Psychiatry*. 2015;2(6):524–31.
 7. Langford R, Bonell CP, Jones HE, Poulou T, Murphy SM, Waters E, et al. The WHO Health Promoting School framework for improving the health and well-being of students and staff. *Cochrane Database Syst Rev*. 2011; 2014(Issue 1):Art No: CD008958.
 8. Vreeman RC, Carroll AE. A systematic review of school-based interventions to prevent bullying. *Arch Pediatr Adolesc Med*. 2007;161(1):78–88.
 9. Smith JD, Schneider BH, Smith PK, Ananiadou K. The effectiveness of whole-school antibullying programs: a synthesis of evaluation research. *Sch Psychol Rev*. 2004;33:547.
 10. Ttofi MM, Farrington DP. Effectiveness of school-based programs to reduce bullying: a systematic and meta-analytic review. *J Exp Criminol*. 2011;7(1): 27–56.
 11. Flay BR, Graumlich S, Segawa E, Burns JL, Holliday MY. Effects of 2 prevention programs on high-risk behaviors among African American youth: a randomized trial. *Arch Pediatr Adolesc Med*. 2004;158(4):377–84.
 12. Patton G, Bond L, Carlin JB, Thomas L, Butler H, Glover S, et al. Promoting social inclusion in schools: group-randomized trial of effects on student health risk behaviour and well-being. *Am J Public Health*. 2006;96(9):1582–7.
 13. Warren EBL, Opondo C, Allen E, Mathiot A, West G, Jamal F, Viner R, Bonell C. Action groups as a participative strategy for leading whole-school health promotion: results on implementation from the INCLUSIVE trial in English secondary schools. *B Educ Res J*. 2019;45(5):748–62.
 14. Bond L, Glover S, Godfrey C, Butler H, Patton GC. Building capacity for system-level change in schools: lessons from the Gatehouse Project. *Health Educ Behav*. 2001;28(3):368–83.
 15. Warren E, Bevilacqua L, Opondo C, Allen E, Mathiot A, West G, et al. Action groups as a participative strategy for leading whole-school health promotion: results on implementation from the INCLUSIVE trial in English secondary schools. *British Educ Res J*. 2019;45(5):979–1000.
 16. Buckley S, Maxwell GM. Respectful schools: restorative practices in education: a summary report. Wellington: Office of the Children's Commissioner and the Institute of Policy Studies, School of Government, Victoria University; 2007.
 17. Kokotsaki D, White C, Hopkins B. Capturing change: a review of the implementation of restorative approaches and its outcomes within a local authority in North East England. *Online Educ Res J*. 2014;5:151.
 18. Kane JG, McCluskey LG, Riddell S, Stead J, Weedon E. Restorative practices in Scottish schools. Edinburgh: Scottish Executive; 2007.
 19. Skinnis L, Du Rose N, Hough M. An evaluation of Bristol RAIS. London: ICPR King's College London; 2009.
 20. Durlak JA, Weissberg RP, Dymnicki AB. The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Dev*. 2011;82(1):405–32.
 21. Brooks F. The link between pupil health and wellbeing and attainment: a briefing for head teachers, governors and staff in education settings: November 2014; 2014.
 22. Challen A, Noden P, West A, Machin S. UK resilience programme evaluation; 2011.
 23. Daly-Smith A, Quarmby T, Archbold V, Routen AC, Morris JL, Gammon C, et al. Implementing physically active learning: future directions for research, policy, and practice. *J Sport Health Sci*. 2020;9(1):41–9.
 24. Daly-Smith A, Quarmby T, Archbold V, Corrigan N, Wilson D, Resaland GK, et al. Using a multi-stakeholder experience-based design process to co-develop the Creating Active Schools Framework. *Int J Behav Nutr Phys Act*. 2020;17(1):13.
 25. Bernstein B. *Class, codes and control: applied studies towards a sociology of language*. Routledge: Psychology Press; 2003.
 26. Bonell C, Allen E, Christie D, Elbourne D, Fletcher A, Grieve R, et al. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial. *Trials*. 2014;15:381.
 27. Markham WA, Aveyard P. A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice. *Soc Sci Med*. 2003;56(6):1209–20.
 28. Jamal F, Fletcher A, Harden A, Wells H, Thomas J, Bonell C. The school environment and student health: a systematic review and meta-ethnography of qualitative research. *BMC Public Health*. 2013;13(1):798.
 29. Bonell C, Allen E, Warren E, McGowan J, Bevilacqua L, Jamal F, et al. Initiating change in the school environment to reduce bullying and aggression: a cluster randomised controlled trial of the Learning Together (LT) intervention in English secondary schools. *Lancet*. 2018;392(10163): 2452–64.
 30. Bonell C, Jamal F, Harden A, Wells H, Parry W, Fletcher A, et al. Systematic review of the effects of schools and school environment interventions on health: evidence mapping and synthesis. *Public Health Res*. 2013;1(1):1–320.
 31. Rutter H, Savona N, Glonti K, Bibby J, Cummins S, Finegood DT, et al. The need for a complex systems model of evidence for public health. *Lancet*. 2017;390(10112):2602–4.
 32. Pawson R, Tilley N. *Realistic evaluation*. London: Sage; 1997.
 33. Van Belle S, Wong G, Westhorp G, Pearson M, Emmel N, Manzano A, et al. Can 'realist' randomized controlled trials be genuinely realist? *Trials*. 2016;17: 313–18.
 34. Marchal B, Westhorp G, Wong G, Van Belle S, Greenhalgh T, Kegels G, et al. Realist RCTs of complex interventions—an oxymoron. *Soc Sci Med*. 2013;94: 124–8.
 35. Sherman LW, Schmidt JD, Rogan DP. *Policing domestic violence: experiments and dilemmas*. New York: Free Press; 1992.
 36. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *Bmj*. 2006;333(7563):346–9.
 37. Bonell C, Allen E, Warren E, McGowan J, Bevilacqua L, Jamal F, et al. Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial. *Lancet*. 2018;392(10163):2452–64.
 38. Schatzman L. *Dimensional analysis: outline in preciform*. San Francisco: Unpublished manuscript University of California; 1980.
 39. Schatzman L. *Dimensional analysis: Notes on an alternative approach to the grounding of theory in qualitative research*. In: Maines DR, editor. *Social Organization and Social Process*. New York: Gruyter Inc; 1991.
 40. Archer M. *Structure, agency and the internal conversation*. Cambridge: Cambridge University Press; 2003.
 41. Giddens A. *The constitution of society*. Cambridge: Polity Press; 1984.
 42. Bonell C, Blakemore SJ, Fletcher A, Patton GC. Role theory of schools and adolescent health. *Lancet Child Adolesc Health*. 2019;3(10):P742–8.
 43. Paille B. *Toxic schools: high-poverty education in New York and Amsterdam*. Chicago: University of Chicago Press; 2013.
 44. Willis P. *Learning to labour: how working class kids get working class jobs*. Aldershot: Saxon House; 1977.
 45. Shearn K, Allmark P, Piercy H, Hirst J. Building realist program theory for large complex and messy interventions. *Int J Qual Methods*. 2017;16(1): 1609406917741796.
 46. Pawson R. *Evidence-based policy: a realist perspective*. Thousand Oaks CA: Sage; 2006.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Introduction to the third study:

In the previous study, I delved into three years of rich qualitative data from three case-studies to provide a portrait of each school. I also synthesised insights from participants' accounts regarding how mechanisms may have been activated to decrease bullying and what contextual features may have deactivated similar mechanisms in other schools. By looking at communities and school environments as disparate as St Anslem's, Meadowood, Harper's, I was also able to explore how students and staff experienced changes in the school environment and what contextual features seemed to remain relatively immutable during the timescale of the project. However, (like everyone's), research participants' experiences and knowledge is partial but the data they provide can be synthesised to show which resources were used, what changed, and how, but it is not well suited to show whether these patterns exist more broadly.

When trials are thoughtfully designed and contain an integrated PE, I have argued that realist questions can be answered as part of an RCT. Within the ongoing debate about realist trials, the previous publication demonstrated empirically that qualitative data from an RCT can be used to construct CMOCs, and therefore, that trials can include sufficient diversity to explore the functioning of mechanisms across diverse contexts. Despite some realist evaluators concern that trials need to be conducted in homogenous settings, the construction of CMOCs was possible because of the heterogeneity in both school environments and within individual participants' accounts. From a research approach that claims to be "whole-heartedly pluralist" [[1, pg. 85](#)] the use of RCTs to assess social change should be unsurprising.

In the dimensional analysis reported in the previous chapter, we developed three CMOCs that can be tested with data from the trial's process and outcome evaluations. The aim of the following paper is to assess whether these hypotheses are found in the trial data more broadly, and if there is evidence that the same mechanisms are activating via other resources in control schools. While I had originally planned on using moderator and mediator analyses, other study team members led on those studies [[278, 294, 304, 306](#)] (see Appendices 7-10) to publish our findings at pace and I was able to contribute to that work. I also realised in the course of the PhD that I actually had little interest in learning statistical analyses and wanted to focus on other approaches which

were better suited to the skills I wanted to develop. It was in this context that I decided that QCA would be an interesting approach within a realist trial due to its alignment with generativist articulations of causation and its ability to distil synergistic pathways to outcomes changing or remaining stable. In their work on social mobilization and homeless, Cress and Snow explain that they chose to carry out a QCA because:

“the factors associated with outcome attainment [homelessness] have typically been analysed in a correlate fashion, while the ways in which they interact with one another has remained less developed...thus the importance of the factors does not reside solely in the strength of their association with a particular outcome, but in the more complex ways they interact with each other in relation to the attainment of various movement outcomes.”[335, pg. 1070]

As described in chapter 2, the use of statistical analysis is fraught with controversy in realist circles,[3, 4] but the use of QCA may be less contentious¹⁸[333, 334, 349], although only one QCA of a realist evaluation of the Swiss environmental impact assessment evaluation was identified.[350] By including indicators of contextual features and mechanisms as conditions within a QCA model, CMOCs may be straightforward to test and may offer greater insights into the interactions between contextual features and mechanisms that give rise to or block improvements in health.

As described in Chapter 3, QCA does not use probabilistic statistics, examine associations between variables or describe net-effects. QCA is used to explore patterns of emergence to unpick how outcomes change or remain the same. The incorporation of Boolean algebra also offers a number of potential benefits. Firstly, it allows for a more nuanced examination of how different combinations of a wider spread of factors might facilitate the generation of an outcome.[337] Secondly, it illustrates the various pathways to ineffectiveness, an understudied area of investigation and shows how a condition being absent may be an equally “active ingredient” in a causal pathway as a present condition. Thirdly, it allows researchers to develop parsimonious solution by removing factors whose presence or absence has no impact. A final distinctive feature of QCA that made it an interesting method to employ was its base in qualitative knowledge of cases and its allowance of quantified qualitative data. Researchers using QCA are

¹⁸ It should be noted that many critical realists are generally opposed to the incorporation of any quantitative analysis within a realist study.[237, 275] Realist evaluators are more likely to acknowledge that numerical analyses may be acceptable.[1]

expected to have a rich knowledge of the cases involved in their analysis. By contributing to data collection and completing the previous two analyses, I had sufficiently deep case-knowledge to use QCA effectively.

It is important to be transparent about contributions to the below study.[\[10\]](#) The STATA data files had already been prepared for the primary outcome evaluation by LSHTM's CTU, led by Professor Elizabeth Allen. Those data were shared by her with me and my supervisors, Professors Chris Bonell (CB) and G.J. Melendez-Torres (GJMT). I developed the hypotheses and proposed possible proxies which were then discussed and agreed upon with CB. I then created dummy tables. When data were from the PE, I had access to the data and extracted them. When the data were from the outcome evaluation or the AG member survey, GJMT extracted the relevant data from the CTU files and sent them to me. I proposed cut-off points with GJMT, which were discussed and agreed. He then calibrated the scores in STATA and sent them back to me. I then used Tosmana[\[338\]](#), a QCA add-in function for Excel to create, analyse and interpret the truth tables. GJMT and I both reviewed truth tables for Boolean minimisation. (This process is explained and described in more detail in the publication's methods section). I led on the analysis, interpretation and writing. Both GJMT and CB advised on the analyses, commented on drafts and approved the manuscripts prior to publication.

The draft of the manuscript in the next chapter is the post-peer-review revision which has been re-submitted to the Journal of School Violence for their further consideration.

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	233090	Title	Ms
First Name(s)	Emily Ashbrook		
Surname/Family Name	Warren		
Thesis Title	Are realist randomised controlled trials possible? INCLUSIVE as a case study of an emerging methodology		
Primary Supervisor	Professor Chris Bonell		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Journal of School Violence
Please list the paper's authors in the intended authorship order:	Warren E, Melendez-Torres G.J., Bonell C
Stage of publication	Undergoing revision

SECTION D – Multi-authored work

<p>For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)</p>	<p>I led on the qualitative analysis upon which these hypotheses are derived. I proposed relevant proxies, prepared dummy data tables, and proposed anchor points. I created the truth tables, led on data analysis, interpretation, and manuscript writing. Professor Melendez-Torres worked with me to decide anchor points, and calibrated all the data in STATA, assisted with the minimization process, and interpretation. He also provided methodological support. Professor Chris Bonell helped identify appropriate proxies. All authors helped write and edit the manuscript and respond to reviewer comments.</p>
---	--

SECTION E

Student Signature	Emily Warren
Date	28/01/2022

Supervisor Signature	Chris Bonell
Date	28/01/2022

Chapter 6: Using fuzzy-set qualitative comparative analysis (fsQCA) to explore the causal pathways to reduced bullying victimization in a whole-school intervention: results from a cluster randomized controlled trial

Emily Warren, G.J. Melendez-Torres, and Chris Bonell

Learning Together is a complex whole-school intervention evaluated using a randomized controlled trial in southeast England which was found to reduce bullying and improve physical and mental health. This paper examines trial data using fuzzy-set qualitative comparative analysis to test hypotheses derived from embedded qualitative research about potential causal pathways. Analyses suggested that the intervention worked via three mechanisms: improving students' commitment to school; improving social skills; and de-escalating conflict and bullying. Evidence from the intervention and control arms show that these mechanisms may also have activated via other resources in schools that did not receive Learning Together resources. The analysis also suggests which contextual features may be important for activation of these mechanisms.

Introduction

Bullying is intentionally hurtful, repetitive, physically, verbally or socially aggressive behaviour targeting those with less power.[16] Up to one third of young people in the UK report bullying [37] and the impacts on physical and mental health can be found decades later.[119, 121] Numerous systematic reviews [16, 79, 180, 351] have examined the impact of school-based anti-bullying interventions such as social and emotional learning (SEL) curricula, restorative practices (RP), and whole-school interventions. A meta-analysis of 213 intervention studies found significantly lower rates of bullying in schools implementing SEL intervention compared to controls.[128] The evidence for RP is growing but remains

inconsistent. Two recent randomized controlled trials (RCTs) either found no significant decrease in bullying[158] or positive impacts which were not sustained.[352] Another RCT of a whole-school intervention including RP in England found significant impacts on bullying victimization and improved mental health compared to the control arm.[278] Recognizing the complexity of bullying as a social phenomenon, anti-bullying interventions have grown increasingly complex in order to provide schools with various resources which may be effective in their particular environment. However, few anti-bullying studies explore how interventions work or why effects vary with place or population. Therefore, examining “whether those interventions shift systems, and how specific conditions of interventions and setting contexts interact to lead to anticipated outcomes” [199, pg 1] is key. Understanding the intersections between resources use, mechanisms, outcomes and how those differ by context needs to be unpacked in ways that illuminate this complexity and advance both evaluations and our understanding of social phenomena.

Summary of Learning Together and its theory of change

Learning Together (LT) is a complex, whole-school intervention aiming to decrease bullying and aggression and improve mental health and wellbeing among secondary-school students in south-east England. Its theory of change is based upon the theory of human functioning and school organisation in which Markham and Aveyard propose that schools are sites in which practical reasoning and social affiliations are developed via instructional (curricular) and regulatory (social norms and behavioural expectations) orders. When students either do not commit to those expectations or are unable to meet their demands, students are less likely to develop practical reasoning and social affiliations. Schools can “reframe” school practices to better address student needs and preferences and erode the “boundaries” in classifications between students and staff, between subjects, and between students’ academic and broader development. [7] Therefore, we hypothesized that implementing LT would improve relationships between students and between students and staff, and improve schools’ responsiveness to student needs and preferences. This would in turn increase student commitment

to school and build their practical reasoning and social affiliation, culminating in reduced aggression and bullying. While each school received the same intervention resources, schools were encouraged to tailor implementation to local needs (See Figure 1 and Box 1 for more details).

[INSERT Figure 1]

[INSERT Box 1]

An RCT of LT found it effective in reducing bullying victimization but not aggression. Students in intervention schools reported higher QoL and wellbeing (as measured the Paediatric Quality of Life Inventory [291] and the Short Warwick Edinburgh Mental Wellbeing Scale [292], respectively) and lower rates of psychological difficulties (as measured by the Strengths and Difficulties Questionnaire[293]) compared to controls.[278] Pre-specified subgroup analyses based on sex, baseline experiences of bullying and aggression, and socio-economic status found the intervention was more effective for boys and those reporting bullying or aggression at baseline. There were no differences in outcomes by socio-economic status. However, these moderator analyses did not aim to examine the mechanisms through which outcomes were generated or what conditions may have made the intervention effective in some, but not all, settings.

Summary of previous qualitative research upon which this study is based

The analysis presented in this paper builds from previously published qualitative research identifying three mechanisms through which bullying may have been reduced.[284] In some schools, bullying may have decreased by building students' commitment to and increasing participation in school. Data indicate commitment was built via students and staff being able to share experiences and develop mutual empathy via working collaboratively on the action group (AG). Secondly, bullying appears to have decreased via improving students' pro-social behaviours. This was primarily described as happening via preventative RP and SEL which was reported to promote better behaviour by addressing what was and was not acceptable. Finally, qualitative data indicate that bullying was decreased via de-escalation among those students most involved in this. Bullies, victims, and

teachers described how bullies faced the consequences of their actions through responsive RP and bullies described how they often felt contrition.[284] The qualitative data were only from a small sample of case-study schools and we were interested in exploring whether these mechanisms appeared important in other schools within the trial.

Realist evaluations

Building on the theory of human functioning and school organisation and the above qualitative analysis, we developed hypotheses about these three mechanisms, i.e. that bullying could be decreased via: building student commitment to school, teaching pro-social skills, and de-escalating bullying amongst a core group of students.[284] Qualitative evidence suggested that the activation of these mechanisms was contextually contingent and that schools were using intervention resources in locally relevant ways.[8, 284] We framed our hypotheses (listed below), as context-mechanism-outcome (CMO) configurations (i.e. hypotheses about how mechanisms interact with context to generate outcomes) which are central to realist evaluation. Realist evaluations generally focus on “what works, for whom, under what conditions, and how”[1]. Within realist evaluation, interventions are understood to “work” via changes in agents’ reasoning as changed by the availability of new resources. Therefore, mechanisms are changes in cognitive or social processes, and not merely the availability or use of intervention resources. Within realist evaluation, mechanisms may interact in complex ways leading to various changes depending on important contextual features.

To assess the plausibility of those hypotheses and their transferability to schools not in receipt of the same resources, we used fuzzy set qualitative comparative analysis (fsQCA). We were interested in whether these mechanisms occurred in intervention schools via provision of intervention resources but also whether control schools were activating the same mechanisms via the use of resources already available in those schools.

Qualitative comparative analysis

QCA was developed by Charles Ragin as a tool for understanding macrosociological change.[337, 353, 354] In contrast to regression-based analysis which examines

statistical associations between multiple variables, QCA employs Boolean algebra (combinations of conditions linked by AND, NOT and OR) to examine what 'pathways' (complex combinations of the presence, absence or combination of factors) appear to result in the manifestation of a certain outcome among a set of cases. Researchers using QCA assume a configurational view of causation whereby multiple conditions combine to generate change.[336, 355] Within QCA, causes are also understood to be either "*sufficient*" or "*necessary*." A condition is *sufficient* for an outcome to occur if the outcome is always present with the condition regardless of the presence of other factors. A condition is *necessary* if the outcome cannot occur when the condition is absent.[356]

The first step in QCA is constructing a *data table* with conditions (in this analysis, contextual features, markers of hypothesized mechanisms, and outcome indicators) as columns and cases (schools) as rows. The two most common versions of QCA are crisp-set (csQCA) and fuzzy-set. In csQCA, conditions and outcomes are binarized as either as 0 or 1 showing that they are either fully cases (1) or fully not cases (0) e.g., a nation either does or does not have free elections. The word "case" in QCA refers to both the units being studied (e.g. people, schools, countries) and the degree to which they represent or contain the conditions of interest. In fsQCA, data are *calibrated* so that all data are made to fit between 0 and 1. To do this, researchers can use *indirect* or *direct calibration*. *Indirect calibration* involves researchers or experts assigning scores (also called truth values) based on their in-depth knowledge of a subject area, commonly 0 (fully not a case), 0.33 (more not a case than a case), 0.67 (more a case than not a case) or 1 (fully a case), although more nuanced truth values can be used. In *direct calibration*, researchers examine the distribution of a condition and select three anchor points: the threshold for full membership (at or above which something is fully a case), the threshold for full non-membership (at or below which something is fully not a case), and the crossover (at which it is most unclear whether or not something is a case). By transforming all scores into truth values which fit between 0 and 1, QCA generates comparable values of 'caseness.' Data may be primary or secondary, and qualitative or quantitative.

After completing the data table with the calibrated scores, the analytic focus moves from individual cases (e.g. countries or schools) to understanding the combinations of conditions that are associated with the outcome examined via a *truth table*.^[336] A truth table presents each different configuration of conditions as a row and reports how many cases are within each set. In both fuzzy- and crisp-set QCA, data in truth tables are presented as binary, with truth values <0.5 being reduced to 0 and values >0.5 becoming 1. Truth tables also report *consistency* and *coverage* scores. Consistency relates to the percentage of cases within each set that also have the same outcome.^[339] If consistency is low, there is weak or contradictory evidence that this pathway consistently leads to the outcome of interest. Coverage relates to how much of the outcome is explained by the model. Low coverage shows that the model is missing key explanations. Configurations can either be positive (all cases within the set have the outcome), negative (all cases within the set do not have the outcome), contradictions (the same combination of conditions produce different results) and remainders (possible configurations with no empirical manifestations to test them). QCA is an iterative method, with theoretically or empirically derived concepts added to the models to improve them.^[336] QCA models can become difficult to interpret with too many conditions so when more are needed to improve consistency and coverage, they are chosen judiciously, based on thick case-knowledge.^[356]

The final step before interpretation is *Boolean minimization*. In QCA, a condition's absence may be as important for causation as its presence. A condition is therefore only minimized or removed from a solution if neither its presence nor absence affects the emergence of the outcome.

QCA and trials

The use of QCA in RCTs is in its infancy.^[357] QCA was chosen as our approach because it allows for the testing of hypotheses which emerged from earlier qualitative analysis to understand how contextual features and markers of mechanisms were causally related to reductions in bullying.

Aims

This paper seeks to assess whether the causal mechanisms hypothesized through earlier qualitative research appear consonant with the pattern of contingencies

found in fsQCA. Based on the above theory and qualitative research, we first hypothesized that when students' commitment to school increased (M), when students' pro-social skills were improved (M), and/or where bullying was de-escalated amongst a core group of students (M), bullying would decrease (O)[\[284\]](#) regardless of which arm schools were allocated to within the trial.

We then further explored each of these three sub-mechanisms as CMO configurations in relation to LT resources and hypothesized that:

- 1) In schools with a pre-existing ethos of wanting to involve students in decision-making (C), improving relationships between students and staff on the AG (M), students feeling like they made a positive contribution to the school via implementation of AG activities (M), and/or feeling the AG connected them to other people in the school to make positive changes (M) will increase commitment to school (O).
- 2) In schools where students lack strong pro-social skills or where the development of pro-social skills is a staff priority (C), and/or where students feel unsafe in school (C), delivering a social and emotional skills-based curriculum (M) and/or implementing preventative RPs (M) will improve pro-social skills (O).
- 3) In schools with high bullying victimization at baseline (C), sufficient staff trained in responsive RP (M), high incidence of the use of responsive RP (M), perpetrators feeling empathy (M), and/or accepting responsibility and punishment for their actions (M) will decrease bullying (O).

Methods

Trial Methods

Details about the trial methods and the intervention are published elsewhere.[\[5, 278\]](#) In brief, LT was evaluated using a cluster RCT in 40 secondary schools. Eligible schools were mainstream state schools (not private schools, pupil referral units or schools for children with special educational needs) in south-east England with a government inspection rating higher than "inadequate". Baseline surveys involved paper-based questionnaires completed by students nearing the end of year 7 (age 11-12) with trained fieldworkers. Similar surveys were conducted at 24- and 36-months post-baseline when students were in years 9 and 10, respectively. Staff

were also surveyed at these time-points. Parents/carers were informed about the study and could withdraw their child. Written consent was gathered from students. Main trial analyses were based on intention-to-treat analyses and focused on primary and secondary outcomes at 36-months.

Data sources

Data used in the overarching model were drawn from baseline and endline student surveys and included measures for mechanisms relating to aggression, school climate, mental health and wellbeing, and bullying victimization as the primary outcome (see Table 1). Specific, validated scales used in the surveys included: the Strengths and Difficulties Questionnaire which includes subscales on behavioural, emotion and peer problems as well as pro-social strengths;[\[293\]](#) the Beyond Blue School Climate Questionnaire which contains subscales on relationships with teachers, student participation in school decision making, commitment to learning, and sense of belonging in the school community;[\[358\]](#) the Edinburgh Study of Youth Transitions and Crime (ESYTC) scale which measures aggression towards staff and fellow students;[\[290\]](#) and the Gatehouse Bullying Scale (GBS) which measures bullying victimization in the last three months.[\[289\]](#)

Data sources used in the sub-mechanisms were drawn from the above sources as well as surveys of AG members, staff surveys, interviews with students involved in restorative conferences, process evaluation records on intervention fidelity, and staff surveys reports on whether and how they used RP. Measures of context were collected in the first year of the trial while outcomes used either endline scores or changes between baseline and endline. Mechanisms were represented by change over time or data collected at 24-months post-baseline (See Tables 2-4).

QCA

To begin the analysis, an overarching model was constructed to assess whether the three hypothesized sub-mechanisms, individually or in combination, led to decreased bullying across contexts. While our overall aim is to understand how the use of intervention resources activated mechanisms to improve outcomes, we also recognize that control schools were making efforts to reduce bullying using resources available to them. Therefore, by building our overarching hypothesis to examine broader social mechanisms (and not merely mechanisms directly arising

from use of LT resources) we are able to assess whether the mechanisms we identified are plausible in reducing bullying, regardless of which specific resources may have enabled their activation. This aligned with the study's realist orientation to understanding the plausibility of mechanisms and how these vary by context.

For the models examining what sub-mechanisms were important in different contexts, we examined mechanisms via markers of increased participation in school decision-making, improved social skills, and decreased aggression, which we theorized would contribute to reductions in bullying. We also introduced measures of contextual conditions in these models. The analysis of these sub-mechanisms focused only on intervention schools because we wanted to understand how LT resources were used by agents within various school environments to reduce bullying.

Building data tables

A recent systematic review has highlighted the need for those establishing anchor points and interpreting the data to have 'thick' knowledge of the cases. [199] For this current study, the lead author was responsible for collecting much of the qualitative data, had worked directly with all the schools in the trial, and all authors contributed to the earlier qualitative analyses which informed our hypotheses. Two study authors (*EW and GJMT*) examined data to decide anchor points. For example, we started by examining the school average change in bullying victimization using the GBS, which varied from -62 to +7%. After examining any natural gaps in the data which may indicate qualitatively different levels of casesness and the conditions' distribution, we discussed at what level of bullying reduction might be of public health significance. We decided that schools with greater than 50% reductions were fully cases, schools with less than 15% decrease in bullying were not fully cases, and schools with 30% reductions were the most ambiguous. It is important to note that even though a 15% reduction in bullying is notable, schools achieving less than that were the least successful in our sample, and cut-offs must be established based on the included data so that analysis can continue. The schools' GBS scores were then directly calibrated in STATA, giving every school a truth value between 0 and 1. This same process of examining scores, establishing cut-offs, and calibrating data was repeated with all

other conditions in our overarching mechanisms model, and later repeated for each sub-mechanism.

Constructing truth tables

Generated using the Tosmana [338] add-in for Microsoft Excel, we assessed each model's consistency and coverage in their respective truth tables. When coverage was too low, new concepts, suggested by our intervention theory of change and qualitative research, were added. To avoid data-dredging, we stopped adding conditions when there were no further measures that aligned with the hypotheses informed by the qualitative findings. For example, in the first iteration of the overarching mechanism, consistency was high at 90% and coverage was moderate at 55%. Therefore, we added indicators for learning conflict resolution and decreasing conduct problems which are both important for improving social skills.[293]

Our first iteration of sub-mechanism 3 lacked sufficient explanatory power. Therefore, we added a measure of the success of RP into our model, reducing this to only include the 14 schools for which interview data provided a marker of this. In these schools, interviews were conducted with students who had been involved in a restorative conference either as a bully or victim. To quantify interview data, we created spreadsheets identifying which school the data came from, key quotes explaining the situation, and any data that expressed feeling empathy (or not) and accepting responsibility and, when applicable, feeling that punishment was fair (or not). These quotes were then given a score of 0 if they did not express any change in attitude, .33 if they expressed very limited change in attitude, .67 if they recognized a change in their attitudes but it was not complete, and 1 if they described the intervention as having a meaningful change in their views and actions.[359] When multiple accounts were taken from one school, the scores were averaged, and then directly calibrated in STATA. These markers of mechanisms were not included in the original models because non-parsimonious models will include more "remainders" or possible configurations without any empirical manifestations, and can become impossible to interpret.[336, 360, 361]

When new conditions were added to the data tables, they were subjected to the same aforementioned calibration process, and truth tables were re-run to assess

the impact of their inclusion on coverage. Tables 1-4 show which variables were included in the original models, and which were added later.

[INSERT TABLES 1-4]

Boolean minimization

We examined all configurations and identified all instances which achieved the same outcome and were the same except for the presence of one variable. Where these were identified, we report the simplified solution. When reading QCA solutions, present conditions are written in capital letters, absent conditions are written in lower-case letters, and * is read as “and.”

Results

Overarching mechanism

Our first model explored our hypothesis that schools can decrease bullying by improving students’ commitment to school, improving social skills, and/or de-escalating bullying, regardless of context. We identified 13 pathways (with data from 21 schools) that did not decrease bullying and 15 pathways (with data from 19 schools) that did. (See Online Appendix Table 1.) These 15 effective causal pathways were minimized to nine solutions. Consistency across solutions was very high (97.43%) meaning that all of the schools that followed one of these effective pathways reduced bullying. Coverage was moderate at 62% meaning that 62% of the decrease in bullying could be explained by these combinations of factors (see Table 5).

[INSERT TABLE 5]

School 31 was the most similar to our hypothesis and had all conditions except using RP to resolve conflict, and experienced one of the greatest decreases in bullying (truth value= 0.9766525). The pathway with the greatest explanatory power (role*rpsolving*CP*aggress) suggests that in schools that did not improve participation, implement RP or decrease aggression, but that did decrease conduct problems, bullying was meaningfully lessened. This pathway explains 20% of the bullying decrease in the trial. A similar configuration was also effective (belong*role*prosocial*rpsolving*CP; coverage = 0.14792643). Decreasing

conduct problems appears to be the most important mechanism for decreasing bullying, activating in 12 of 15 effective solutions. Other configurations required the activation of multiple mechanisms. For example, in schools that did not improve students' sense of belonging or participation in school, and did not decrease aggression but improved pro-sociality and taught students to resolve conflict (belong*role*PROSOCIAL*RPSOLVING*aggression) explained 14% of the model's effectiveness.

Sub-mechanisms

We then examined each sub-mechanism, including contextual features.

Sub-mechanism 1: Improved commitment

Our first sub-mechanism relates to pathways through which schools could improve students' commitment to school. Our analysis found evidence for eight pathways that did not improve participation and two that did. However, the contextual feature (baseline ethos of involving students in decision making) was not necessary in either effective pathway. Excluding the contextual feature, schools 22 and 27 met the conditions for our hypothesized mechanisms and both increased students' commitment to school. After Boolean minimization, the reduced causal pathway can be expressed as ACTIONS*ATTITUDE CHANGE→PARTICIPATION meaning that students feeling they made a positive contribution through the implementation of AG activities and students reporting that the AG helped them connect with other students to change the school led to increased student participation in decision-making. This effect was felt in four schools. Consistency was good at 84.10% but coverage was low at 32.40%. This means that the majority of the ways through which schools improved commitment were achieved outside of these conditions. (See Table 5 and Online Appendix Table 2)

Sub-mechanism 2: Improved pro-social skills

No school had all of our hypotheses' conditions (weak prosocial skills, students feeling unsafe at school, the school delivering the social emotional learning curriculum, and using preventative RPs). Two sets of configurations were similar to our model, each containing one of the two contextual features and both

mechanisms activated, but only one configuration (FEEL UNSAFE*CURRICULUM*PREVENTATIVE RP) led to improvements in pro-social behaviour (see Online Appendix Table 3). Six pathways affecting 13 schools were identified that did not lead to improvements in pro-social skills. Seven schools improved pro-social skills via five other pathways.

The data indicate that students feeling unsafe at school was a more important contextual feature than students lacking pro-social strengths or having staff value their development. Feeling unsafe was an important condition in four of five configurations while lacking pro-social skills was active in only two. Interestingly, there was one configuration in which none of our hypothesized mechanisms activated but pro-social skills were still improved, indicating that other mechanisms, disconnected from the trial, improved pro-social skills.

The delivery of the SEL curricula was only present in one effective pathway but was present in half of the ineffective pathways, indicating that the curriculum had a negligible or even a detrimental impact on improving social skills. Consistency was acceptable at 76.34% and coverage was moderate at 54.27%. The pathway with the greatest explanatory power (38%) was explained by students having underdeveloped pro-social skills or their development was seen as a priority by staff, not delivering the curriculum, and using preventative RP (see Table 5). This aligns closely with earlier qualitative research that indicated that teachers often did not like the curriculum but felt RP was useful in improving student behaviour.[\[8, 284\]](#)

Sub-mechanisms 3: De-escalate conflict amongst a core group of students

Of the 14 schools with data from bullying incidents that were responded to with restorative conferences, we identified seven pathways (with data from eight schools) that did not decrease bullying and five (with data from six schools) that did (see Online Appendix Table 4). School 3 met all of our hypothesized conditions and mechanisms and bullying decreased. High baseline victimization was present in five of the ineffective combinations indicating that it may not be important for the activation of the investigated mechanisms. Consistency was high at 90.24% and coverage was moderate at 59.7%. The pathway that explained the greatest decrease in bullying (33% coverage) was not having high bullying at baseline,

having sufficient staff trained in RP, but not needing perpetrators of bullying to feel contrition. Another effective configuration was not having high victimization at baseline, training staff in RP, and students not learning empathy or expressing contrition (23% coverage) (see Table 5 and Online Appendices Table 4). In both of the above configurations, training staff in RP may have caused a decline in bullying by modifying and correcting many small instances of poor behaviour which can create a climate where bullying is tolerated.

Discussion

Summary of key findings

This analysis showed that LT provided resources which could be used to reduce bullying and that QCA is an effective approach for unpicking how and under what conditions outcomes changed. Our overarching model indicates that of the three investigated sub-mechanisms, the most consistently effective appears to be strengthening students' social skills, as at least one indicator of improved pro-sociality was indicated in all of the effective solutions, the most common being decreased conduct problems. Various intervention resources, including training in preventative and responsive RPs may have contributed to the activation of this mechanism. Indicators of improving commitment were present in six pathways and de-escalated bullying was found in four. When exploring the sub-mechanisms, de-escalating bullying had the highest coverage in the combined solution, indicating that training in the use of RP and teaching empathy and contrition can all contribute to decreasing bullying. The presence of hypothesized contextual features was less important than the activation of hypothesized mechanisms in the generation of improved outcomes. Evidence from the overarching hypothesis shows that control schools were able to activate the same mechanisms using other resources.

While no school met all of the conditions of our overarching hypothesis, the school that most nearly did so also experienced one of the largest decreases in bullying victimization, suggesting that our overarching hypothesis is a highly plausible pathway through which bullying can be decreased. Two schools improved commitment via the predicted mechanisms but did not have the contextual feature that the qualitative data indicated would be important for its activation. To

increase commitment, the two most important mechanisms appeared to be creating new roles for students and participation in the AG changing broader attitudes towards school. Evidence also indicates that preventative RPs help develop student pro-social skills. Finally, the evidence fully supported our hypothesis about de-escalating bullying.

QCA enabled us to look beyond monocausal explanations of causality to focus on generative explanations in terms of complex combinations of contextual features and mechanisms, suggesting there were multiple pathways to the same outcome. [349] Our research also showed that while one part of a sub-mechanism may be sufficient to decrease bullying in some schools, multiple mechanisms have to be activated together to disrupt the mechanisms which generate bullying in other schools.[362] Within our three sub-mechanisms, we found evidence of impact for 15 of 20 intervention schools, with only two schools activating more than one sub-mechanism. In practice this may indicate that schools can select which activities to focus on depending on their needs and abilities, and that multiple activities, acting independently or synergistically, may decrease bullying. In some schools this may have meant focusing on students who were regularly aggressive while in others it meant creating opportunities for students and staff to build bonds outside of hierarchical classroom settings. Within this study's realist lens, this finding was anticipated.

We were able to identify consistent explanations across both arms of the trial, indicating that our theory of change was helpful in explaining outcomes. Coverage was lower, advancing our belief that schools were likely undertaking other activities not related to our intervention which decreased bullying.

Weaknesses and strengths of this study

Our study has important limitations. Firstly, our over-arching model did not account for the contextual features earlier qualitative analysis indicated would be important for the activation of mechanisms. Within an already-large model, additional variables would have made interpretation difficult. Secondly, some of the mechanisms we identified as important were not predicted before the trial began so we lacked measures and therefore used imperfect proxies in these analyses. In two instances, we had to quantify interview data to develop post-hoc

markers of the mechanisms that the qualitative data indicated was crucial for change, which may mean that the measures are of limited validity as the interview guide did not include specific questions about empathy or contrition. Thirdly, our truth tables show that we identified contradictory configurations whereby the same combination of conditions led to different outcomes in different schools. As we were unable to expand the models further without data dredging, they remain unresolved. Fourthly, our process evaluation asked school staff about whether addressing bullying was a priority but we lacked sufficient detail to report on what other activities may have been ongoing to reduce bullying at the same time as the intervention. Finally, it is possible that some pathways that were identified as effective were coincidental rather than causal.

Where possible, we sought to bolster our research against QCA's well-known shortcomings. For example, accounting for the passage of time is difficult in QCA where data are generally cross-sectional. To maximize the strengths of the longitudinal data, all of our contextual features in our CMOs were taken from data in the first year of the study, and our outcomes related to either the percent change over three years or prevalence at endline.

One key strength of this study is the use of data from both trial arms. Other QCA studies have been nested within larger RCTs but have only focused on pathways to change within the intervention arm.^[363] By exploring the overarching mechanism also drawing on data from control schools, we are able to examine the social processes through which improvements might be made without focusing solely on intervention resources, enabling us to assess whether or not our hypotheses were generalizable to schools in the control arm. Another strength of this analysis is that plausible pathways outside of our hypotheses were highlighted. Finally, this study is part of a theoretically informed evaluation and builds on an earlier qualitative study which drew on 66 interviews in order to inform the CMOs tested here.

Conclusion

While our sample was not always large enough to have empirical manifestations of our exact hypotheses, our CMO configurations were generally supported. Evidence from this study suggests that student participation in decision-making may be an avenue through which bullying can be decreased. Preventative RPs may be

sufficient to improve pro-social skills among students, and training staff in responsive RP appears to de-escalate conflict amongst a core group of aggressive students. The SEL curricula appears to be the least effective resources provided through the LT intervention. Even when the hypothesized contextual features were not present, mechanisms were often still able to activate, indicating that a wide range of schools could benefit from the implementation of either preventative and/or responsive RP depending on the local manifestations of bullying. Our analysis, informed by realist evaluation, suggests that when given resources, agents will deploy them in locally relevant ways. While some schools needed to activate numerous mechanisms to improve outcomes, in others, changes could be achieved more easily. Our analysis also showed that using different resources, control schools also achieved reductions in victimization- although slighter- via the same mechanisms, indicating that our hypotheses are plausible and potentially generalizable. QCA is a useful approach within trials and it can be used to explore phenomena in both trial arms.

Box 1:

LT was delivered over three years. Resources included an intervention manual; annual reports based on students' needs as reported in an annual survey; a trained facilitator to guide intervention delivery in the first two years; and a yearly social and emotional learning curriculum. In the first year, a half-day training in preventative restorative practice (RP) was offered to all staff, and schools sent 5-10 staff members to a three-day training on responsive RP. These resources were meant to enable the following processes: form action groups (AGs) with at least six students and six staff to meet twice per term to review the needs reports, implement the AG's decisions on ways to improve the school, implement the curriculum, and review and re-write rules and policies to be more restorative. Schools were asked to implement preventative RPs to improve behaviour and responsive RPs to address bullying and aggression. These resources were intended to work synergistically to improve the school environment.

Table 1: Overall mechanism, indicators, variables, and whether they were included in the original model

	Sub-mechanism 1		Sub-mechanism 2			Sub-mechanism 3	Outcome
	Improved commitment		Improved social skills			De-escalate bullying	Decreased bullying victimization
Indicator	Increased overall student belonging	Creating a role for to participate in school	Improved pro-social skills	Learning conflict resolution	Decreased conduct problem	Reduced perpetration	Decreased victimization
Abbreviation	belong	role	prosocial	rp solving	cp	aggression	decreased bullying
Data source (difference between 1-36 months)	BBSCQ belonging subscale	BBSQC student active participation at school subscale	Selected SDQ prosocial items (1,4,9,17,20)	Students who report teachers help resolve conflict	SDQ conduct problems subscale	ESYTC measure of bullying perpetration	GBS (School % difference 0-36 months)
Included in original model or added to improve coverage	Original	Original	Original	Added to improve coverage	Added to improve coverage	Original	Original

Table 2 for improved commitment: Indicators, data sources, and whether they were included in the original model

Context (C), Mechanism (M), or Outcome (O)	C	M	M	M	O
Indicator	Pre-existing ethos of wishing to	Good relationships between staff and students on AGM	Students feel they make a positive contribution to the school via	AG participants initiate change in	AGM increases participation of other students in decisions

	involve students in decision-making		implementation of AG activities	students' attitudes to school	
Abbreviation	decision-making	relationships	actions	attitude change	participation
Relevant intervention resource		Facilitator, annual student needs survey, preventative RP training	Facilitator, NAR	Facilitator, NAR	
Data sources	BBSCQ participation subscale at baseline	AGM survey (end of year 1) Score with a point for agreeing with any of the following: "I got positive responses when I expressed my own attitudes and ideas on the Action Group"; "I found the Action Group to be exciting and energizing"; "This Action Group taught me how to work well together with others"; and "This Action Group helped me connect with other people in my school to help others".	AGM surveys (end of year 1) "Do you think the AG made sure that these actions were implemented?" Percentage of students who answered "Yes" vs "No" and "Not sure"	AGM survey (year 1) "This Action Group helped me connect with other people in my school to help others" Percentage of students who answered "Yes" vs "No" and "Not sure"	BBSCQ (school % difference 0-36 months)
Included in original model or added to improve coverage	Original	Original	Original	Added to improve coverage	Original

Table 3 for improved social skills: Indicators, data sources, and whether they were included in the original model

Context (C), Mechanism (M), or Outcome (O)	C	C	M	M	O
---	---	---	---	---	---

Indicator	Students lack strong pro-social skills, or their development is seen as a priority by staff	Students feel unsafe in school	Deliver SEL skills curriculum with fidelity	Preventative RPs have been used	Improved pro-social skills
Abbreviation	weak pro-social	feel unsafe	curriculum	preventative rp	improve prosocial
Relevant intervention resource			Curriculum	Preventative and responsive RP training	
Data source	School average baseline SDQ	Student survey question 55 at baseline: "Do you feel safe in school?" Percentage of students who responded "Never" vs "some of the time", "most of the time", and "all of the time"	Delivered 5+ hours or units of SEL curriculum in Y1 and Y2	Staff survey question Q32 at endline: "Teachers and students at this school get together to build better relationships" and question 33 "Teachers and students at this school get together to discuss their views and feelings" (Answers: "Often" vs "Sometimes" or "Never")	SDQ-pro-social subscale (school % difference 0-36 months)
Included in original model or added to improve coverage	Original	Added to improve coverage	Original	Original	Original

Table 4 for De-escalated bullying: Indicators, data sources, and whether they were included in the original model

Context (C), Mechanism (M), or Outcome (O)	C	M	M	M	M	O
---	---	---	---	---	---	---

Indicator	High baseline bullying victimization	Sufficient staff trained in RP	Implement responsive RP	Perpetrators feel empathy	Perpetrators accept responsibility and accept punishment	Decreased bullying
Abbreviation	bullying	rp training	responsive rp	empathy	contrition	decreased bullying
Relevant intervention resource		Responsive RP training	Responsive RP training	Responsive RP training	Responsive RP training	
Variable	GBS at baseline. Threshold: School score at baseline > median across all schools	At least 6 members of staff participated in 3-day training	From the endline staff survey: "If there is trouble at this school, staff respond by": and anyone who answers "Talking to those involved to help them get on better"	Student interviews	Student interviews	GBS (School % difference 0-36 months)
Included in original model or added to improve coverage	Original	Original	Original	Added to improve coverage	Added to improve coverage	Original

Table 5: Consistency and coverage scores for effective solutions

Mechanism	Consistency	Coverage
Overall mechanism: Effective solutions to reduce bullying victimization		
Combined solutions	0.97437423	0.61665392
belong*ROLE*prosocial*RPSOLVING*cp*AGGRESS	0.99540198	0.02784096

belong*ROLE*PROSOCIAL*RPSOLVING*CP*AGGRESS	1	0.05866053
BELONG*ROLE*PROSOCIAL*rpsolving*cp*aggress	0.99647295	0.0465431
belong*role*prosocial*rpsolving*CP	0.95749176	0.14792643
belong*role*PROSOCIAL*RPSOLVING*aggress	0.98276007	0.14313276
belong*ROLE*prosocial*CP*aggress	0.97100782	0.12016959
BELONG*PROSOCIAL*rpsolving*CP*AGGRESS	1	0.11216037
BELONG*ROLE*rpsolving*CP*AGGRESS	1	0.1324797
role*rpsolving*CP*aggress	0.95651352	0.20270701
Sub-mechanism 1: Effective solutions for improving commitment		
decision-making*ACTIONS*ATTITUDE CHANGE	0.84102321	0.32393599
Sub-mechanism 2: Effective solutions for improving pro-social skills		
Combined solutions	0.76434785	0.54527509
weak pro-social*curriculum*PREVENTATIVE RP	0.83939826	0.3778989
weak pro-social*FEEL UNSAFE*PREVENTATIVE RP	0.75935143	0.30734947
WEAK PRO-SOCIAL*FEEL UNSAFE*curriculum	0.86973155	0.25098297
Sub-mechanism 3: Effective solutions for de-escalating conflict		
Combined solutions	0.90241832	0.59704226
bullying*rp training*responsive rp*EMPATHY*CONTRITION	0.88330477	0.10710326
BULLYING*RP TRAINING*IMPLEMENT RESPONSIVE RP*EMPATHY*CONTRITION	0.85339141	0.10188263
bullying*RP TRAINING*empathy*contrition	0.90876937	0.3278009
bullying*RP TRAINING*responsive rp* empathy	0.93179297	0.22316746

Capital letters indicate the presence of a condition; lowercase letters indicate the absence of a condition. * = and.

ONLINE APPENDICES

Online Appendix Table 1: Truth table for overarching mechanisms

Improved commitment		Improved social skills			De-escalation	Outcome		
Belong	Role	Pro-social	RP solving	CP	Aggression	Decreased bullying	Consistency	Frequency (school identifiers)
0	0	0	0	0	0	0	0.4017148	3 (7, 6, 15)
0	0	1	0	0	0	0	0.44614342	2 (26, 33)
1	0	0	0	0	0	0	0.50450158	1 (40)
0	1	0	1	0	0	0	0.67300642	2 (39, 22)
0	1	0	0	0	0	0	0.69604999	1 (17)
0	0	0	1	0	0	0	0.71058667	2 (18, 4)*
0	0	0	1	1	0	0	0.75201011	1 (29)
1	1	0	0	0	0	0	0.79608303	1 (24)
1	0	0	1	0	0	0	0.79884046	1 (23)
1	1	1	1	0	1	0	0.82398456	3 (38, 12, 19)*
1	1	0	1	1	0	0	0.8564707	2 (30, 9)
1	0	1	1	1	0	0	0.8625111	1 (20)
1	1	1	1	0	0	0	0.88112772	1 (35)
0	0	1	0	1	0	1	0.91093093	1 (13)
0	0	0	0	1	0	1	0.95655191	2 (36, 21)
0	0	1	1	0	0	1	0.95869929	1 (2)

0	1	0	1	1	0	1	0.98954034	1 (14)
0	1	0	1	0	1	1	0.99540198	1 (16)
0	1	0	0	1	0	1	0.99626642	2 (8, 37)
1	1	1	0	0	0	1	0.99647295	1 (27)
1	0	1	0	1	0	1	0.99910313	1 (10)
1	0	0	0	1	0	1	0.99912763	1 (34)
0	0	0	0	1	1	1	1	1 (3)
0	0	1	1	1	0	1	1	1(5, 11)
0	1	1 (32)						
1	0	1	0	1	1	1	1	1 (28)
1	1	0	0	1	1	1	1	2 (25, 1)
1	1	1	0	1	1	1	1	1 (31)

Effective solutions in bold; contradictions indicated with *

Online Appendix Table 2: Truth Table for sub-mechanism 1 (Improving commitment)

Decision-making	Relationships	Actions	Attitude change	Participation	Consistency	Frequency (school identifiers)
0	0	0	0	0	0.77994579	2 (1, 3)
0	0	0	1	0	0.75885803	1 (24)
1	0	0	0	0	0.51764214	2 (11, 23)
1	0	0	1	0	0.716672	2 (33, 19)*
1	0	1	1	0	0.48022598	1 (39)
1	1	0	1	0	0.55138689	2 (13, 30)
1	1	1	0	0	0.59032303	1 (9)
1	1	1	1	0	0.46901244	5 (2, 18, 26, 28, 38)*
0	0	1	1	1	0.82203442	2 (10, 25)
0	1	1	1	1	0.82687396	2 (27, 22)

Effective solutions in **bold**; contradictions indicated with *

Online Appendix Table 3: Truth Table for sub-mechanisms 2 (Improving pro-social skills)

Weak pro-social	Feel unsafe	curriculum	Preventative RP	Improved pro-social	Consistency	Frequency (school identifiers)
0	0	1	0	0	0.16182378	2 (24, 9)
0	1	0	0	0	0.61754107	4 (23, 25, 3, 1)*
1	0	0	0	0	0.48359427	1 (10)
1	0	0	1	0	0.61548901	3 (13, 11, 33)*
1	0	1	0	0	0.46858984	1 (18)
1	0	1	1	0	0.6870141	2 (27, 39)*
0	0	0	1	1	0.85340101	2 (19, 30)*
0	1	0	1	1	0.87240565	2 (22, 2)*
0	1	1	1	1	0.92279029	1 (38)
1	1	0	0	1	0.83872306	1 (28)
1	1	0	1	1	0.98480994	1 (26)

Effective solutions in **bold**; contradictions indicated with *

Online Appendix Table 4: Truth Table for sub-mechanisms 3 (De-escalation of conflict)

Bullying	RP training	Responsive RP	Empathy	Contrition	Decreased bullying	Consistency	Frequency (school identifiers)
0	0	0	1	0	0	0.44274211	1 (19)
1	0	1	1	0	0	0.48113209	1 (26)
1	0	0	1	0	0	0.53866839	2 (18, 25)
1	0	0	0	0	0	0.55126637	1 (33)
1	0	1	0	0	0	0.74444433	1 (30)
1	1	1	1	0	0	0.78494948	1 (10)
0	0	1	0	0	0	0.78650504	1 (2)
1	1	1	1	1	1	0.85339141	1 (3)
0	1	0	0	0	1	0.87700456	1 (1)
0	0	0	1	1	1	0.88330477	1 (28)
0	1	1	0	0	1	0.89185798	2 (13, 27)
0	1	0	0	1	1	1	1 (24)

Effective solutions in **bold**

DISCUSSION

Introduction to the Discussion chapter:

As described in Chapter 2, realist evaluators opposed to RCTs argue that realist trials are “oxymoronic”:[3] randomisation and control stifle context; trials depend upon constant conjunctions; they cannot account for complexity; they ignore agency; and they are incompatible with a generativist ontology. We hope that the QCA allayed some of these concerns. QCA enabled us to examine evidence that the presence or absence of important contextual features might have on the activation of various mechanisms, either working individually or in conjunction. This aligns with generativist understanding of causation. Detractors of realist trials suggest that randomisation and control stifles one’s ability to understand and explore the complexity of “real-world” investigations.[3] Van Belle et al assert that:

“Given the need for randomisation and control in an RCT, only relatively few and simple CMO configurations can be tested at a time. At best, then, the RCT may help us in assessing the relative contribution of mechanisms to outcome patterns if the causal configuration is uniform but not when it is likely that different mechanisms will generate different outcomes in different circumstances, as is the rule rather than the exception in any health intervention.”[4, pg. 4-5]

The previous chapter suggests that despite their concerns, neither the use of randomisation nor control groups limited our ability to develop a generativist understanding of causation or limited the testing to few or simple CMOs. Helpfully, it also demonstrated that RCTs can be designed to enable the exploration of multiple, complex configurations to assess the impact of context and how and under what circumstances mechanisms activate. It also shows that the causal pattern does not need to be uniform and that multiple pathways, both to improvements and to things remaining unchanged, can be identified.

This study also helps us address the concern of opponents of realist trials relating to complexity, agency and adaptations. By using qualitative data to develop our hypotheses about whether or not certain mechanisms were activated, we were able to

account for agency directly from participants who explained why their behaviour or attitudes did or did not change as a result of interacting with intervention resources. Fidelity data, taken from a mix of qualitative and quantitative sources, also enabled us to understand local adaptations, especially in relation to the delivery of RPs and the SEL curriculum. By testing our overarching mechanism with data from both trial arms we also made a clear declaration that we do not need to believe that LT resources alone were causing a decrease in bullying: part of our aim was to assess the social mechanisms, regardless of which resources may have led to the activation of those mechanisms in order to improve our theory. We did not operationalise mechanisms merely as fidelity or the availability or use of intervention resources. The mechanisms we tested related primarily to changes in people's reasoning or psychological processes. Our findings do indicate, however, that training staff in RPs, developing their skills in leading student/staff groups, and providing them with ready-made curricular resources enabled their empowerment to improve their school environment and decrease bullying in ways that they believe are likely to be effective in their school. These changes were more likely to occur in settings that received intervention resources than those that did not (as the overall trial findings demonstrated). The smaller mechanisms delved into the interplay between programme theory, relating to intervention specific resources and processes, and mid-range theory, exploring the social phenomena more broadly. The study also examined the role that the availability of intervention resources played in activating hypothesised mechanisms. This will be helpful when synthesising all the work on the trial together to refine the study's theory of change and improve the mid-range theory.

As discussed earlier, in 2004, Oakley et al argued that PEs need to move beyond evaluating acceptability and implementation processes to evaluating the impact of these processes on outcomes.[\[217\]](#) So far in this thesis, we have evaluated fidelity and the acceptability of LT and the processes that staff and students were expected to enact as part of their participation in the trial. Using qualitative data from case-study schools, we have then completed an in-depth analysis of participant accounts of their school and community environments, and the processes they enacted with LT resources. We have drawn on this to develop hypotheses about: the mechanisms through which change may have occurred; what contextual factors or agent's decisions may have enabled or inhibited change; and how context appeared to affect their emergence. Finally, we

tested those hypotheses and found that our CMOCs were largely supported by the data and were highly plausible explanations for the change that was identified in the trial's primary outcome evaluation.[\[278\]](#)

Having already considered the question about the philosophical coherence of realist RCTs in Chapter 2 and in a review of social science and philosophy literature on positivism[\[259\]](#) (see Appendix 3), I now return to the original debate to primarily focus on the practical challenges that realist evaluators predicted would make a realist trial unfeasible and incoherent, and reflect on conducting the first realist trial. Part of this reflection includes recommendations to researchers interested in realist trials about what they can do to avoid some of the challenges we faced.

In the below publication, we also refer to a number of studies, the full text versions of which can be found in Appendices 3 and 7-11.

Are realist randomised controlled trials possible? A reflection on the INCLUSIVE evaluation of a whole-school, bullying-prevention intervention



London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	233090	Title	Ms
First Name(s)	Emily Ashbrook		
Surname/Family Name	Warren		
Thesis Title	Are realist randomised controlled trials possible? INCLUSIVE as a case study of an emerging methodology		
Primary Supervisor	Professor Chris Bonell		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Trials		
When was the work published?	Jan 2022		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I drafted the commentary. Professor GJ Melendez-Torres edited the manuscript and provided points for further consideration. Professor Chris Bonell edited the manuscript and helped shape the arguments.
--	--

SECTION E

Student Signature	Emily Warren
Date	28/01/2022

Supervisor Signature	Chris Bonell
Date	28/01/2022

COMMENTARY

Open Access

Are realist randomised controlled trials possible? A reflection on the INCLUSIVE evaluation of a whole-school, bullying-prevention intervention



Emily A. Warren^{1*}, G. J. Melendez-Torres² and Chris Bonell¹

Abstract

We previously proposed that realist randomised controlled trials could be used to evaluate how, for whom and under what conditions complex interventions can be used to activate mechanisms to improve health. While this idea was accepted by some, it was also met with resistance, particularly from some realist evaluators who believe that trials are inextricably positivist and dependent on constant conjunctions to understand causation, and that realist trials are unfeasible because participants and contexts will be insufficiently diverse to enable the testing of context-mechanism-outcome configurations. In this paper, we reflect on analyses of qualitative and quantitative data from the Initiating Change Locally in Bullying and Aggression through the School Environment (INCLUSIVE) trial, and whether these are useful and aligned with realism. We summarise the concerns expressed by realists and reflect on the philosophical and practical challenges that we encountered and whether or not they are related to the trial's design. Finally, we reflect on the trial's weaknesses and highlight areas that future researchers might consider when running realist trials. We conclude that realist randomised controlled trials are philosophically coherent, practically feasible, and can produce nuanced findings.

Keywords: RCTs, Realist evaluation, Randomised trials, Complex interventions, Adolescent health, Bullying, School environment

Introduction

In 2012, the INCLUSIVE trial was presented as being the first realist randomised controlled trial (RCT) [1]. Realist trials aim to examine how intervention resources introduced into various contexts enable the activation of contextually contingent mechanisms which generate improvements in health, and assess how those vary by context. Realist evaluation within RCTs should minimise bias and confounding in these analyses of effects.

However, proposals for realist trials have been met with concerns about the philosophical compatibility between realism and RCTs, and practical concerns, particularly in relation to the assessment of mechanisms and whether sufficiently diverse settings can be included to test context-mechanism-outcome configurations (CMOCs). Since this debate began, we have completed our trial [2] and continue to analyse data from it. In this paper, we reflect on how we engaged with realist evaluation, what challenges we faced and whether or not we were able to conduct an informative realist RCT.

The INCLUSIVE trial evaluated Learning Together (LT), a complex, whole-school intervention that provided resources intended to enable secondary schools to

* Correspondence: Emily.warren@lshtm.ac.uk

¹Department of Public Health, Environments and Society, Faculty of Public Health and Policy, London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK

Full list of author information is available at the end of the article



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

reduce bullying and aggression. These resources included an intervention manual, an annual needs assessment report (NAR) generated from an annual student survey, a social and emotional learning curriculum, an external facilitator in the first 2 years of the intervention and training on restorative practice (a half-day training introducing restorative practices for all staff members and a 3-day intensive training for selected staff). These resources were provided to enable the following school processes: convening an action group with at least six students and six staff members to meet at least once every half-term; this group reviewing rules and policies to ensure they were supportive of restorative practices; this group deciding and implementing local actions based on the NAR; staff using restorative practice to address student conflict or misbehaviour (convening meetings between bullies and victims to understand the source of problems, give the victim the opportunity to explain how they feel and give the bully the opportunity to listen, and make amends); and staff teaching the curriculum.

The theory of change for LT was informed by a mid-range theory known as the theory of human functioning and school organisation [3]. This proposes that schools have an instructional order (concerning academic learning) and regulatory order (concerning the social norms, behavioural expectations and shared values). Commitment to these orders can be increased by “reframing” school practices on student needs and “eroding” the boundaries that separate students from staff, student groups from each other, students’ intellectual learning from their personal development and the school from the surrounding community. The theory proposes that reframing and boundary erosion will particularly engender the commitment of students from deprived backgrounds for whom school cultures may be particularly alienating and for whom engagement with education may be more challenging. Building commitment to school will, the theory proposes, equip students with the skills and social relationships so that they avoid participation in anti-school groups and risk-taking behaviours [4, 5].

Informed by this mid-range theory, LT’s theory of change proposes that the intervention resources will help schools build student commitment via reframing school practices around student needs and eroding boundaries between staff/students and different areas of the curriculum. This is theorised as being achieved via the action groups bringing staff and students together in a constructive environment to make collaborative decisions and improve school practices, restorative practices focusing on students’ needs for genuine conflict resolution and the curriculum addressing student needs for social and emotional skills and eroding the boundaries

between students’ personal development and academic learning.

INCLUSIVE was a 3-year cluster RCT of the LT intervention including a mixed-method process evaluation [1, 6], designed to be the first realist RCT [1]. The trial aimed to embrace realist approaches by using qualitative research to assess, refine and augment the starting theory of change and inform a number of CMOCs and using moderation and mediation analyses to test these CMOCs. These ideas were quickly incorporated into the Medical Research Council’s guidelines on process evaluations for complex interventions [7] but were met with criticism from some realist evaluators [8, 9]. In this paper, we seek to reflect on (1) the methods and findings of our analyses, identifying where these aligned with realist aims and approaches; (2) the challenges raised about realist trials, whether we encountered these and how we addressed them; and (3) whether or not we were able to conduct a realist trial which generated useful information on what works, for whom, and under what conditions.

The methods and findings of the INCLUSIVE trial

We have reported analyses of the outcome and process evaluation data using thematic content analysis [10], a variant of grounded theory called dimensional analysis [11], moderation [2], mediation [12], moderated-mediation [13] and qualitative comparative analyses (QCA) [14]. Some publications have explicitly referred to realist evaluation [11, 13, 14] while others have not. Likewise, some of these methods, such as moderator and mediator analyses [15, 16], are controversial within realist circles, while others are not [17–19].

The trial’s overall analysis of primary outcomes reported a significant difference in bullying victimisation at 36-month follow-up between schools allocated to the intervention arm compared to the control, but no difference in aggressive behaviours. In terms of secondary outcomes, the intervention was associated with benefits in terms of quality of life, mental well-being, psychological difficulties, smoking tobacco, drinking alcohol and drug use [2]. These findings are important in terms of overall public health impacts but are not aligned with realist questions on what works for whom, under what conditions and how [15]. Pre-specified subgroup (moderator) analyses showed that LT was more effective for boys, for students who had previously been bullied and for those with higher reported levels of aggression at baseline. However, contrary to our theory of change, no differences in impacts by socio-economic status were found [2]. Such subgroup analyses are common in trials and describe “for whom” health was improved, but are not realist in orientation because they are generally not focused on testing CMOCs. However, in the case of INCLUSIVE, our hypotheses about the intervention having

greater impacts among those of lower socio-economic status were based on a CMOC concerning how intervention impacts would be greater among disadvantaged students who are more likely to benefit from an intervention aiming to build commitment to school.

Our next attempt to understand mechanisms involved using a causal-steps mediation analysis [20] to assess whether our a priori theories about whether changes to school organisation, student commitment to school and involvement in anti-school peer groups were implicated in overall mechanisms (but not whether these varied between contexts). Based on our mid-range theory and theory of change, we hypothesised that intervention effects on bullying victimisation at final follow-up (36 months) would be mediated by school climate (as reported by students using the Beyond Blue School Climate Questionnaire [21]), school organisation (as reported by staff using a novel measure created by the INCLUSIVE trial team [22]) and involvement with delinquent peers (as measured by the Young People's Development Programme measure [23]) all measured at interim follow-up at 24 months. We found that only contact with delinquent peers was impacted by the intervention at 24 months. Intervention impacts on school climate did emerge but not until 36 months. Adjustment for these mediators did not reduce the association between intervention allocation and bullying victimisation, suggesting that, when examining all schools together, there was no evidence of changes in school organisation, student commitment or student involvement in anti-school peer groups being implicated in mechanisms generating outcomes [12]. However, these analyses did not tell us whether these mechanisms might be operating in some schools but not others.

Before we could explore this question of mechanisms operating differently in different schools, we decided we needed to better understand these mechanisms so that we could refine our CMOCs prior to further statistical analyses. Therefore, we undertook analyses of qualitative data to better understand how implementation and mechanisms might vary across schools. We first assessed how implementation fidelity of action groups varied between schools, also exploring how participants described their experiences on action groups, what factors influenced implementation and what consequences they had for schools. We found that in schools where the action group was led by senior staff who were not overwhelmed with other pressures, action groups proved a powerful motor of whole-school change. Having staff-members consistently attend meetings and communicate respectfully with students was described as improving relationships between staff and students, increasing student self-confidence and motivating students to work harder in class [10]. This informed our refinement of

CMOCs by helping us understand the ways in which LT resources were used differently in different schools.

We undertook further qualitative analyses to explore mechanisms in more depth and refine our ideas about how and under what conditions these might operate in schools. This involved analysis of qualitative data from case-study schools using "dimensional analysis", a variant of grounded theory which aims to understand phenomenon in terms of their context, conditions, processes and consequences [24, 25] (a framework with obvious resonance with realist CMOC terminology despite the use of slightly different terms). This analysis identified three mechanisms whereby the intervention might reduce bullying differently in different schools, each consisting of smaller sub-mechanisms. The first mechanism involved a process of increasing commitment to school by giving students new roles, a forum to share their experiences of being at the school and working with teachers to address shared problems. Such processes could generate consequences of building respectful and warm relationships with staff and increasing students' sense of belonging at school. This was likely to ensue in conditions in which schools had the capacity and space to engage in such elaborate processes. We thought of these in terms of schools having pre-existing good leadership, less distractions from other problems and a pre-existing inclusive ethos to build on. The second mechanism involved a process of building healthy relationships and behaviours by modelling and teaching pro-social skills via restorative practices, with the consequence of reducing misbehaviour and teaching non-violent conflict management. Such processes required staff who were committed to implementing restorative practice and were more likely to be transformative in schools where most student did not already possess strong pro-social skills. The third mechanism involved a process of de-escalating bullying among a core group of aggressive students via creating a space in which perpetrators could learn about the impacts of their behaviour. Such processes had the consequences of these students learning to empathise, experiencing shame, expressing contrition and accepting responsibility for their actions. Again, such processes were more likely in aggressive or violent schools where committed staff recognised the need and had the capacity to implement restorative practice [11]. Thus, the qualitative data suggested much more detailed ideas about mechanisms and in which schools these mechanisms would generate outcomes [11].

The hypotheses generated through the above qualitative analyses were assessed in two further quantitative analyses using different methods. In the first instance, we used moderated-mediation analyses to explore the first mechanism described above. Specifically, we assessed whether student sense of belonging at interim

follow-up might be a mediator of intervention effects on bullying and mental health at final follow-up in schools with certain contextual features. We hypothesised that this would be the case in schools with strong leadership (as indicated by government inspection judgements collected at baseline), low baseline rates of bullying and high baseline student inclusion as measured in our questionnaires. Analysis showed that in schools with these features but not in others, student belonging at interim follow-up did indeed mediate reductions in bullying [13]. Reductions in bullying occurred in other schools but were not mediated by student belonging. Thus, this analysis supported our CMOc that increased student belonging is implicated in mechanisms reducing bullying but only when schools possess the prior capacity, culture and space to promote student belonging via elaborate processes of student engagement. We concluded that, in other schools, other mechanisms, perhaps aligned with mechanisms 2 and 3 described above, were generating reductions in bullying.

In the second instance, we used QCA to explore the complex pathways between allocation to the intervention arm and changes in bullying. Whereas mediation and moderation analyses rely on probabilistic statistics and can only examine the inter-relationships between a small number of variables, QCA instead examines how more complex combinations of multiple conditions appear to enable or preclude the emergence of an outcome, using Boolean logic. A benefit of QCA is that it not only shows the possible pathways to an outcome, but it also shows the pathways that do not lead to the outcome. Our QCA suggested that, as we expected, schools did not need to activate all of the mechanisms identified in the qualitative research to decrease bullying, and that under the correct conditions, often the activation of a single mechanism was sufficient to reduce bullying. Because the data were from a trial with a comparison group of schools, we were also able to explore whether similar mechanisms might occur in schools not in receipt of intervention resources, bolstering our belief that the mechanisms we identified were plausible, transferable, causal and in realist terms, emerging from the realm of the real.

In the next section, we revisit the debate about realist trials and reflect on whether we experienced anticipated challenges and how we addressed them.

Concerns about realist trials, whether we encountered them, and how we responded

Realist evaluation and critical realism is a broad church with internal disagreements, especially in relation to the use of quantitative data and the usefulness of trials. Central to all interpretations of realism, however, are three interconnected beliefs: reality exists and is independent of human knowledge (ontological realism); knowledge is

always incomplete and dependant on the context of its discovery (epistemic relativism); and the rational adjudication between competing claims is possible because reality is intransitive and our means of understanding is transitive and can thus be improved when new knowledge is found (judgmental rationality).

RCTs are not inimical to these tenets of realism. The summary of the studies above suggests that RCTs are a study design which can be used to gather various types of data and analyse these using various methods. While the overall analyses of trial effect sizes clearly do not align with realist evaluations' concerns with what works, for whom and how, our subsequent analyses suggest that RCTs can nonetheless provide data for analyses which do align with realist concerns, while minimising bias and confounding.

However, several concerns have been expressed by realist evaluators about realist trials, which we now consider. These concerns fall under two key themes. Firstly, some realists regard RCTs as irretrievably positivist in philosophy and reliant on successivism to understand causality, and so are incongruent with realist analysis. Secondly, some realists argue that, in practical terms, RCTs are too narrow in scope to enable realist analyses. Below, we summarise these concerns, consider whether they arose within our trial and describe how we responded to these challenges.

Concerns about positivism and successivism

The first concern is that RCTs are positivist [9, 15, 26–29]. We have already published a paper on the key tenets of positivism, considering whether or not trials in general are, of necessity or in practice, positivist [30]. We will not repeat those arguments at length but provide a short summary and then consider the case of our own trial. The philosophical and social science literature delineates four key tenets of positivism thus (1) scientific knowledge is derived from direct, sensory observation; (2) theoretical terms must directly equate with empirical measurements with no reference to deeper, unobservable mechanisms of causation; (3) the objective of positivist inquiry is to generate universally applicable laws; and (4) the same methods can be used in the natural and social sciences.

In our previous paper, we argued that, in regard to the first tenet, trials more often use a hypothetico-deductive than an inductive approach, using data not to build theory but rather to assess the falsifiability of hypotheses generated from a priori theory. In response to the second tenet, many trials are theorised purely in terms of the hypothesised association between variables, but this is not a necessary feature. Trials may evaluate interventions informed by theories of change derived from mid-range theory describing deeper mechanisms which need

not align with empirical measures. In regard to the third assertion, few trialists claim their results are universally generalisable and many identify factors likely to define the limits of transferability. Some trials, notably realist trials, aim to provide guidance on transferability not in terms of statistical generalisability but via developing or refining theory of how interventions generate certain outcomes in certain settings. Finally, although trials may be used in both the natural and social sciences, trials of social interventions are distinct in their inclusion of qualitative data to explore hermeneutic questions of meaning and agency which are not relevant, for example, in trials of purely natural science (e.g. agricultural) interventions.

Applying these arguments to our own trial, it is clear that INCLUSIVE was hypothetico-deductive in orientation, deriving our hypotheses from a priori theory of change based on a mid-range sociological theory which engaged with the deep mechanisms by which outcomes are generated and were not reducible to associations between empirical constructs. We also aimed to develop findings which might be contingently, but certainly not universally, transferable to other contexts dependent on specific theorised factors. Finally, we used a variety of methods including those rooted in a hermeneutic approach such as interviews, focus groups and semi-structured observations to understand social phenomena.

Related to concerns about trials being positivist (and in particular to the lack of deep theorisation) are some realists' concerns about how trialists view the world. Realist evaluators commonly argue that trialists think in terms of "interventions working" and are therefore insensitive to the fact that outcomes are actually the result of changes in peoples' reasoning and actions in response to the availability of novel resources [9, 15]. We acknowledge that trialists (and other evaluators) often write in terms of "intervention X causing outcome Y" but we also believe that this is generally a linguistic short-cut that avoids the consistent wordiness which would be required to remind readers that it is how people employ intervention resources which might generate outcomes.

The debate about realist trials has also revealed a concern relating to positivism, the use of statistical associations between allocation to intervention/control arms, and measures of outcomes as a basis for assessing causality. Although not strictly part of the tenets of positivism, realists have criticised trialists for understanding causation through a "successionist" focus on constant conjunction, arguing that this approach fails to appreciate that, in "open systems", simple regularities rarely occur [15]:

... events arise from the workings of mechanisms which derived from the structures of objects, and they take place within geo-historical contexts. This contrasts with approaches which treat the world as if it were no more than patterns of events, to be registered by recording punctiform data regarding 'variables' and looking for regularities among them... Given the variety and changeability of the contexts of social life, this absence of regular associations between 'causes' and 'effects' should be expected [16], pp. 15–16

We agree that trial analyses of overall population effects do not align with realist concerns but counter that other analyses are possible within trial designs that do provide evidence useful for realist questions. We hope that the examples provided above of our own analyses support this point. We would also point out that use of probabilistic statistical measures assessing the associations between two variables (or whether such associations are moderated or mediated by third variables) does not imply a belief that causation can only be considered in terms of constant conjunctions. Indeed, the very use of statistical analyses of the regularity of such conjunctions recognises that these are not constant. The use of statistical moderation analyses in particular reflects a recognition that any conjunctions are contingent on other factors. The use of QCA is also possible within trials, as we have demonstrated, and this rests on an assumption that causality is best assessed by exploring the contingent inter-relationships between multiple factors [19]. Questions about causal attribution are central to trial analyses not because interventions are thought to be the exclusive, determining source of causation but because trials seek to explore how the mechanisms triggered by the introduction of new resources into contexts interacts with all the other mechanisms operating in that context to generate new outcomes. In this sense, trials measure added-value, not unique causation.

The final related concern is about aggregation. Marchal et al. argue that even if a process evaluation attempts to study mechanisms, "such information is lost in the aggregation process required to give RCTs their power" [9], pp. 125–126. While generating an effect size at the aggregate level is important in trials, the same data sources can be used to answer questions where it would be nonsensical to focus on net-effects. For example, in the abovementioned QCA, we did not aggregate data above the level of individual schools.

Concerns about the practical feasibility of realist RCTs

The second area of concern relates to whether or not RCTs can practically provide the necessary data to address realist questions. These concerns were difficult to

respond to before we finished our analyses, but we are now able to reflect on the challenges we faced and how we addressed them. These concerns include whether randomisation and control stifle our ability to explore CMOCs; RCTs are insufficiently theorised; and trials are only concerned with attribution.

A key concern expressed by those who feel realist RCTs are ill-conceived is that randomisation and control stifle our ability to test CMOCs [8, 9, 15]. This is a significant concern and requires being broken down into its component parts. Firstly, random allocation is important because it ensures that comparisons between intervention and control sites are fair and minimise bias and confounding [13]. As Bonell et al. have previously argued, “Randomisation is merely a practical tool to reduce confounding. It does not fundamentally change the nature of the way we view or research the social world, or affect how we will use comparative empirical data to test hypotheses about mechanisms” [31], pp. 3. Many realist evaluations employ natural experiments which, like RCTs, involve an internal or external comparison group. The only difference in RCTs is that such comparison groups are constructed in such a way that comparisons are balanced. Secondly, control groups are scientifically useful because estimating the effectiveness of interventions is important, and its presence does not diminish our ability to employ other methods or answer other questions.

Another issue related to control is the concern that, in RCTs, the recruitment of participants (individuals or clusters of individuals, for example in schools or villages) is too tightly controlled so that these are insufficiently diverse to allow for cross-contextual comparisons [9], which are necessary for exploring CMOCs. This concern reflects the obviously insufficient diversity in many trials, particularly in biomedical efficacy studies where certain populations are routinely under-recruited or actively excluded [32]. Such homogeneity is not, however, a necessary or desirable feature of RCTs, especially pragmatic trials of public health and other social interventions, where the aim is often to ensure participants reflect the population from which they were recruited. In the INCLUSIVE trial and with the strong support of the funder, we aimed to recruit a diversity of schools and students that reflected the profile of schools and students in England. Our schools were representative across a range of factors including size, population demographic factors, deprivation and educational performance. Participating schools were, however, more likely to have a positive government inspection rating compared to other schools. Within our random allocation of schools to intervention or control group, we stratified randomisation by single-sex versus mixed-sex entry, school-level socio-economic deprivation and student

examination attainment to ensure that the trial arms were balanced according to these factors. This stratification meant that, although both intervention and control groups were highly diverse according to these factors, they contained the same range of diversity. This diversity meant that in our moderation and moderated-mediation analyses and QCA, our sample was diverse enough to allow us to explore how indicators of mechanisms and outcomes varied across a diversity of school contexts. This diversity also meant that we could explore contextual contingencies in our qualitative analysis. By purposively selecting schools based on their contextual diversity, we were able to explore what contextual features of a school or community appear to be key to the intervention being implemented (or not), and how and for what purposes intervention resources were used. This allowed us to develop emergent CMOCs [11]. This is not to say that all trials are successful in recruiting diverse samples, but the lack is a weakness in specific studies and not an inherent feature of the RCTs.

Marchal et al. have expressed concerns that RCTs are unable to explore mechanisms and argued, “Even if evaluations of implementation, process, and context are added [to a trial], they can elucidate just that—the intensity, fidelity, and actual process of implementation, and the context in which the intervention took place” [9]. LT was a complex intervention, comprising multiple components and enabling local staff to implement actions appropriate to their school. Our evaluation was therefore built around the assumption that a vast array of mechanisms would be activated by the availability of novel resources because agents would use them in various ways based on their context, which would generate different outcomes in different schools. Our conceptually rich, a priori theory of change facilitated the exploration of these through quantitative and qualitative research. This theoretical underpinning enabled us to identify suitable quantitative measures to include in student and staff surveys. It also enabled us to include suitable prompts to explore in the qualitative data collection guides. This allowed us to focus not only how schools implemented the intervention but also how intervention activities triggered mechanisms in their school. For example, two unanticipated mechanisms to reducing bullying emerged from interviews with students who had participated in restorative conferences: learning empathy and accepting their punishment as fair. These mechanisms were much more likely to activate for students with weaker social skills who benefitted from a more direct lesson in social skills and were less effective in changing behaviour in schools where students knew that their behaviour was unacceptable when they chose to engage in it [11]. Thus, process evaluation data was not just used to study implementation and fidelity,

but enabled the discovery and refinement of theorised mechanisms.

Are realist trials possible and do they generate useful findings?

We believe that the INCLUSIVE trial's research programme demonstrates that RCTs can provide evidence that is philosophically appropriate for answering realist questions about interventions. Realist trials are possible. The trial we conducted evaluated an intervention with a theory of change based on deep sociological theory of the mechanisms that generate bullying. The trial generated data which allowed for the refinement and testing of CMOCs relating to our intervention. Through it, we discovered that in some schools benefiting from strong management, inclusive cultures and minimal distraction from acute problems, schools could enact complex processes of student involvement which triggered mechanisms of building student belonging in school, in turn generating reductions in bullying and improvements in mental health. In other, more challenged schools, staff implementing processes of restorative practice was sufficient to enable students to develop the skills and attitudes needed to avoid or terminate conflict, which also generated reductions in bullying and improvements in mental health. This allowed us to refine our starting theory of change to provide a much more nuanced picture of what worked for whom and how. Our research also generated nuanced findings which could inform practical intervention modifications, and identify potentially appropriate or inappropriate contexts for intervention transfer.

It is important to note that while some of the aforementioned analyses were explicitly realist, neither the analytic methods we used nor the sorts of questions we sought to explore are unique to realists. For example, researchers on the Aban Aya Youth Project have used growth mixture modelling techniques and found that young men at higher risk of violent trajectories gained the most preventative benefit from a whole-school intervention [33]. Analysis of the KiVa anti-bullying intervention identified both individual and classroom-level mediators which reduced the risk of bullying. By finding that bullies are more likely to offend in contexts where peers encourage violent behaviours [34], researchers were able to specify more clearly how changing peer norms can contribute to decreasing bullying, and incorporated this into new theories and novel interventions. Hence, RCTs need not be explicitly realist in orientation to generate analyses of interest to realists. Nonetheless, we think employing an explicitly realist position does facilitate a more comprehensive assessment of how outcomes are generated by contextually contingent mechanisms.

To incorporate the benefits of realism within the structure provided by a trial, a number of considerations need to be addressed while the trial is being planned. The resources provided in an intervention and the evaluation should both be informed by an intervention theory of change informed by an appropriately selected mid-range theory. Between what Robert Merton called "piecemeal empiricism" and grand theory, mid-range theories are specific to the phenomena of interest but still have sufficient analytic purchase to be generalisable [35]. Based on this mid-range theory, realist trialists need to be explicit about why all intervention resources are being included, and what mechanisms their use is anticipated to activate, and how this varies by population and place. Thus, unlike most conventional theories of change, those used in realist studies must engage with how context and mechanisms interact to generate outcomes. Realist trials should include moderation analyses to shed light on such interactions. Traditionally in trials, moderators examined might include sex, age and socio-economic status, but in realist trials more specific indicators should be included, informed by the theory of change. In the INCLUSIVE trial, these included baseline experiences with bullying and aggression at the individual-level, and school leaderships, ethos, and value-added score at the school-level. As discussed earlier, it is important for realist trials to recruit sufficiently diverse samples of people and/or clusters to explore how context and mechanisms interact to generate outcomes.

Finally, the process evaluations of realist RCTs will focus not only on questions of intervention feasibility, fidelity and acceptability, but also on mechanisms. In a realist trial, diverse stakeholders and participants are asked to describe their context, their positionality, their experiences using intervention resources and what they perceive as having occurred as a result of this use. Realist interviewing, in which participants help develop or refine the study's logic model [36] can also be used. Participants can be asked about the processes through which they believe change is happening and what the consequences of these processes may be. For example, in INCLUSIVE, instead of speaking about decreasing bullying, students would speak about liking teachers more or getting along better with others in class, which we then theorised would lead to decreased bullying.

There are numerous benefits of incorporating realist approaches into trials. By focusing on what works, for whom, under what conditions and how, trialists' attention is continually focused on these more specific evaluation questions. However, the benefits also extend beyond evaluation to enable a deeper exploration into the phenomena of interest. While the primary function of INCLUSIVE was to evaluate LT, we also deepened our understanding of bullying, the impact of the school

environment on improving mental health and how empathy and forgiveness affect the development of supportive peer-networks. We would argue that our study has demonstrated the value of using RCT designs within realist evaluation. Crucially, randomisation minimises bias and confounding in estimates of intervention effects and moderators, ensuring that we can provide the best possible assessment of what works for whom and under what circumstances.

While detailed findings about the role of context, the activation of mechanisms and explorations of who benefited from LT were written about in the course of the trial, that does not mean that the study could not have been improved. Despite having context-specific hypotheses in our theory of change, the intervention's starting logic model was overly simplistic and did not depict differences in how LT was theorised to activate different mechanisms in different contexts. Our logic model did not reflect that one of the objectives of LT as a whole-school intervention was to change the context of the intervention's implementation throughout the trial. This was especially important because of the variety of schools in the trial. The context at the beginning of the trial for some schools was similar to the context of other schools at final follow-up. Moreover, the logic model was linear, despite acknowledging that complex interventions often contain feedback loops and work in non-linear ways.

Qualitative topic guides would have been improved by focusing more on mechanisms and less on implementation and fidelity, on which we were able to gather from other sources. The baseline and follow-up surveys did not contain measures on all hypothesised mechanisms, such as student-centred framing in relation to teaching and learning. This presented particular challenges with the moderated-mediation and qualitative comparative analyses, in which we either could not explore all the CMOCs which we generated from prior theory and qualitative analyses. However, these limitations were not caused by our use of an RCT design. Despite the challenges arising from insufficient measures or unpredicted mechanisms, the articulation of CMOCs contributed to the development of a broader, realist theory about school environments, bullying as a social phenomenon and how and for whom the introduction of interventions can improve health. Even though we were unable to test all of our hypotheses, they can be used to improve theory and future intervention development.

It is also important to be clear that some of our hypotheses were wrong. Based on the mid-range theory that informed our theory of change, we anticipated greater benefits for socio-economically disadvantaged children [3] but the evidence did not support this [2]. The hypotheses derived from qualitative data and tested

using QCA also showed that while participants were able to express a narrative about how LT was used in their school and led to changes, the quantitative evidence did not always bear this out. Indeed, some of the mechanisms that people predicted would be most important had average or no impact [14]. This was not an unexpected finding as within a realist paradigm, knowledge is always partial and perspectival [37].

We deviated from the three-step structure that was proposed in the team's articulation of what a realist trial would look like [38]. In the original proposal, we planned to create a logic model and theory of change, and use those to develop a priori CMO hypotheses. In phase two, the trial's process evaluation would be conducted, and the qualitative data would be used to refine those hypotheses. In stage three, data from the process and outcome evaluations would be brought together and the refined CMOs would be tested with moderator and mediator analyses to refine the theory of change. We completed phase one, but our next step was unpacking implementation and assessing contextual variation, and why it was more or less acceptable (and used) by some people and in some places [10]. Rather than simply refine CMOs in light of qualitative data, we remained sensitised to our theory of change and used dimensional analysis to explore participants' accounts to build emergent CMOs [11]. Those CMOs were then tested using QCA [14]. Separate to this sequential set of analyses, we also ran moderator [2], mediator [12], and moderated-mediation [13] analyses which more closely followed the description of what realist trials might look like. This reflected a complex multi-collaboration acting to generate useful findings at speed. Many of the aforementioned analyses were not part of the trial's original protocol and were exploratory in nature. Therefore, it would be helpful if future trials of whole-school anti-bullying interventions included analogous analyses into their protocol to assess whether our findings are confirmed.

Finally, it is also important to note that our team's understanding of realism matured as we carried out this work. In original papers, we wrote that "Realist evaluators have viewed interventions as 'working' by introducing *mechanisms* that interact with features of the *context* to produce *outcomes*" [38], pg 2. This was incorrect: resources not mechanisms are introduced into a context.

Despite the aforementioned limitations, we were able to answer detailed questions about how, for whom, under what conditions and to what extent bullying was reduced following the distribution of LT resources to schools in the intervention arm, and what environmental or inter-personal features seemed to affect the generation of those and other outcomes. A simple but

important reflection on this process is that RCTs are what researchers make of them. They can be designed to merely assess overall intervention effects, or they can be designed to answer questions which are central to realist enquiry. Most RCTs fall somewhere between those two extremes but crucially, it is not the study design but the detailed planning of theorisation, data collection and analyses that determines what questions a trial may answer.

Acknowledgements

We are grateful to the staff and students of participating schools and for their completion of surveys and participation in interviews and focus groups. We are also grateful for the advice and support of our Trial Steering Committee and Data Monitoring Committee.

Authors' contributions

EW made substantial contributions to conception and design of the work and analysis and interpretation of the data and drafted the manuscript. GJMT contributed to the design, analysis and interpretation of the data and revised the manuscript. CB made substantial contributions to conception and design of the work and substantially revised the manuscript. The authors read and approved the final manuscript.

Funding

Funding for the intervention and evaluation came from the Education Endowment Foundation and the National Institute of Health Research in England under its Public Health Research Board (12/153/60).

Availability of data and materials

Not available

Declarations

Ethics approval and consent to participate

The trial was approved by the UCL Ethics Committee (ref: 5248/001)

Consent for publication

NA

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Public Health, Environments and Society, Faculty of Public Health and Policy, London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK. ²Institute of Health Research, College of Medicine and Health, University of Exeter, Exeter, UK

Received: 25 October 2021 Accepted: 9 December 2021

Published online: 28 January 2022

References

- Bonell C, Fletcher A, Morton M, Lorenc T, Moore L. Realist randomised controlled trials: a new approach to evaluating complex public health interventions. *Soc Sci Med.* 2012;75(12):2299–306. <https://doi.org/10.1016/j.socscimed.2012.08.032>.
- Bonell C, Allen E, Warren E, McGowan J, Bevilacqua L, Jamal F, et al. Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial. *Lancet.* 2018;392(10163):2452–64. [https://doi.org/10.1016/S0140-6736\(18\)31782-3](https://doi.org/10.1016/S0140-6736(18)31782-3).
- Markham WA, Aveyard P. A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice. *Soc Sci Med.* 2003;56(6):1209–20. [https://doi.org/10.1016/S0277-9536\(02\)00120-X](https://doi.org/10.1016/S0277-9536(02)00120-X).
- Willis, P.E. and P. Willis, Learning to labor: How working class kids get working class jobs. 1981: Columbia University Press.
- Jamal F, Fletcher A, Harden A, Wells H, Thomas J, Bonell C. The school environment and student health: a systematic review and meta-ethnography of qualitative research. *BMC Public Health.* 2013;13(1):798. <https://doi.org/10.1186/1471-2458-13-798>.
- Bonell C, Allen E, Christie D, Elbourne D, Fletcher A, Grieve R, et al. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial. *Trials.* 2014;15(1):1. <https://doi.org/10.1186/1745-6215-15-381>.
- Moore GF, et al. Process evaluation of complex interventions: Medical Research Council guidance. *bmj.* 2015;350:h1258.
- Van Belle S, et al. Can 'realist' randomised controlled trials be genuinely realist? *Trials.* 2016;17(1). <https://doi.org/10.1186/s13063-016-1407-0>.
- Marchal B, Westhorp G, Wong G, van Belle S, Greenhalgh T, Kegels G, et al. Realist RCTs of complex interventions—an oxymoron. *Soc Sci Med.* 2013;94:125–6. <https://doi.org/10.1016/j.socscimed.2013.06.025>.
- Warren E, Bevilacqua L, Opondo C, Allen E, Mathiot A, West G, et al. Action groups as a participative strategy for leading whole-school health promotion: results on implementation from the INCLUSIVE trial in English secondary schools. *Brit Educ Res J.* 2019;45(5):979–1000. <https://doi.org/10.1002/berj.3547>.
- Warren E, et al. Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention. *Trials.* 2020;21(1):1–14.
- Bonell C, Allen E, Opondo C, Warren E, Elbourne DR, Sturgess J, et al. Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention. *J Epidemiol Community Health.* 2019;73(5):455–64. <https://doi.org/10.1136/jech-2018-211443>.
- Melendez-Torres G, et al. Moderated mediation analyses to assess intervention mechanisms for impacts on victimisation, psycho-social problems and mental wellbeing: evidence from the INCLUSIVE realist randomized trial. *Soc Sci Med.* 2021;113984. <https://doi.org/10.1016/j.socscimed.2021.113984>.
- Warren E, Melendez-Torres GJ, Bonell C. Using fuzzy set qualitative comparative analysis (fsQCA) to explore the causal pathways to reduced bullying victimization in a whole-school intervention: results from a cluster randomized controlled trial. *J School Violence (under review).* 2021.
- Pawson, R. and N. Tilley, Realistic evaluation. 1997: Sage.
- Sayer, A. Realism and social science. 1999: Sage.
- Befani B, Ledermann S, Sager F. Realistic evaluation and QCA: conceptual parallels and an empirical application. *Evaluation.* 2007;13(2):171–92. <https://doi.org/10.1177/1356389007075222>.
- Oliver C. Critical realist grounded theory: a new approach for social work research. *Brit J Soc Work.* 2012;42(2):371–87. <https://doi.org/10.1093/bjsw/bcr064>.
- Hanckel B, Petticrew M, Thomas J, Green J. The use of Qualitative Comparative Analysis (QCA) to address causality in complex systems: a systematic review of research on public health interventions. *BMC public health.* 2021;21(1):1–22. <https://doi.org/10.1186/s12889-021-10926-2>.
- Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* 1986;51(6):1173–82. <https://doi.org/10.1037/0022-3514.51.6.1173>.
- Sawyer MG, Pfeiffer S, Spence SH, Bond L, Graetz B, Kay D, et al. School-based prevention of depression: a randomised controlled study of the beyondblue schools research initiative. *J Child Psychol Psychiatry.* 2010; 51(2):199–209. <https://doi.org/10.1111/j.1469-7610.2009.02136.x>.
- Shackleton N, Fletcher A, Jamal F, Markham W, Aveyard P, Mathiot A, et al. A new measure of unhealthy school environments and its implications for critical assessments of health promotion in schools. *Crit Public Health.* 2017; 27(2):248–62. <https://doi.org/10.1080/09581596.2016.1191619>.
- Wiggins M, Bonell C, Sawtell M, Austerberry H, Burchett H, Allen E, et al. Health outcomes of youth development programme in England: prospective matched comparison study. *Bmj.* 2009;339(jul07):b2534. <https://doi.org/10.1136/bmj.b2534>.
- Schatzman L. Dimensional analysis: outline in prec form. University of California, San Francisco: Unpublished manuscript; 1980.
- Schatzman L. Dimensional analysis: notes on an alternative approach to the grounding of theory in qualitative research. *Soci Org Soc Process.* 1991:303–14.
- Green, J. and N. Thorogood, Qualitative methods for health research. 2013: Sage.
- Pearce W, Raman S. The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance. *Policy sciences.* 2014; 47(4):387–402. <https://doi.org/10.1007/s11077-014-9208-3>.

28. Rowe M, Oltmann C. Randomised controlled trials in educational research: ontological and epistemological limitations. *Afr J Health Prof Educ*. 2016; 8(1):6–8. <https://doi.org/10.7196/AJHPE.2016.v8i1.683>.
29. Hinds K, Dickson K. Realist synthesis: a critique and an alternative. *J Crit Realism*. 2021;20(1):1–17. <https://doi.org/10.1080/14767430.2020.1860425>.
30. Bonell C, Moore G, Warren E, Moore L. Are randomised controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services. *Trials*. 2018;19(1):238. <https://doi.org/10.1186/s13063-018-2589-4>.
31. Bonell C, Warren E, Fletcher A, Viner R. Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al. *Trials*. 2016;17(1):478. <https://doi.org/10.1186/s13063-016-1613-9>.
32. Perez, C.C., *Invisible women: exposing data bias in a world designed for men*. 2019: Random House.
33. Segawa E, Ngwe JE, Li Y, Flay BR, Aban Aya Coinvestigators. Evaluation of the effects of the Aban Aya Youth Project in reducing violence among African American adolescent males using latent class growth mixture modeling techniques. *Evaluation review*. 2005;29(2):128–48. <https://doi.org/10.1177/0193841X04271095>.
34. Saarento S, Boulton AJ, Salmivalli C. Reducing bullying and victimization: Student- and classroom-level mechanisms of change. *J Abnorm Child Psychol*. 2015;43(1):61–76. <https://doi.org/10.1007/s10802-013-9841-x>.
35. Merton, R.K., *On sociological theories of the middle range* [1949]. 1949; na.
36. Manzano A. The craft of interviewing in realist evaluation. *Evaluation*. 2016; 22(3):342–60. <https://doi.org/10.1177/1356389016638615>.
37. Bhaskar, R., *The possibility of naturalism: a philosophical critique of the contemporary human sciences*. 2014: Routledge.
38. Jamal F, Fletcher A, Shackleton N, Elbourne D, Viner R, Bonell C. The three stages of building and testing mid-level theories in a realist RCT: a theoretical and methodological case-example. *Trials*. 2015;16(1):1. <https://doi.org/10.1186/s13063-015-0980-y>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Conclusion

The aim of this thesis was to conduct analyses of the first attempted realist RCT and provide evidence regarding whether or not realist trials are feasible, can enable deeper understanding of how intervention works, for whom, under what circumstances, and how. The above paper summarised the work undertaken, concluding that realist trials are feasible, philosophically coherent and generate useful findings and commented on some potential implications for research.

By assessing implementation via the AGs and qualitative data from a diverse range of case study schools, we were able to show how the contexts in which LT was implemented activated a range of mechanisms and that mechanisms which were vital in some places or populations were less important or absent in others. Our QCA was the first study to my knowledge that used both arms of a trial to assess the generalisability of mechanisms identified from the intervention arm to the control arm to assess whether the theorized mechanisms appear to create similar outcomes via different resources in similar contexts.

Moderator, mediator, and moderated-mediation analyses cannot explore as many features simultaneously as a QCA and does not explore pathways in as much detail as is possible using Boolean operators, but it is also less liable to chance coincidence. Despite the notion of realist trials being called ‘oxymoronic’ [3] our QCA and moderated-mediation analysis uses the trial’s experimental design to explore theoretically informed, contingent mechanisms and the contexts of their emergence. [306]

Encouragingly, the qualitative analyses, the QCA, and the moderated mediation analyses all tell a similar story about what works, for whom, under what conditions and how.

The previous publication also comments on the implications for research. Below, I discuss implications for policy and practice and research, before I move to considering the limitations of LT and INCLUSIVE. Reflection on the weaknesses in each study have been explored in each publication.

Implications for practice, policy, and research:

In relation to potential implications for practice, the data support further consideration of some key issues. One of these considerations, currently under-explored in the

literature, is that intervention activities not expressly focused on bullying appear nonetheless to have significant potential to reduce it. When led by a competent and respected senior member of staff, AGs activated powerful mechanisms through which young people who felt disillusioned with school were able to connect to staff members over a shared purpose. Creating a space in which they had input and in which they could see teachers and school staff not merely as authority figures but people who were putting in effort to listen to them and understand their perspective, humanised staff and created bonds which motivated students to behave better both in class and to their peers more broadly. This was most likely to occur in schools with highly motivated staff and/or where school leaders acknowledged that improvements were needed. Contextual features, including low staff efficiency, de-prioritisation by the head teacher, and consistently facing acute and insurmountable amounts of other work, often overpowered this mechanisms and meant that change did not occur.[8] While I did not lead on the moderated mediation analyses, we similarly found that developing a sense of belonging to the school community mediated reductions in bullying in schools with strong leadership, below median levels of bullying and above median levels of school inclusivity at baseline.[306]

Not all schools were able to implement LT successfully. In a study that recruited both Ofsted rated 'outstanding' grammar schools and schools with low staff retention and high unmet needs, it was clear from the PE data that some schools required more support than others. Therefore, in future iterations of LT or similar interventions, implementation could be phased to better support schools with greater needs and/or lower capacity. Because of its broader application and usefulness across a number of challenges facing students and teachers, I would recommend beginning with RPs and then work up to implementing the AG. One challenge facing secondary schools in the UK is the expectation that schools are not only sites for curricular education, they are also sites in which complex social needs, including safe-guarding, mental health, family social support and reporting of children who are perceived at risk of extremism are meant to be addressed.[364] In some schools where students had complex unmet needs, teachers were unable to implement an intensive intervention despite having the greatest need for its benefits.

In relation to implications for research, INCLUSIVE showed that bullying-specific components of LT, such as RP conferences, were important for students who, for various reasons, did not understand or care that their actions were harmful. Having a guided and private conversation with the victim was often described by aggressors as a powerful source of constructive or integrative shame about their behaviour and motivated positive personal growth. This was not universal and some students appeared to be unmoved by the distress they had caused. Future evaluation should explore whether or not RPs have the potential to teach empathy via repetition. In two schools, remorse developed days after the meeting, so it may be that with longitudinal data, researchers can assess whether a dose-response relationship exists and whether or not that relationship is moderated by mental health or personality characteristics at baseline. I would hypothesise that students who engage in bullying because of a deficit in social skills will be significantly more amenable to behaviour change than students who engage in bullying and have psychopathic, narcissistic or sadistic tendencies and/or traits.

Finally, future research programmes should incorporate theories that address the various levels of causation. LT was built upon the sociological theory of human functioning and school organisation.[7] This was useful in explaining how schools as complex systems can be altered to improve students commitment to school and improve health. To improve future iterations of LT, other appropriately selected theory should be incorporated to better understand outcomes that emerges from various layers of causal 'depth.' [241] For example, evidence showed that many bullies learned empathy when they were confronted by seeing how upset their peers were as a consequences of their bullying. This is a causal mechanism that emerged at the social level and social theories can help us understand and explain this event. At the same time, other causal factors, emerging from psychology and biology were also likely activating to cause changes and these can be incorporated into evaluations. These could explore the psychology of forgiveness,[365] the role that adolescent development has on behaviour,[366] and how bullying affects cortisol levels.[112, 367] Westthorp describes this concept as a "theory ladder" in which various theories are used to correspond to the different levels of causation, ranging from the micro to macro.[368, 369]

Limitations

There are a number of limitations related to the intervention, evaluation and this thesis. The complex nature of LT made it difficult for some schools to implement fully. In England, where there is intense pressure on schools to maximise the academic attainment of their students, some schools did not want to prioritize an intervention the objectives of which were viewed as not aligned with their own.[\[370\]](#) This was a challenge despite recruitment materials and agreements with head teachers being explicit about the study aims. Coordinating the delivery of intervention resources to schools was also difficult and likely limited the impact that certain components had in some schools. For example, despite intervention delivery commencing in the autumn of 2015, some schools did not receive their whole-school training in RP until the spring, meaning that the staff lacked key knowledge connected to the trial's objectives for nearly one third of the trial. Had this been implemented earlier across all schools, the effect sizes from the trial may have been larger.

The SEL component of LT would benefit from a number of improvements. Firstly, the materials were not professional looking and some teachers found typographic mistakes in them. The curriculum was also modified from the Gatehouse Project intervention[\[84\]](#) which was originally intended for Australian students and was updated by a teacher who was not a SEL specialist. Given the diversity of the schools in the trial, the trial team faced challenges in trying to pitch the SEL curriculum at an appropriate level. Some schools complained that the intervention was infantilizing to their students while others thought it was impossibly complex for students to grasp. In future work, either multiple iterations could be created so that schools have access to resources that fit their needs or schools could be encouraged to use the resources provided to them as a template to modify based on their unique needs.

Finally, LT was primarily delivered by freelance facilitators and managed by a research team. It is possible that had LT been implemented by an independent organisation, the intervention would have been delivered better and we would have better data to understand issues around scalability.

INCLUSIVE also had weaknesses that are relevant to this study. Despite having a clear logic model and theory of change, not all of the anticipated mechanisms had corresponding measures incorporated into the student surveys which measured outcomes. Therefore, some mechanisms could not be tested, or could only be tested

with measures of questionable validity. For example, the QCA study would have been strengthened had validated measures for empathy, contrition and/or forgiveness been incorporated into relevant data collection tools. In the moderated mediation paper, we could model outcomes that emerged from increasing belonging but we lacked sufficient measures to describe the impact that RP had via increased empathy. While our interpretation is reasonable, it lacks definitive proof. Future trials can address this by analysing qualitative data prospectively so that there is sufficient time to introduce new measures into the interim and final follow-ups.

The study had a detailed PE protocol but the data gathered through interviews and focus group discussions were insufficiently focused on potential mechanisms in the first two years of the study. Finally, the trial overburdened schools with too much data collection. Adjustments were made to the INCLUSIVE protocol after we became aware that we were risking respondent fatigue with an overly intensive data collection schedule. Future PEs may need to build in a degree of flexibility to assess the experiences of the key staff that researchers depend on to coordinate data collection and ensure that they are not overwhelmed.

Despite these limitations, this thesis has shown that realist trials are possible and that data collected within an RCT can be used to understand what works, for whom, under what conditions, and how. This can be used to improve our understanding of a given social phenomenon, improve upon existing intervention and mid-range theories, guide intervention refinement, and provide evidence on generalisability and transferability.

References:

1. Pawson, R. and N. Tilley, *Realistic evaluation*. 1997: Sage.
2. Bonell, C., et al., *Realist randomised controlled trials: a new approach to evaluating complex public health interventions*. *Social science & medicine*, 2012. **75**(12): p. 2299-2306.
3. Marchal, B., et al., *Realist RCTs of complex interventions—an oxymoron*. *Social Science & Medicine*, 2013. **94**: p. 124-128.
4. Van Belle, S., et al., *Can 'realist' randomized controlled trials be genuinely realist?* *Trials*, 2016.
5. Bonell, C., et al., *Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial*. *Trials*, 2014. **15**(1): p. 1.
6. Bonell, C., et al., *Initiating change locally in bullying and aggression through the school environment (INCLUSIVE) trial: update to cluster randomised controlled trial protocol*. *Trials*, 2017. **18**(1): p. 238.
7. Markham, W.A. and P. Aveyard, *A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice*. *Social science & medicine*, 2003. **56**(6): p. 1209-1220.
8. Warren, E., et al., *Action groups as a participative strategy for leading whole-school health promotion: Results on implementation from the INCLUSIVE trial in English secondary schools*. *British Educational Research Journal*, 2019.
9. Warren, E., et al., *Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention*. *Trials*, 2020. **21**(1): p. 1-14.
10. Warren, E., G.J. Melendez-Torres, and C. Bonell, *Using fuzzy set qualitative comparative analysis (fsQCA) to explore the causal pathways to reduced bullying victimization in a whole-school intervention: results from a cluster randomized controlled trial*. *Journal of School Violence* (under review), 2021.
11. Warren, E., G.J. Melendez-Torres, and C. Bonell, *Are realist randomised controlled trials possible? A reflection on the INCLUSIVE evaluation of a whole-school, bullying-prevention intervention* *Trials*, 2022. **23**(82).
12. GOV.UK. *Bullying at school*. 23/03/2018]; Available from: <https://www.gov.uk/bullying-at-school/bullying-a-definition>.
13. Farrington, D.P., *Understanding and preventing bullying*. *Crime and justice*, 1993. **17**: p. 381-458.
14. Olweus, D., *Bully/victim problems among schoolchildren: Basic facts and effects of a school based intervention program*. *The development and treatment of childhood aggression*, 1991. **17**(17): p. 411-448.
15. Gaffney, H., M.M. Ttofi, and D.P. Farrington, *Evaluating the effectiveness of school-bullying prevention programs: An updated meta-analytical review*. *Aggression and violent behavior*, 2019. **45**: p. 111-133.
16. Cantone, E., et al., *Interventions on bullying and cyberbullying in schools: A systematic review*. *Clinical practice and epidemiology in mental health: CP & EMH*, 2015. **11**(Suppl 1 M4): p. 58.
17. Chamberlain, T., et al., *Tellus4 national report*. London: Department for Children, Schools and Families, 2010.
18. Roe, S. and J. Ashe, *Young people and crime: Findings from the 2006 Offending, Crime and Justice survey*. 2008.
19. Patchin, J.W. and S. Hinduja, *Bullies move beyond the schoolyard: A preliminary look at cyberbullying*. *Youth violence and juvenile justice*, 2006. **4**(2): p. 148-169.

20. Kowalski, R.M., et al., *Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth*. Psychological bulletin, 2014. **140**(4): p. 1073.
21. Cappadocia, M.C., W.M. Craig, and D. Pepler, *Cyberbullying: Prevalence, stability, and risk factors during adolescence*. Canadian Journal of School Psychology, 2013. **28**(2): p. 171-192.
22. Camerini, A.-L., et al., *Cyberbullying perpetration and victimization among children and adolescents: A systematic review of longitudinal studies*. Telematics and Informatics, 2020. **49**: p. 101362.
23. Sharp, S. and P. Smith, *School bullying: Insights and perspectives*. 2002: Routledge.
24. Carlon, A., S. Freedman, and N. Skaar, *FORGIVENESS EDUCATION AND BULLYING PREVENTION AND INTERVENTION: BENEFITS FOR BULLIES, VICTIMS AND SCHOOLS*.
25. Wolke, D., et al., *Impact of bullying in childhood on adult health, wealth, crime, and social outcomes*. Psychological science, 2013. **24**(10): p. 1958-1970.
26. Coyle, S., A. Cipra, and S.Y. Rueger, *Bullying types and roles in early adolescence: Latent classes of perpetrators and victims*. Journal of School Psychology, 2021. **89**: p. 51-71.
27. Haynie, D.L., et al., *Bullies, victims, and bully/victims: Distinct groups of at-risk youth*. The Journal of Early Adolescence, 2001. **21**(1): p. 29-49.
28. Baldry, A.C., D.P. Farrington, and A. Sorrentino, *School bullying and cyberbullying among boys and girls: Roles and overlap*. Journal of Aggression, Maltreatment & Trauma, 2017. **26**(9): p. 937-951.
29. Thomas, H.J., J.P. Connor, and J.G. Scott, *Integrating traditional bullying and cyberbullying: challenges of definition and measurement in adolescents—a review*. Educational psychology review, 2015. **27**(1): p. 135-152.
30. Walters, G.D., *School-age bullying victimization and perpetration: A meta-analysis of prospective studies and research*. Trauma, Violence, & Abuse, 2020: p. 1524838020906513.
31. Bronfenbrenner, U., *The ecology of human development: Experiments by nature and design*. 1979: Harvard university press.
32. Bronfenbrenner, U., *Making human beings human: Bioecological perspectives on human development*. 2005: sage.
33. Bronfenbrenner, U. and P. Morris, *The bioecological model of human development*, in *Handbook of Child Psychology*, R.M. Lerner and W. Damon, Editors. 2006, John Wiley & Sons: Hoboken, New Jersey. p. 793-828.
34. Divecha, D. and M. Brackett, *Rethinking school-based bullying prevention through the lens of social and emotional learning: A bioecological perspective*. International journal of bullying prevention, 2020. **2**(2): p. 93-113.
35. Valle, J.E., L.C. Williams, and A.C. Stelko-Pereira, *Whole-school antibullying interventions: A systematic review of 20 years of publications*. Psychology in the Schools, 2020. **57**(6): p. 868-883.
36. Inchley, J. and D. Currie, *Growing up unequal: gender and socioeconomic differences in young people's health and well-being. Health Behaviour in School-aged Children (HBSC) study: international report from the 2013/2014 survey. WHO, 2016*. Health policy for children and adolescents, 2016(7).
37. Brooks, F., et al., *HBSC England national report 2014*. Hatfield, UK: BMJ Publishing Group, 2015.
38. van Geel, M., et al., *Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism*. Personality and individual differences, 2017. **106**: p. 231-235.
39. Smokowski, P.R. and K.H. Kopasz, *Bullying in school: An overview of types, effects, family characteristics, and intervention strategies*. Children & Schools, 2005. **27**(2): p. 101-110.
40. Wright, M.F., *Predictors of anonymous cyber aggression: The role of adolescents' beliefs about anonymity, aggression, and the permanency of digital content*. Cyberpsychology, Behavior, and Social Networking, 2014. **17**(7): p. 431-438.

41. Fergusson, D.M., L.J. Horwood, and N. Swain-Campbell, *Cannabis use and psychosocial adjustment in adolescence and young adulthood*. *Addiction*, 2002. **97**(9): p. 1123-1135.
42. McGue, M. and W.G. Iacono, *The adolescent origins of substance use disorders*. *International journal of methods in psychiatric research*, 2008. **17**(S1): p. S30-S38.
43. Carlyle, K.E. and K.J. Steinman, *Demographic differences in the prevalence, co-occurrence, and correlates of adolescent bullying at school*. *Journal of School Health*, 2007. **77**(9): p. 623-629.
44. Ttofi, M.M. and D.P. Farrington, *Risk and protective factors, longitudinal research, and bullying prevention*. *New directions for youth development*, 2012. **2012**(133): p. 85-98.
45. Juvonen, J., S. Graham, and M.A. Schuster, *Bullying among young adolescents: The strong, the weak, and the troubled*. *Pediatrics*, 2003. **112**(6): p. 1231-1237.
46. Brimblecombe, N., et al., *Long term economic impact associated with childhood bullying victimisation*. *Social Science & Medicine*, 2018. **208**: p. 134-141.
47. Yeager, D.S., et al., *Declines in efficacy of anti-bullying programs among older adolescents: Theory and a three-level meta-analysis*. *Journal of applied developmental psychology*, 2015. **37**: p. 36-51.
48. Dake, J.A., J.H. Price, and S.K. Telljohann, *The nature and extent of bullying at school*. *Journal of school health*, 2003. **73**(5): p. 173-180.
49. Olafsen, R.N. and V. Viemerö, *Bully/victim problems and coping with stress in school among 10-to 12-year-old pupils in Åland, Finland*. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 2000. **26**(1): p. 57-65.
50. Fujikawa, S., et al., *Bullying Across Late Childhood and Early Adolescence: A Prospective Cohort of Students Assessed Annually From Grades 3 to 8*. *Academic Pediatrics*, 2021. **21**(2): p. 344-351.
51. Goldweber, A., T.E. Waasdorp, and C.P. Bradshaw, *Examining the link between forms of bullying behaviors and perceptions of safety and belonging among secondary school students*. *Journal of school psychology*, 2013. **51**(4): p. 469-485.
52. Sumter, S.R., et al., *Developmental trajectories of peer victimization: Off-line and online experiences during adolescence*. *Journal of Adolescent Health*, 2012. **50**(6): p. 607-613.
53. Zych, I., D.P. Farrington, and M.M. Ttofi, *Protective factors against bullying and cyberbullying: A systematic review of meta-analyses*. *Aggression and Violent Behavior*, 2019. **45**: p. 4-19.
54. Georgiou, S.N. and P. Stavrinides, *Parenting at home and bullying at school*. *Social psychology of education*, 2013. **16**(2): p. 165-179.
55. Nocentini, A., et al., *Parents, family characteristics and bullying behavior: A systematic review*. *Aggression and violent behavior*, 2019. **45**: p. 41-50.
56. Ahmed, E. and V. Braithwaite, *Forgiveness, reconciliation, and shame: Three key variables in reducing school bullying*. *Journal of social issues*, 2006. **62**(2): p. 347-370.
57. Georgiou, S.N., *Bullying and victimization at school: The role of mothers*. *British Journal of Educational Psychology*, 2008. **78**(1): p. 109-125.
58. Jansen, D.E., et al., *Early risk factors for being a bully, victim, or bully/victim in late elementary and early secondary education. The longitudinal TRAILS study*. *BMC public health*, 2011. **11**(1): p. 440.
59. Bevilacqua, L., et al., *The role of family and school-level factors in bullying and cyberbullying: a cross-sectional study*. *BMC pediatrics*, 2017. **17**(1): p. 160.
60. Sellström, E. and S. Bremberg, *Is there a "school effect" on pupil outcomes? A review of multilevel studies*. *Journal of Epidemiology & Community Health*, 2006. **60**(2): p. 149-155.
61. Fletcher, A., et al., *Cannabis use and 'safe' identities in an inner-city school risk environment*. *International Journal of Drug Policy*, 2009. **20**(3): p. 244-250.
62. Jamal, F., et al., *The school environment and student health: a systematic review and meta-ethnography of qualitative research*. *BMC Public Health*, 2013. **13**(1): p. 798.

63. Bonell, C., et al., *Theories of how the school environment impacts on student health: Systematic review and synthesis*. Health & place, 2013. **24**: p. 242-249.
64. Paille, B., *Toxic schools: high-poverty education in New York and Amsterdam*. 2013: University of Chicago Press.
65. Dance, L.J., *Tough fronts: The impact of street culture on schooling*. 2012: Routledge.
66. Viner, R.M., et al., *Adolescence and the social determinants of health*. The Lancet, 2012. **379**(9826): p. 1641-1652.
67. Flay, B.R., et al., *Effects of 2 prevention programs on high-risk behaviors among African American youth: a randomized trial*. Archives of Pediatrics & Adolescent Medicine, 2004. **158**(4): p. 377-384.
68. Resnick, M.D., et al., *Protecting adolescents from harm: findings from the National Longitudinal Study on Adolescent Health*. Jama, 1997. **278**(10): p. 823-832.
69. Patton, G.C., et al., *Our future: a Lancet commission on adolescent health and wellbeing*. The Lancet, 2016.
70. Jain, S., et al., *School climate and physical adolescent relationship abuse: differences by sex, socioeconomic status, and bullying*. Journal of adolescence, 2018. **66**: p. 71-82.
71. Wentzel, K.R., C.M. Barry, and K.A. Caldwell, *Friendships in Middle School: Influences on Motivation and School Adjustment*. Journal of educational psychology, 2004. **96**(2): p. 195.
72. Wentzel, K.R., *Relations of social goal pursuit to social acceptance, classroom behavior, and perceived social support*. Journal of Educational Psychology, 1994. **86**(2): p. 173.
73. Wentzel, K.R., *Social relationships and motivation in middle school: The role of parents, teachers, and peers*. Journal of educational psychology, 1998. **90**(2): p. 202.
74. Van Ryzin, M.J., A.A. Gravely, and C.J. Roseth, *Autonomy, belongingness, and engagement in school as contributors to adolescent psychological well-being*. Journal of youth and adolescence, 2009. **38**(1): p. 1-12.
75. Vieno, A., et al., *Social support, sense of community in school, and self-efficacy as resources during early adolescence: an integrative model*. American Journal of Community Psychology, 2007. **39**(1-2): p. 177-190.
76. Bond, L., et al., *Social and school connectedness in early secondary school as predictors of late teenage substance use, mental health, and academic outcomes*. Journal of Adolescent Health, 2007. **40**(4): p. 357. e9-357. e18.
77. Rasmussen, M., et al., *School connectedness and daily smoking among boys and girls: the influence of parental smoking norms*. The European Journal of Public Health, 2005. **15**(6): p. 607-612.
78. Johnson, S.L., *Improving the school environment to reduce school violence: A review of the literature*. Journal of school health, 2009. **79**(10): p. 451-465.
79. Shackleton, N., et al., *Systematic review of reviews of observational studies of school-level effects on sexual health, violence and substance use*. Health & place, 2016. **39**: p. 168-176.
80. Bonell, C., et al., *The effects on student health of interventions modifying the school environment: systematic review*. Journal of epidemiology and community health, 2013. **67**(8): p. 677-681.
81. Markham, W.A., et al., *Does school ethos explain the relationship between value-added education and teenage substance use? A cohort study*. Social science & medicine, 2012. **75**(1): p. 69-76.
82. Tobler, A.L., et al., *Preventing the link between SES and high-risk behaviors: "Value-added" education, drug use and delinquency in high-risk, urban schools*. Prevention Science, 2011. **12**(2): p. 211-221.
83. Bonell, C., et al., *The effects of the school environment on student health: a systematic review of multi-level studies*. Health & place, 2013. **21**: p. 180-191.

84. Bond, L., et al., *The Gatehouse Project: can a multilevel school intervention affect emotional wellbeing and health risk behaviours?* Journal of epidemiology and community health, 2004. **58**(12): p. 997-1003.
85. Bond, L., et al., *Building capacity for system-level change in schools: lessons from the Gatehouse Project.* Health Education & Behavior, 2001. **28**(3): p. 368-383.
86. Bond, L., et al., *Long-Term Impact of the Gatehouse Project on Cannabis Use of 16-Year-Olds in Australia.* Journal of School Health, 2004. **74**(1): p. 23-29.
87. Patton, G.C., et al., *Promoting social inclusion in schools: a group-randomized trial of effects on student health risk behavior and well-being.* American journal of public health, 2006. **96**(9): p. 1582-1587.
88. Hale, D.R. and R.M. Viner, *The correlates and course of multiple health risk behaviour in adolescence.* BMC public health, 2016. **16**(1): p. 1.
89. Moffitt, T.E., *Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy.* Psychological review, 1993. **100**(4): p. 674.
90. Larson, R.W., et al., *Changes in adolescents' daily interactions with their families from ages 10 to 18: Disengagement and transformation.* Developmental Psychology, 1996. **32**(4): p. 744.
91. Steer, A., *learning behavior: lessons learned a review of behavior standards and practices in out schools*, S.a.F. Department for Children, Editor. 2009: London.
92. Moore, S.E., et al., *Consequences of bullying victimization in childhood and adolescence: A systematic review and meta-analysis.* World journal of psychiatry, 2017. **7**(1): p. 60.
93. Hair, E.C., et al., *Risky behaviors in late adolescence: co-occurrence, predictors, and consequences.* Journal of Adolescent Health, 2009. **45**(3): p. 253-261.
94. Fisher, H.L., et al., *Bullying victimisation and risk of self harm in early adolescence: longitudinal cohort study.* bmj, 2012. **344**: p. e2683.
95. Winsper, C., et al., *Involvement in bullying and suicide-related behavior at 11 years: a prospective birth cohort study.* Journal of the American Academy of Child & Adolescent Psychiatry, 2012. **51**(3): p. 271-282. e3.
96. Cosma, A., et al., *Trends in bullying victimization in Scottish adolescents 1994–2014: changing associations with mental well-being.* International journal of public health, 2017. **62**(6): p. 639-646.
97. Moore, S.E., et al., *Adolescent peer aggression and its association with mental health and substance use in an Australian cohort.* Journal of adolescence, 2014. **37**(1): p. 11-21.
98. Calvete, E., I. Orue, and M. Gámez-Guadix, *Cyberbullying victimization and depression in adolescents: The mediating role of body image and cognitive schemas in a one-year prospective study.* European Journal on Criminal Policy and Research, 2016. **22**(2): p. 271-284.
99. Cole, D.A., et al., *Longitudinal and incremental relation of cybervictimization to negative self-cognitions and depressive symptoms in young adolescents.* Journal of abnormal child psychology, 2016. **44**(7): p. 1321-1332.
100. Espinoza, G., *Daily cybervictimization among Latino adolescents: Links with emotional, physical and school adjustment.* Journal of Applied Developmental Psychology, 2015. **38**: p. 39-48.
101. Bonanno, R.A. and S. Hymel, *Cyber bullying and internalizing difficulties: Above and beyond the impact of traditional forms of bullying.* Journal of youth and adolescence, 2013. **42**(5): p. 685-697.
102. Smokowski, P.R., C.B. Evans, and K.L. Cotter, *The differential impacts of episodic, chronic, and cumulative physical bullying and cyberbullying: The effects of victimization on the school experiences, social support, and mental health of rural adolescents.* Violence and Victims, 2014. **29**(6): p. 1029-1046.

103. Nansel, T.R., et al., *Cross-national consistency in the relationship between bullying behaviors and psychosocial adjustment*. Archives of Pediatrics & Adolescent Medicine, 2004. **158**(8): p. 730-736.
104. Wolke, D., et al., *Bullying and victimization of primary school children in England and Germany: Prevalence and school factors*. British Journal of Psychology, 2001. **92**(4): p. 673-696.
105. Nansel, T.R., et al., *Relationships between bullying and violence among US youth*. Archives of Pediatrics & Adolescent Medicine, 2003. **157**(4): p. 348-353.
106. Radliff, K.M., et al., *Illuminating the relationship between bullying and substance use among middle and high school youth*. Addictive behaviors, 2012. **37**(4): p. 569-572.
107. Tharp-Taylor, S., A. Haviland, and E.J. D'Amico, *Victimization from mental and physical bullying and substance use in early adolescence*. Addictive behaviors, 2009. **34**(6-7): p. 561-567.
108. Fletcher, A., et al., *How might schools influence young people's drug use? Development of theory from qualitative case-study research*. Journal of Adolescent Health, 2009. **45**(2): p. 126-132.
109. Mishna, F., et al., *Risk factors for involvement in cyber bullying: Victims, bullies and bully-victims*. Children and Youth Services Review, 2012. **34**(1): p. 63-70.
110. Demaray, M.K. and C.K. Malecki, *Perceptions of the frequency and importance of social support by students classified as victims, bullies, and bully/victims in an urban middle school*. School Psychology Review, 2003. **32**(3): p. 471-490.
111. Jamal, F., et al., *The social ecology of girls' bullying practices: exploratory research in two London schools*. Sociology of health & illness, 2015. **37**(5): p. 731-744.
112. González-Cabrera, J., et al., *Relationship between cyberbullying roles, cortisol secretion and psychological stress*. Computers in Human Behavior, 2017. **70**: p. 153-160.
113. Zych, I., et al., *School bullying and dating violence in adolescents: A systematic review and meta-analysis*. Trauma, Violence, & Abuse, 2021. **22**(2): p. 397-412.
114. Boxer, P., et al., *Exposure to low-level aggression in schools associations with aggressive behaviour, future expectation, and perceived safety*. Violence and victims, 2003. **18**(6): p. 691-705.
115. Kochenderfer, B.J. and G.W. Ladd, *Peer victimization: Cause or consequence of school maladjustment?* Child development, 1996. **67**(4): p. 1305-1317.
116. Nansel, T.R., et al., *Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment*. Jama, 2001. **285**(16): p. 2094-2100.
117. Glew, G.M., et al., *Bullying, psychosocial adjustment, and academic performance in elementary school*. Archives of pediatrics & adolescent medicine, 2005. **159**(11): p. 1026-1031.
118. Yang, J., X. Wang, and L. Lei, *Perceived school climate and adolescents' bullying perpetration: A moderated mediation model of moral disengagement and peers' defending*. Children and youth services review, 2020. **109**: p. 104716.
119. Price, L.H., et al., *Telomeres and early-life stress: an overview*. Biological psychiatry, 2013. **73**(1): p. 15-23.
120. Vreeman, R.C. and A.E. Carroll, *A systematic review of school-based interventions to prevent bullying*. Archives of Pediatrics & Adolescent Medicine, 2007. **161**(1): p. 78-88.
121. Copeland, W.E., et al., *Adult psychiatric outcomes of bullying and being bullied by peers in childhood and adolescence*. JAMA psychiatry, 2013. **70**(4): p. 419-426.
122. Lereya, S.T., et al., *Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries*. The Lancet Psychiatry, 2015. **2**(6): p. 524-531.
123. Beinar, S., et al., *Youth at Risk?: A National Survey of Risk Factors, Protective Factors and Problem Behaviour Among Young People in England, Scotland and Wales*. 2002: Communities that Care London.

124. Jantzer, V., et al., *The cost incurred by victims of bullying from a societal perspective: estimates based on a German online survey of adolescents*. European child & adolescent psychiatry, 2019. **28**(4): p. 585-594.
125. Takizawa, R., B. Maughan, and L. Arseneault, *Adult health outcomes of childhood bullying victimization: evidence from a five-decade longitudinal British birth cohort*. American journal of psychiatry, 2014. **171**(7): p. 777-784.
126. DCSF, *Your child, your schools, our future: building a 21st century schools system*. 2009, London: TSO.
127. Health, D.o., *Healthy Lives, Brighter Futures—The Strategy for Children's and Younger People's Health*. 2009, DH London.
128. Durlak, J.A., et al., *The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions*. Child development, 2011. **82**(1): p. 405-432.
129. Bonell, C., et al., *Systematic review of the effects of schools and school environment interventions on health: evidence mapping and synthesis*. Public Health Research, 2013. **1**(1).
130. Shackleton, N., et al., *School-Based Interventions Going Beyond Health Education to Promote Adolescent Health: Systematic Review of Reviews*. Journal of Adolescent Health, 2016. **58**(4): p. 382-396.
131. Smith, J.D., et al., *The effectiveness of whole-school antibullying programs: A synthesis of evaluation research*. School psychology review, 2004. **33**(4): p. 547.
132. Ttofi, M.M. and D.P. Farrington, *Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review*. Journal of Experimental Criminology, 2011. **7**(1): p. 27-56.
133. Wells, J., J. Barlow, and S. Stewart-Brown, *A systematic review of universal approaches to mental health promotion in schools*. Health Education, 2003. **103**(4): p. 197-220.
134. Lipsey, M.W. and D.B. Wilson, *Practical meta-analysis*. 2001: Sage Publications, Inc.
135. Olweus, D. and S.P. Limber, *The Olweus Bullying Prevention Program: Implementation and Evaluation over Two Decades*, in *Handbook of bullying in schools*. 2009, Routledge. p. 387-412.
136. Ng, E.D., J.Y.X. Chua, and S. Shorey, *The effectiveness of educational interventions on traditional bullying and cyberbullying among adolescents: A systematic review and meta-analysis*. Trauma, Violence, & Abuse, 2020: p. 1524838020933867.
137. Gaffney, H., D.P. Farrington, and M.M. Ttofi, *Examining the effectiveness of school-bullying intervention programs globally: A meta-analysis*. International Journal of Bullying Prevention, 2019. **1**(1): p. 14-31.
138. Taylor, R.D., et al., *Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects*. Child development, 2017. **88**(4): p. 1156-1171.
139. Silva, J.L.d., et al., *Results from interventions addressing social skills to reduce school bullying: a systematic review with meta-analysis*. Trends in Psychology, 2018. **26**: p. 509-522.
140. Sutton, J., P.K. Smith, and J. Swettenham, *Bullying and 'theory of mind': A critique of the 'social skills deficit' view of anti-social behaviour*. Social development, 1999. **8**(1): p. 117-127.
141. Fonagy, P., et al., *A cluster randomized controlled trial of child-focused psychiatric consultation and a school systems-focused intervention to reduce aggression*. Journal of Child Psychology and Psychiatry, 2009. **50**(5): p. 607-616.
142. Wigelsworth, M., N. Humphrey, and A. Lendrum, *Evaluation of a school-wide preventive intervention for adolescents: The secondary social and emotional aspects of learning (SEAL) programme*. School Mental Health, 2013. **5**(2): p. 96-109.
143. Brooks, F., *The link between pupil health and wellbeing and attainment: a briefing for head teachers, governors and staff in education settings: November 2014*. 2014.

144. Blank, L., et al., *Promoting well-being by changing behaviour: A systematic review and narrative synthesis of the effectiveness of whole secondary school behavioural interventions*. *Mental Health Review Journal*, 2010. **15**(2): p. 43-53.
145. Hallam, S., J. Rhamie, and J. Shaw, *Evaluation of the primary behaviour and attendance pilot*. 2006: Department for Education and Skills London,, UK.
146. Challen, A., et al., *UK resilience programme evaluation*. 2011.
147. Van Ness, D.W. and K.H. Strong, *Restoring justice: An introduction to restorative justice*. 2014: Routledge.
148. Latimer, J., C. Dowden, and D. Muise, *The effectiveness of restorative justice practices: A meta-analysis*. *The prison journal*, 2005. **85**(2): p. 127-144.
149. Morris, A. and G. Maxwell, *Restorative justice for juveniles: Conferencing, mediation and circles*. 2001: Bloomsbury Publishing.
150. Wallis, P., *Understanding restorative justice: How empathy can close the gap created by crime*. 2014: Policy Press.
151. Braithwaite, J. *Restorative justice: Theories and worries*. in *Visiting Experts' Papers: 123rd International Senior Seminar, Resource Material Series*. 2004.
152. Van Dyke, C.J. and M.J. Elias, *How forgiveness, purpose, and religiosity are related to the mental health and well-being of youth: A review of the literature*. *Mental Health, Religion and Culture*, 2007. **10**(4): p. 395-415.
153. Worthington, E.L. and M. Scherer, *Forgiveness is an emotion-focused coping strategy that can reduce health risks and promote health resilience: Theory, review, and hypotheses*. *Psychology & Health*, 2004. **19**(3): p. 385-405.
154. van der Valk, A., *There Are No Bullies: Just Children Who Bully And You Can Help Them*. *The Education Digest*, 2014. **79**(8): p. 39.
155. Kokotsaki, D., C. White, and B. Hopkins, *Capturing change: a review of the implementation of Restorative Approaches and its outcomes within a local authority in North East England*. *Online educational research journal.*, 2014. **5**(5): p. 151.
156. Acosta, J.D., et al., *A cluster-randomized trial of restorative practices: An illustration to spur high-quality research and evaluation*. *Journal of Educational and Psychological Consultation*, 2016. **26**(4): p. 413-430.
157. Hopkins, B., *Just schools. A whole school approach to restorative justice* (London, Jessica Kingsley), 2004.
158. Acosta, J., et al., *Evaluation of a Whole-School Change Intervention: Findings from a Two-Year Cluster-Randomized Trial of the Restorative Practices Intervention*. *Journal of youth and adolescence*, 2019. **48**(5): p. 876-890.
159. Tyler, T.R., *Restorative justice and procedural justice: Dealing with rule breaking*. *Journal of social issues*, 2006. **62**(2): p. 307.
160. Darling-Hammond, S., et al., *Effectiveness of restorative justice in US K-12 schools: A review of quantitative research*. *Contemporary School Psychology*, 2020. **24**: p. 295-308.
161. Karp, D.R. and B. Breslin, *Restorative justice in school communities*. *Youth & Society*, 2001. **33**(2): p. 249-272.
162. Anfara Jr, V.A., K.R. Evans, and J.N. Lester, *Restorative justice in education: What we know so far*. *Middle School Journal*, 2013. **44**(5): p. 57-63.
163. Song, S.Y. and S.M. Swearer, *The cart before the horse: The challenge and promise of restorative justice consultation in schools*. 2016, Taylor & Francis.
164. Sumner, M.D., C.J. Silverman, and M.L. Frampton, *School-based restorative justice as an alternative to zero-tolerance policies: Lessons from West Oakland*. Thelton E. Henderson Center for Social Justice, 2010.
165. Wong, D.S., et al., *Program effectiveness of a restorative whole-school approach for tackling school bullying in Hong Kong*. *International journal of offender therapy and comparative criminology*, 2011. **55**(6): p. 846-862.

166. Mas-Expósito, L., et al., *Implementation of Whole School Restorative Approaches to Promote Positive Youth Development: Review of Relevant Literature and Practice Guidelines*. Education Sciences, 2022. **12**(187).
167. Buckley, S. and G.M. Maxwell, *Respectful schools: Restorative practices in education: A summary report*. 2007: Office of the Children's Commissioner and the Institute of Policy Studies, School of Government, Victoria University, Wellington.
168. Learning, M.D.o.C.a.F., *A three-year evaluation of alternative approaches to suspensions and expulsions: a report to the Legislature*. 1998.
169. Board, Y.J., *National evaluation of the restorative justice in schools programme*. London: Youth Justice Board. Whose Justice, 2004.
170. Skinnis, L. and M. Hough, *An Evaluation of Bristol RAiS*. London: ICPR, King's College London, 2009.
171. Kane, J., et al., *Restorative practices in Scottish schools*. Edinburgh: Scottish Executive, 2007.
172. Moore, G.F., et al., *School composition, school culture and socioeconomic inequalities in young people's health: Multi-level analysis of the Health Behaviour in School-aged Children (HBSC) survey in Wales*. British educational research journal, 2017. **43**(2): p. 310-329.
173. Dahlgren, G. and M. Whitehead, *Policies and strategies to promote social equity in health*. Stockholm: Institute for future studies, 1991.
174. Fletcher, A., et al., *Involving young people in changing their school environment to make it safer: findings from a process evaluation in english secondary schools*. Health Education, 2015. **115**(3/4): p. 322-338.
175. Cross, D., et al., *The Friendly Schools Friendly Families programme: Three-year bullying behaviour outcomes in primary school children*. International Journal of Educational Research, 2012. **53**: p. 394-406.
176. Langford, R., et al., *The WHO Health Promoting School framework for improving the health and well-being of students and their academic achievement*. The Cochrane Library, 2014.
177. Fraguas, D., et al., *Assessment of school anti-bullying interventions: a meta-analysis of randomized clinical trials*. JAMA pediatrics, 2021. **175**(1): p. 44-55.
178. Ponsford, R., et al., *Whole-school interventions promoting student commitment to school to prevent substance use and violence and improve educational attainment (In press)*. 2021.
179. Langford, R., et al., *The World Health Organization's Health Promoting Schools framework: a Cochrane systematic review and meta-analysis*. BMC public health, 2015. **15**(1): p. 130.
180. Gaffney, H., M.M. Ttofi, and D.P. Farrington, *Effectiveness of school-based programs to reduce bullying perpetration and victimization: An updated systematic review and meta-analysis*. Campbell Systematic Reviews, 2021. **17**(2): p. e1143.
181. Riestenberg, N., *In-school behavior intervention grants: A three-year evaluation of alternative approaches to suspensions and expulsions*. Roseville, MN: Minnesota Dept. of Children, Families & Learning, 2001.
182. Shaw, G., *Restorative practices in Australian schools: Changing relationships, changing culture*. Conflict Resolution Quarterly, 2007. **25**(1): p. 127-135.
183. Farrington, D.P. and M.M. Ttofi, *School-based programs to reduce bullying and victimization*. The Campbell Collaboration, 2009. **6**: p. 1-149.
184. Durlak, J.A. and E.P. DuPre, *Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation*. American journal of community psychology, 2008. **41**(3-4): p. 327-350.
185. Smith, P.K., C. Salmivalli, and H. Cowie, *Effectiveness of school-based programs to reduce bullying: A commentary*. Journal of Experimental Criminology, 2012. **8**(4): p. 433-441.
186. Copeland, W.E., et al., *Childhood and adolescent psychiatric disorders as predictors of young adult disorders*. Archives of general psychiatry, 2009. **66**(7): p. 764-772.

187. Segawa, E., et al., *Evaluation of the effects of the Aban Aya Youth Project in reducing violence among African American adolescent males using latent class growth mixture modeling techniques*. Evaluation review, 2005. **29**(2): p. 128-148.
188. Baron, R.M. and D.A. Kenny, *The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations*. Journal of personality and social psychology, 1986. **51**(6): p. 1173.
189. Gardner, F., J. Burton, and I. Klimes, *Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: outcomes and mechanisms of change*. Journal of child psychology and psychiatry, 2006. **47**(11): p. 1123-1132.
190. Garandeau, C.F. and C. Salmivalli, *Can healthier contexts be harmful? A new perspective on the plight of victims of bullying*. Child Development Perspectives, 2019. **13**(3): p. 147-152.
191. Salmivalli, C., *Peer victimization and adjustment in young adulthood: commentary on the special section*. Journal of abnormal child psychology, 2018. **46**(1): p. 67-72.
192. Huitsing, G., et al., *The healthy context paradox: victims' adjustment during an anti-bullying intervention*. Journal of Child and Family Studies, 2019. **28**(9): p. 2499-2509.
193. Melendez-Torres, G.J., et al., *Locating and testing the healthy context paradox: examples from the INCLUSIVE trial (In press)*. BMC Medical Research Methodology, 2022.
194. Patton, G.C., et al., *The Gatehouse Project: a systematic approach to mental health promotion in secondary schools*. Australian & New Zealand Journal of Psychiatry, 2000. **34**(4): p. 586-593.
195. Moore, G.F., et al., *Process evaluation of complex interventions: Medical Research Council guidance*. bmj, 2015. **350**: p. h1258.
196. Rutter, H., et al., *The need for a complex systems model of evidence for public health*. The Lancet, 2017. **390**(10112): p. 2602-2604.
197. Craig, P., et al., *Developing and evaluating complex interventions*. Medical Research Council, UK, 2011.
198. Saarento, S., A.J. Boulton, and C. Salmivalli, *Reducing bullying and victimization: Student-and classroom-level mechanisms of change*. Journal of abnormal child psychology, 2015. **43**(1): p. 61-76.
199. Hanckel, B., et al., *The use of Qualitative Comparative Analysis (QCA) to address causality in complex systems: a systematic review of research on public health interventions*. BMC public health, 2021. **21**(1): p. 1-22.
200. Keshavarz, N., et al., *Schools as social complex adaptive systems: a new way to understand the challenges of introducing the health promoting schools concept*. Social science & medicine, 2010. **70**(10): p. 1467-1474.
201. Popper, K.R., *Objective knowledge: An evolutionary approach*. 1972.
202. Bonell, C., G. Melendez-Torres, and S. Quilley, *The potential role for sociologists in designing RCTs and of RCTs in refining sociological theory: A commentary on Deaton and Cartwright*. Social science & medicine (1982), 2018.
203. Moore, G.F. and R.E. Evans, *What theory, for whom and in which context? Reflections on the application of theory in the development and evaluation of complex population health interventions*. SSM-population health, 2017. **3**: p. 132-135.
204. Merton, R.K., *On sociological theories of the middle range [1949]*. 1949: na.
205. Breuer, E., et al., *Using theory of change to design and evaluate public health interventions: a systematic review*. Implementation Science, 2015. **11**(1): p. 1-17.
206. Shearn, K., et al., *Building realist program theory for large complex and messy interventions*. International Journal of Qualitative Methods, 2017. **16**(1): p. 1609406917741796.
207. Hill, A.B., *The clinical trial*. British Medical Bulletin, 1951. **7**(4): p. 278-282.
208. Services, M.R.C.H. and P.H.R. Board, *A framework for development and evaluation of RCTs for complex interventions to improve health*. 2000: Medical Research Council.

209. Chalmers, I., *Fisher and Bradford Hill: theory and pragmatism?* International journal of epidemiology, 2003. **32**(6): p. 922-924.
210. Evans, J., *Epidemiology in Practice: Randomised Controlled Trials*. Community eye health, 1998. **11**(26): p. 26.
211. Oakley, A., *Who's afraid of the randomized controlled trial? Some dilemmas of the scientific method and 'good' research practice*. 2004: Routledge.
212. O' Cathain, A., E. Murphy, and J. Nicholl, *Three techniques for integrating data in mixed methods studies*. Bmj, 2010. **341**: p. c4587.
213. Yin, R.K., *Mixed methods research: Are the methods genuinely integrated or merely parallel*. Research in the Schools, 2006. **13**(1): p. 41-47.
214. Oakley, A., et al., *Process evaluation in randomised controlled trials of complex interventions*. BMJ: British Medical Journal, 2006. **332**(7538): p. 413.
215. Evans, R., J. Scourfield, and S. Murphy, *Pragmatic, formative process evaluations of complex interventions and why we need more of them*. 2015, BMJ Publishing Group Ltd.
216. Moore, G., et al., *Process evaluation of complex interventions*, Medical Research Council, Editor. 2015.
217. Oakley, A., et al., *Evaluating Processes: A Case Study of a Randomized Controlled Trial of Sex Education*. Evaluation, 2004. **10**(4): p. 440-462.
218. Suchman, E., *Evaluative Research: Principles and Practice in Public Service and Social Action Progr*. 1968: Russell Sage Foundation.
219. Steckler, A.B., L. Linnan, and B. Israel, *Process evaluation for public health interventions and research*. 2002: Jossey-Bass San Francisco, CA.
220. Strange, V., et al., *Integrating process with outcome data in a randomized controlled trial of sex education*. Evaluation, 2006. **12**(3): p. 330-352.
221. Pocock, S.J. and M. Abdalla, *The hope and the hazards of using compliance data in randomized controlled trials*. Statistics in medicine, 1998. **17**(3): p. 303-317.
222. Parry-Langdon, N., et al., *Process evaluation of health promotion interventions*. Policy & Politics, 2003. **31**(2): p. 207-216.
223. Wight, D. and A. Obasi, *Unpacking the black box: the importance of process data to explain outcomes*. 2003.
224. Filstead, W.J., *Qualitative methods: A needed perspective in evaluation research*. Qualitative and quantitative methods in evaluation research, 1979: p. 33-48.
225. Stephenson, J., et al., *The long-term effects of a peer-led sex education programme (RIPPLE): a cluster randomised trial in schools in England*. PLoS Med, 2008. **5**(11): p. e224.
226. Patton, M.Q., *A context and boundaries for a theory-driven approach to validity*. Evaluation and Program Planning, 1989. **12**(4): p. 375-377.
227. Stame, N., *Theory-based evaluation and types of complexity*. Evaluation, 2004. **10**(1): p. 58-76.
228. Perry, C.L., et al., *The Child and Adolescent Trial for Cardiovascular Health (CATCH): intervention, implementation, and feasibility for elementary schools in the United States*. Health Education & Behavior, 1997. **24**(6): p. 716-735.
229. Baranowski, T., C. Anderson, and C. Carmack, *Mediating variable framework in physical activity interventions: How are we doing? How might we do better?* American journal of preventive medicine, 1998. **15**(4): p. 266-297.
230. Grant, A., et al., *Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting*. Trials, 2013. **14**(1): p. 15.
231. Sherman, L.W., J.D. Schmidt, and D.P. Rogan, *Policing domestic violence: Experiments and dilemmas*. 1992: Free Press.
232. Tilley, N. *Realistic evaluation: an overview*. in *founding conference of the Danish Evaluation Society*. 2000.
233. Pawson, R. and N. Tilley, *Realist evaluation*. 2004, London: Sage Publications Google Scholar.

234. Porter, S., T. McConnell, and J. Reid, *The possibility of critical realist randomised controlled trials*. *Trials*, 2017. **18**(1): p. 133.
235. Hinds, K. and K. Dickson, *Realist synthesis: a critique and an alternative*. *Journal of Critical Realism*, 2021. **20**(1): p. 1-17.
236. Porter, S., *The uncritical realism of realist evaluation*. *Evaluation*, 2015. **21**(1): p. 65-82.
237. Bhaskar, R., *The possibility of naturalism: A philosophical critique of the contemporary human sciences*. 2014: Routledge.
238. King, A., *The odd couple: Margaret Archer, Anthony Giddens and British social theory*. *The British journal of sociology*, 2010. **61**: p. 253-260.
239. Archer, M.S., *Realist social theory: The morphogenetic approach*. 1995: Cambridge university press.
240. Giddens, A., *The constitution of society: Outline of the theory of structuration*. 1984: Univ of California Press.
241. Bhaskar, R., *A Realist Theory of Science*. 1975: Verso.
242. Lawson, T., *Economics and reality*. 1997: Routledge.
243. Dalkin, S.M., et al., *What's in a mechanism? Development of a key concept in realist evaluation*. *Implementation Science*, 2015. **10**(1): p. 49.
244. Lemire, S., et al., *What is this thing called a mechanism? Findings from a review of realist evaluations*. *New Directions for Evaluation*, 2020. **2020**(167): p. 73-86.
245. Porter, S., *Realist evaluation: an immanent critique*. *Nursing Philosophy*, 2015. **16**(4): p. 239-251.
246. Dalkin, S.M., et al., *Reducing inequalities in care for patients with non-malignant diseases: Insights from a realist evaluation of an integrated palliative care pathway*. *Palliative medicine*, 2016. **30**(7): p. 690-697.
247. Ford, J.A., et al., *Access to primary care for socio-economically disadvantaged older people in rural areas: exploring realist theory using structural equation modelling in a linked dataset*. *BMC medical research methodology*, 2018. **18**(1): p. 57.
248. Ravn, R., *Testing mechanisms in large-N realistic evaluations*. *Evaluation*, 2019. **25**(2): p. 171-188.
249. Greenhalgh, T., et al., *How do you modernize a health service? A realist evaluation of whole-scale transformation in London*. *The Milbank Quarterly*, 2009. **87**(2): p. 391-416.
250. Hawkins, A.J., *Realist evaluation and randomised controlled trials for testing program theory in complex social systems*. *Evaluation*, 2016. **22**(3): p. 270-285.
251. Mirzoev, T., et al., *Study protocol: realist evaluation of effectiveness and sustainability of a community health workers programme in improving maternal and child health in Nigeria*. *Implementation Science*, 2016. **11**(1): p. 83.
252. Gilmore, B., et al., *Improving the performance of community health workers in humanitarian emergencies: a realist evaluation protocol for the PIECES programme*. *BMJ open*, 2016. **6**(8).
253. Farrell, G., A. Tseloni, and N. Tilley, *Signature dish: Triangulation from data signatures to examine the role of security in falling crime*. *Methodological Innovations*, 2016. **9**: p. 2059799115622754.
254. Porter, S., et al., *A critical realist evaluation of a music therapy intervention in palliative care*. *BMC Palliative Care*, 2017. **16**(1): p. 70.
255. McConnell, T., et al., *Evaluation of the effectiveness of music therapy in improving the quality of life of palliative care patients: a randomised controlled pilot and feasibility study*. *Pilot and feasibility studies*, 2016. **2**(1): p. 1-8.
256. Randell, R., et al., *A realist process evaluation of robot-assisted surgery: integration into routine practice and impacts on communication, collaboration and decision-making*. *Health Services and Delivery Research*, 2017. **5**(20).

257. Rycroft-Malone, J., et al., *A realist process evaluation within the Facilitating Implementation of Research Evidence (FIRE) cluster randomised controlled international trial: an exemplar*. Implementation Science, 2018. **13**(1): p. 138.
258. Byng, R., et al., *Exposing the key functions of a complex intervention for shared care in mental health: case study of a process evaluation*. BMC Health Services Research, 2008. **8**(1): p. 274.
259. Bonell, C., et al., *Are randomised controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services*. Trials, 2018. **19**(1): p. 238.
260. Green, J. and N. Thorogood, *Qualitative methods for health research*. 2013: Sage.
261. Pearce, W. and S. Raman, *The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance*. Policy sciences, 2014. **47**(4): p. 387-402.
262. Rowe, M. and C. Oltmann, *Randomised controlled trials in educational research: Ontological and epistemological limitations*. African Journal of Health Professions Education, 2016. **8**(1): p. 6-8.
263. Greenhalgh, J. and A. Manzano, *Understanding 'context' in realist evaluation and synthesis*. International Journal of Social Research Methodology, 2021: p. 1-13.
264. Hume, D., *1739. A treatise of human nature*. London: John Noon, 1978.
265. Perez, C.C., *Invisible women: Exposing data bias in a world designed for men*. 2019: Random House.
266. Salas, X.R., *The ineffectiveness and unintended consequences of the public health war on obesity*. Canadian Journal of Public Health, 2015. **106**(2): p. E79.
267. Mahoney, J. *Beyond correlational analysis: Recent innovations in theory and method*. in *Sociological forum*. 2001. JSTOR.
268. Chen, H.-T. and P.H. Rossi, *The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science*. Social forces, 1980. **59**(1): p. 106-122.
269. Chen, H.-T. and P.H. Rossi, *Evaluating with sense: The theory-driven approach*. Evaluation review, 1983. **7**(3): p. 283-302.
270. Jamal, F., et al., *The three stages of building and testing mid-level theories in a realist RCT: a theoretical and methodological case-example*. Trials, 2015. **16**(1): p. 1.
271. Bonell, C., et al., *Methods don't make assumptions, researchers do: a response to Marchal et al*. Social science & medicine (1982), 2013. **94**: p. 81-82.
272. Bonell, C., et al., *Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al*. Trials, 2016. **17**(1): p. 478.
273. Harachi, T.W., et al., *Opening the black box: using process evaluation measures to assess implementation and theory building*. American journal of community psychology, 1999. **27**(5): p. 711-731.
274. Chen, H.T., *Theory-driven evaluations*. 1990: Sage.
275. Sayer, A., *Realism and social science*. 1999: Sage.
276. Popper, K., *The Poverty of Historicism*. Economica, 1944. **11**(42): p. 86-103.
277. Hill, A.B., *The environment and disease: association or causation?* 1965, Sage Publications.
278. Bonell, C., et al., *Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial*. The Lancet, 2018. **392**(10163): p. 2452-2464.
279. Fletcher, A., et al., *Realist complex intervention science: Applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions*. Evaluation, 2016. **22**(3): p. 286-303.
280. Banerjee, A.V. and E. Duflo, *The experimental approach to development economics*. Annu. Rev. Econ., 2009. **1**(1): p. 151-178.
281. Van Bavel, J.J., et al., *Contextual sensitivity in scientific reproducibility*. Proceedings of the National Academy of Sciences, 2016. **113**(23): p. 6454-6459.

282. Kerr, N.L., et al., *Addressing replicability concerns via adversarial collaboration: Discovering hidden moderators of the minimal intergroup discrimination effect*. Journal of Experimental Social Psychology, 2018. **78**: p. 66-76.
283. Go, J.C., *Bite-Sized lecture 05: The Critical Realist Theory of Ontological Emergence*, C.R.-A.P. Network, Editor. 2019.
284. Warren, E., et al., *Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention*. Trials, 2020. **21**(1): p. 1-14.
285. Nussbaum, M., *Aristotelian social democracy*, in *Liberalism and the Good*, G.M. Mara and H.S. Richardson, Editors. 1990, Routledge: London. p. 203-251.
286. Erikson, E.H., *Identity: Youth and crisis*. 1968: WW Norton & company.
287. Bernstein, B., *Class, codes and control: Applied studies towards a sociology of language*. Vol. 2. 2003: Psychology Press.
288. Bernstein, B., *Class, codes and control. Vol 3: Toward a theory of educational transmission*. 1975, Londres: Routledge and Kegan Paul.
289. Bond, L., et al., *A comparison of the Gatehouse Bullying Scale and the Peer Relations Questionnaire for students in secondary school*. Journal of School Health, 2007. **77**(2): p. 75-79.
290. Smith, D.J., *School experience and delinquency at ages 13 to 16*. 2006: University of Edinburgh, Centre for Law and Society Edinburgh.
291. Varni, J.W., et al., *The PedsQL™* 4.0 as a pediatric population health measure: feasibility, reliability, and validity*. Ambulatory pediatrics, 2003. **3**(6): p. 329-341.
292. Clarke, A., et al., *Warwick-Edinburgh Mental Well-being Scale (WEMWBS): validated for teenage school students in England and Scotland. A mixed methods assessment*. BMC public health, 2011. **11**(1): p. 487.
293. Goodman, R., *The Strengths and Difficulties Questionnaire: a research note*. Journal of child psychology and psychiatry, 1997. **38**(5): p. 581-586.
294. Bonell, C., et al., *Modifying the secondary school environment to reduce bullying and aggression: the INCLUSIVE cluster RCT*. Public Health Research, 2019. **7**(18): p. 1-164.
295. Jackson II, R.L. and M.A. Hogg, *Encyclopedia of identity*. Vol. 1. 2010: Sage.
296. Kant, I., *Critique of Pure Reason (translated and edited by Paul Guyer & Allen W. Wood)*. 1998.
297. Harrison, V.S., *Internal realism and the problem of religious diversity*. Philosophia, 2006. **34**(3): p. 287-301.
298. Go, J.C., *Bite-size lecture 01: Roy Bhaskar's notion of judgemental rationality*. 2019: Critical Realism-Asia Pacific Network.
299. Stones, R., *Sociological reasoning: Towards a post-modern sociology*. 1996: Macmillan International Higher Education.
300. Moffatt, S., et al., *Using quantitative and qualitative data in health services research—what happens when mixed method findings conflict?*[ISRCTN61522618]. BMC Health Services Research, 2006. **6**(1): p. 28.
301. Maxwell, J.A. and K. Mittapalli, *Realism as a stance for mixed methods research*. Handbook of mixed methods in social & behavioral research, 2010: p. 145-168.
302. Morgan, D.L., *Practical strategies for combining qualitative and quantitative methods: Applications to health research*. Qualitative health research, 1998. **8**(3): p. 362-376.
303. Manzano, A., *The craft of interviewing in realist evaluation*. Evaluation, 2016. **22**(3): p. 342-360.
304. Bonell, C., et al., *Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention*. J Epidemiol Community Health, 2019. **73**(5): p. 455-464.

305. Shackleton, N., et al., *A new measure of unhealthy school environments and its implications for critical assessments of health promotion in schools*. *Critical public health*, 2017. **27**(2): p. 248-262.
306. Melendez-Torres, G., et al., *Moderated mediation analyses to assess intervention mechanisms for impacts on victimisation, psycho-social problems and mental wellbeing: evidence from the INCLUSIVE realist randomized trial*. *Social Science & Medicine*, 2021: p. 113984.
307. Frymier, A.B., G.M. Shulman, and M. Houser, *The development of a learner empowerment measure*. *Communication education*, 1996. **45**(3): p. 181-199.
308. Maslach, C. and S.E. Jackson, *The measurement of experienced burnout*. *Journal of organizational behavior*, 1981. **2**(2): p. 99-113.
309. Farivar, S.S., W.E. Cunningham, and R.D. Hays, *Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V. 1*. *Health and quality of life outcomes*, 2007. **5**(1): p. 1-8.
310. Q.I.P., *NVivo (Version 12)*. 2018.
311. Marchal, B., et al., *Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research*. *Evaluation*, 2012. **18**(2): p. 192-212.
312. Jagosh, J., et al., *Critical reflections on realist review: insights from customizing the methodology to the needs of participatory research assessment*. *Research Synthesis Methods*, 2014. **5**(2): p. 131-141.
313. Mukumbang, F.C., et al., *An exploration of group-based HIV/AIDS treatment and care models in Sub-Saharan Africa using a realist evaluation (Intervention-Context-Actor-Mechanism-Outcome) heuristic tool: a systematic review*. *Implementation Science*, 2017. **12**(1): p. 1-20.
314. Charmaz, K., *Constructing grounded theory: A practical guide through qualitative analysis*. 2006: Sage.
315. Oliver, C., *Critical realist grounded theory: A new approach for social work research*. *British Journal of Social Work*, 2012. **42**(2): p. 371-387.
316. Glaser, B.G. and A.L. Strauss, *The discovery of grounded theory: strategies for qualitative theory*. New Brunswick: Aldine Transaction, 1967.
317. Mead, G.H., *Mind, self and society*. Vol. 111. 1934: Chicago University of Chicago Press.
318. Geertz, C., *The interpretation of cultures: Selected essays*. Vol. 5019. 1973: Basic books.
319. Urquhart, C., *The evolving nature of grounded theory method: The case of the information systems discipline*. *The Sage handbook of grounded theory*, 2007: p. 339-359.
320. Charmaz, K., *Constructing grounded theory*. 2014: sage.
321. Glaser, B.G., *Doing grounded theory: Issues and discussions*. 1998: Sociology Press.
322. Schatzman, L., *Dimensional analysis: Outline in preciform*. Unpublished manuscript. University of California, San Francisco, 1980.
323. Schatzman, L., *Dimensional analysis: Notes on an alternative approach to the grounding of theory in qualitative research*. *Social organization and social process: Essays in honor of Anselm Strauss*, 1991: p. 303-314.
324. Melendez-Torres, G.J., *Substance use, situational characteristics and sexual outcomes in men who have sex with men*. 2014, Oxford University, UK.
325. Kools, S., et al., *Dimensional analysis: Broadening the conception of grounded theory*. *Qualitative Health Research*, 1996. **6**(3): p. 312-330.
326. Byng, R., I. Norman, and S. Redfern, *Using realistic evaluation to evaluate a practice-level intervention to improve primary healthcare for patients with long-term mental illness*. *Evaluation*, 2005. **11**(1): p. 69-93.
327. Marchal, B., M. Dedzo, and G. Kegels, *A realist evaluation of the management of a well-performing regional hospital in Ghana*. *BMC health services research*, 2010. **10**(1): p. 1-14.
328. Rycroft-Malone, J., et al., *A realistic evaluation: the case of protocol-based care*. *Implementation science*, 2010. **5**(1): p. 1-14.

329. Wand, T., K. White, and J. Patching, *Applying a realist (ic) framework to the evaluation of a new model of emergency department based mental health nursing practice*. *Nursing inquiry*, 2010. **17**(3): p. 231-239.
330. Wang, R. and J.H. Ware, *Detecting moderator effects using subgroup analyses*. *Prevention Science*, 2013. **14**(2): p. 111-120.
331. Kraemer, H.C., et al., *Mediators and moderators of treatment effects in randomized clinical trials*. *Archives of general psychiatry*, 2002. **59**(10): p. 877-883.
332. Emsley, R., G. Dunn, and I.R. White, *Mediation and moderation of treatment effects in randomised controlled trials of complex interventions*. *Statistical Methods in Medical Research*, 2010. **19**(3): p. 237-270.
333. Befani, B., S. Ledermann, and F. Sager, *Realistic evaluation and QCA: conceptual parallels and an empirical application*. *Evaluation*, 2007. **13**(2): p. 171-192.
334. Schneider, C.Q., *Realists and Idealists in QCA*. *Political Analysis*, 2018. **26**(2): p. 246-254.
335. Cress, D.M. and D.A. Snow, *The outcomes of homeless mobilization: The influence of organization, disruption, political mediation, and framing*. *American Journal of Sociology*, 2000. **105**(4): p. 1063-1104.
336. Thomas, J., A. O'Mara-Eves, and G. Brunton, *Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example*. *Systematic reviews*, 2014. **3**(1): p. 67.
337. Ragin, C.C., *Redesigning social inquiry: Fuzzy sets and beyond*. 2009: University of Chicago Press.
338. Cronqvist, L., *Tosmana*. Tool for Small-N Analysis [Computer Programme], Version, 2011. **1**(2.0).
339. Roig-Tierno, N., T.F. Gonzalez-Cruz, and J. Llopis-Martinez, *An overview of qualitative comparative analysis: A bibliometric analysis*. *Journal of Innovation & Knowledge*, 2017. **2**(1): p. 15-23.
340. Cilliers, P., *Complexity, deconstruction and relativism*. *Theory, culture & society*, 2005. **22**(5): p. 255-267.
341. CONDON, R.E., *Type III error*. *Archives of Surgery*, 1986. **121**(8): p. 877-878.
342. Schwartz, S. and K.M. Carpenter, *The right answer for the wrong question: consequences of type III error for public health research*. *American journal of public health*, 1999. **89**(8): p. 1175-1180.
343. Hawe, P., A. Shiell, and T. Riley, *Complex interventions: how "out of control" can a randomised controlled trial be?* *Penelope*, 2004. **328**: p. 1561-1563.
344. Toledo, L., et al., *Knowledge, attitudes, and experiences of HIV pre-exposure prophylaxis (PrEP) trial participants in Botswana*. *World journal of AIDS*, 2015. **5**(2): p. 10.
345. Meyers, K., et al., *Lessons for patient education around long-acting injectable PrEP: findings from a mixed-method study of phase II trial participants*. *AIDS and Behavior*, 2018. **22**(4): p. 1209-1216.
346. Bonell, C., G.J. Melendez-Torres, and E. Warren, *Methodological reflections using qualitative research to explore causal mechanisms of complex health interventions*. *Evaluation (in press)*, 2022.
347. Pawson, R., *Theorizing the interview*. *British Journal of Sociology*, 1996: p. 295-314.
348. Dey, I., *Grounding grounded theory: Guidelines for qualitative inquiry*. 1999. San Diego, CA: Academic.
349. Sager, F. and C. Andereggen, *Dealing with complex causality in realist synthesis: The promise of qualitative comparative analysis*. *American journal of evaluation*, 2012. **33**(1): p. 60-78.
350. Sager, F. and W. Schenkel, *Evaluation UVP*, F.a.L. Swiss Agency for the Environment, Editor. 2004: Bern.

351. Evans, C.B., M.W. Fraser, and K.L. Cotter, *The effectiveness of school-based bullying prevention programs: A systematic review*. *Aggression and Violent Behavior*, 2014. **19**(5): p. 532-544.
352. Cross, D., et al., *Impact of the Friendly Schools whole-school intervention on transition to secondary school and adolescent bullying behaviour*. *European Journal of Education*, 2018. **53**(4): p. 495-513.
353. Ragin, C.C., *Turning the tables: How case-oriented research challenges*. *Rethinking social inquiry: Diverse tools, shared standards*, 2004. **123**.
354. Ragin, C.C. and H.S. Becker, *What is a case?: exploring the foundations of social inquiry*. 1992: Cambridge university press.
355. Melendez-Torres, G., et al., *Developing and testing intervention theory by incorporating a views synthesis into a qualitative comparative analysis of intervention effectiveness*. *Research synthesis methods*, 2019. **10**(3): p. 389-397.
356. Rihoux, B. and C.C. Ragin, *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. 2008: Sage Publications.
357. Kien, C., et al., *Pathways leading to success and non-success: a process evaluation of a cluster randomized physical activity health promotion program applying fuzzy-set qualitative comparative analysis*. *BMC public health*, 2018. **18**(1): p. 1386.
358. Sawyer, M.G., et al., *School-based prevention of depression: a randomised controlled study of the beyondblue schools research initiative*. *Journal of Child Psychology and Psychiatry*, 2010. **51**(2): p. 199-209.
359. Pratchett, L., et al., *Preventing Support for Violent Extremism through Community Interventions: A Review of the Evidence-Rapid Evidence Assessment Full Final Report*. 2009.
360. Melendez-Torres, G., et al., *Weight management programmes: re-analysis of a systematic review to identify pathways to effectiveness*. *Health Expectations*, 2018. **21**(3): p. 574-584.
361. Marx, A. and A. Dusa, *Crisp-set qualitative comparative analysis (csQCA), contradictions and consistency benchmarks for model specification*. *Methodological innovations online*, 2011. **6**(2): p. 103-148.
362. Bonell, C., et al., *Will it work here? A realist approach to local decisions about implementing interventions evaluated as effective elsewhere*. *J Epidemiol Community Health*, 2020.
363. Short, K., P. Eadie, and L. Kemp, *Influential factor combinations leading to language outcomes following a home visiting intervention: A qualitative comparative analysis (QCA)*. *International Journal of Language & Communication Disorders*, 2020. **55**(6): p. 936-954.
364. Education, D.f., *The prevent duty: Departmental advice for schools and childcare providers*. Department of Education (UK), 2015: p. 1-11.
365. Exline, J.J., et al., *Forgiveness and justice: A research agenda for social and personality psychology*. *Personality and social psychology Review*, 2003. **7**(4): p. 337-348.
366. Bonell, C., et al., *Role theory of schools and adolescent health*. *The Lancet Child & Adolescent Health*, 2019.
367. Ouellet-Morin, I., et al., *Increased serotonin transporter gene (SERT) DNA methylation is associated with bullying victimization and blunted cortisol response to stress in childhood: a longitudinal study of discordant monozygotic twins*. *Psychological medicine*, 2013. **43**(9): p. 1813-1823.
368. Westhorp, G., *Developing complexity-consistent theory in a realist investigation*. *Evaluation*, 2013. **19**(4): p. 364-382.
369. Westhorp, G., *Using complexity-consistent theory for evaluating complex systems*. *Evaluation*, 2012. **18**(4): p. 405-420.
370. Bonell, C., et al., *Why schools should promote students' health and wellbeing*. *Bmj*, 2014. **348**(7958): p. g3078.

Complete references

1. Pawson, R. and N. Tilley, *Realistic evaluation*. 1997: Sage.
2. Bonell, C., et al., *Realist randomised controlled trials: a new approach to evaluating complex public health interventions*. *Social science & medicine*, 2012. **75**(12): p. 2299-2306.
3. Marchal, B., et al., *Realist RCTs of complex interventions—an oxymoron*. *Social Science & Medicine*, 2013. **94**: p. 124-128.
4. Van Belle, S., et al., *Can 'realist' randomized controlled trials be genuinely realist?* *Trials*, 2016.
5. Bonell, C., et al., *Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial*. *Trials*, 2014. **15**(1): p. 1.
6. Bonell, C., et al., *Initiating change locally in bullying and aggression through the school environment (INCLUSIVE) trial: update to cluster randomised controlled trial protocol*. *Trials*, 2017. **18**(1): p. 238.
7. Markham, W.A. and P. Aveyard, *A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice*. *Social science & medicine*, 2003. **56**(6): p. 1209-1220.
8. Warren, E., et al., *Action groups as a participative strategy for leading whole-school health promotion: Results on implementation from the INCLUSIVE trial in English secondary schools*. *British Educational Research Journal*, 2019.
9. Warren, E., et al., *Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention*. *Trials*, 2020. **21**(1): p. 1-14.
10. Warren, E., G.J. Melendez-Torres, and C. Bonell, *Using fuzzy set qualitative comparative analysis (fsQCA) to explore the causal pathways to reduced bullying victimization in a whole-school intervention: results from a cluster randomized controlled trial*. *Journal of School Violence* (under review), 2021.
11. Warren, E., G.J. Melendez-Torres, and C. Bonell, *Are realist randomised controlled trials possible? A reflection on the INCLUSIVE evaluation of a whole-school, bullying-prevention intervention* *Trials*, 2022. **23**(82).
12. GOV.UK. *Bullying at school*. 23/03/2018]; Available from: <https://www.gov.uk/bullying-at-school/bullying-a-definition>.
13. Farrington, D.P., *Understanding and preventing bullying*. *Crime and justice*, 1993. **17**: p. 381-458.
14. Olweus, D., *Bully/victim problems among schoolchildren: Basic facts and effects of a school based intervention program*. *The development and treatment of childhood aggression*, 1991. **17**(17): p. 411-448.
15. Gaffney, H., M.M. Ttofi, and D.P. Farrington, *Evaluating the effectiveness of school-bullying prevention programs: An updated meta-analytical review*. *Aggression and violent behavior*, 2019. **45**: p. 111-133.
16. Cantone, E., et al., *Interventions on bullying and cyberbullying in schools: A systematic review*. *Clinical practice and epidemiology in mental health: CP & EMH*, 2015. **11**(Suppl 1 M4): p. 58.
17. Chamberlain, T., et al., *Tellus4 national report*. London: Department for Children, Schools and Families, 2010.
18. Roe, S. and J. Ashe, *Young people and crime: Findings from the 2006 Offending, Crime and Justice survey*. 2008.
19. Patchin, J.W. and S. Hinduja, *Bullies move beyond the schoolyard: A preliminary look at cyberbullying*. *Youth violence and juvenile justice*, 2006. **4**(2): p. 148-169.

20. Kowalski, R.M., et al., *Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth*. Psychological bulletin, 2014. **140**(4): p. 1073.
21. Cappadocia, M.C., W.M. Craig, and D. Pepler, *Cyberbullying: Prevalence, stability, and risk factors during adolescence*. Canadian Journal of School Psychology, 2013. **28**(2): p. 171-192.
22. Camerini, A.-L., et al., *Cyberbullying perpetration and victimization among children and adolescents: A systematic review of longitudinal studies*. Telematics and Informatics, 2020. **49**: p. 101362.
23. Sharp, S. and P. Smith, *School bullying: Insights and perspectives*. 2002: Routledge.
24. Carlon, A., S. Freedman, and N. Skaar, *FORGIVENESS EDUCATION AND BULLYING PREVENTION AND INTERVENTION: BENEFITS FOR BULLIES, VICTIMS AND SCHOOLS*.
25. Wolke, D., et al., *Impact of bullying in childhood on adult health, wealth, crime, and social outcomes*. Psychological science, 2013. **24**(10): p. 1958-1970.
26. Coyle, S., A. Cipra, and S.Y. Rueger, *Bullying types and roles in early adolescence: Latent classes of perpetrators and victims*. Journal of School Psychology, 2021. **89**: p. 51-71.
27. Haynie, D.L., et al., *Bullies, victims, and bully/victims: Distinct groups of at-risk youth*. The Journal of Early Adolescence, 2001. **21**(1): p. 29-49.
28. Baldry, A.C., D.P. Farrington, and A. Sorrentino, *School bullying and cyberbullying among boys and girls: Roles and overlap*. Journal of Aggression, Maltreatment & Trauma, 2017. **26**(9): p. 937-951.
29. Thomas, H.J., J.P. Connor, and J.G. Scott, *Integrating traditional bullying and cyberbullying: challenges of definition and measurement in adolescents—a review*. Educational psychology review, 2015. **27**(1): p. 135-152.
30. Walters, G.D., *School-age bullying victimization and perpetration: A meta-analysis of prospective studies and research*. Trauma, Violence, & Abuse, 2020: p. 1524838020906513.
31. Bronfenbrenner, U., *The ecology of human development: Experiments by nature and design*. 1979: Harvard university press.
32. Bronfenbrenner, U., *Making human beings human: Bioecological perspectives on human development*. 2005: sage.
33. Bronfenbrenner, U. and P. Morris, *The bioecological model of human development*, in *Handbook of Child Psychology*, R.M. Lerner and W. Damon, Editors. 2006, John Wiley & Sons: Hoboken, New Jersey. p. 793-828.
34. Divecha, D. and M. Brackett, *Rethinking school-based bullying prevention through the lens of social and emotional learning: A bioecological perspective*. International journal of bullying prevention, 2020. **2**(2): p. 93-113.
35. Valle, J.E., L.C. Williams, and A.C. Stelko-Pereira, *Whole-school antibullying interventions: A systematic review of 20 years of publications*. Psychology in the Schools, 2020. **57**(6): p. 868-883.
36. Inchley, J. and D. Currie, *Growing up unequal: gender and socioeconomic differences in young people's health and well-being. Health Behaviour in School-aged Children (HBSC) study: international report from the 2013/2014 survey. WHO, 2016*. Health policy for children and adolescents, 2016(7).
37. Brooks, F., et al., *HBSC England national report 2014*. Hatfield, UK: BMJ Publishing Group, 2015.
38. van Geel, M., et al., *Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism*. Personality and individual differences, 2017. **106**: p. 231-235.
39. Smokowski, P.R. and K.H. Kopasz, *Bullying in school: An overview of types, effects, family characteristics, and intervention strategies*. Children & Schools, 2005. **27**(2): p. 101-110.
40. Wright, M.F., *Predictors of anonymous cyber aggression: The role of adolescents' beliefs about anonymity, aggression, and the permanency of digital content*. Cyberpsychology, Behavior, and Social Networking, 2014. **17**(7): p. 431-438.

41. Fergusson, D.M., L.J. Horwood, and N. Swain-Campbell, *Cannabis use and psychosocial adjustment in adolescence and young adulthood*. *Addiction*, 2002. **97**(9): p. 1123-1135.
42. McGue, M. and W.G. Iacono, *The adolescent origins of substance use disorders*. *International journal of methods in psychiatric research*, 2008. **17**(S1): p. S30-S38.
43. Carlyle, K.E. and K.J. Steinman, *Demographic differences in the prevalence, co-occurrence, and correlates of adolescent bullying at school*. *Journal of School Health*, 2007. **77**(9): p. 623-629.
44. Ttofi, M.M. and D.P. Farrington, *Risk and protective factors, longitudinal research, and bullying prevention*. *New directions for youth development*, 2012. **2012**(133): p. 85-98.
45. Juvonen, J., S. Graham, and M.A. Schuster, *Bullying among young adolescents: The strong, the weak, and the troubled*. *Pediatrics*, 2003. **112**(6): p. 1231-1237.
46. Brimblecombe, N., et al., *Long term economic impact associated with childhood bullying victimisation*. *Social Science & Medicine*, 2018. **208**: p. 134-141.
47. Yeager, D.S., et al., *Declines in efficacy of anti-bullying programs among older adolescents: Theory and a three-level meta-analysis*. *Journal of applied developmental psychology*, 2015. **37**: p. 36-51.
48. Dake, J.A., J.H. Price, and S.K. Telljohann, *The nature and extent of bullying at school*. *Journal of school health*, 2003. **73**(5): p. 173-180.
49. Olafsen, R.N. and V. Viemerö, *Bully/victim problems and coping with stress in school among 10-to 12-year-old pupils in Åland, Finland*. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 2000. **26**(1): p. 57-65.
50. Fujikawa, S., et al., *Bullying Across Late Childhood and Early Adolescence: A Prospective Cohort of Students Assessed Annually From Grades 3 to 8*. *Academic Pediatrics*, 2021. **21**(2): p. 344-351.
51. Goldweber, A., T.E. Waasdorp, and C.P. Bradshaw, *Examining the link between forms of bullying behaviors and perceptions of safety and belonging among secondary school students*. *Journal of school psychology*, 2013. **51**(4): p. 469-485.
52. Sumter, S.R., et al., *Developmental trajectories of peer victimization: Off-line and online experiences during adolescence*. *Journal of Adolescent Health*, 2012. **50**(6): p. 607-613.
53. Zych, I., D.P. Farrington, and M.M. Ttofi, *Protective factors against bullying and cyberbullying: A systematic review of meta-analyses*. *Aggression and Violent Behavior*, 2019. **45**: p. 4-19.
54. Georgiou, S.N. and P. Stavrinides, *Parenting at home and bullying at school*. *Social psychology of education*, 2013. **16**(2): p. 165-179.
55. Nocentini, A., et al., *Parents, family characteristics and bullying behavior: A systematic review*. *Aggression and violent behavior*, 2019. **45**: p. 41-50.
56. Ahmed, E. and V. Braithwaite, *Forgiveness, reconciliation, and shame: Three key variables in reducing school bullying*. *Journal of social issues*, 2006. **62**(2): p. 347-370.
57. Georgiou, S.N., *Bullying and victimization at school: The role of mothers*. *British Journal of Educational Psychology*, 2008. **78**(1): p. 109-125.
58. Jansen, D.E., et al., *Early risk factors for being a bully, victim, or bully/victim in late elementary and early secondary education. The longitudinal TRAILS study*. *BMC public health*, 2011. **11**(1): p. 440.
59. Bevilacqua, L., et al., *The role of family and school-level factors in bullying and cyberbullying: a cross-sectional study*. *BMC pediatrics*, 2017. **17**(1): p. 160.
60. Sellström, E. and S. Bremberg, *Is there a "school effect" on pupil outcomes? A review of multilevel studies*. *Journal of Epidemiology & Community Health*, 2006. **60**(2): p. 149-155.
61. Fletcher, A., et al., *Cannabis use and 'safe' identities in an inner-city school risk environment*. *International Journal of Drug Policy*, 2009. **20**(3): p. 244-250.
62. Jamal, F., et al., *The school environment and student health: a systematic review and meta-ethnography of qualitative research*. *BMC Public Health*, 2013. **13**(1): p. 798.

63. Bonell, C., et al., *Theories of how the school environment impacts on student health: Systematic review and synthesis*. Health & place, 2013. **24**: p. 242-249.
64. Paille, B., *Toxic schools: high-poverty education in New York and Amsterdam*. 2013: University of Chicago Press.
65. Dance, L.J., *Tough fronts: The impact of street culture on schooling*. 2012: Routledge.
66. Viner, R.M., et al., *Adolescence and the social determinants of health*. The Lancet, 2012. **379**(9826): p. 1641-1652.
67. Flay, B.R., et al., *Effects of 2 prevention programs on high-risk behaviors among African American youth: a randomized trial*. Archives of Pediatrics & Adolescent Medicine, 2004. **158**(4): p. 377-384.
68. Resnick, M.D., et al., *Protecting adolescents from harm: findings from the National Longitudinal Study on Adolescent Health*. Jama, 1997. **278**(10): p. 823-832.
69. Patton, G.C., et al., *Our future: a Lancet commission on adolescent health and wellbeing*. The Lancet, 2016.
70. Jain, S., et al., *School climate and physical adolescent relationship abuse: differences by sex, socioeconomic status, and bullying*. Journal of adolescence, 2018. **66**: p. 71-82.
71. Wentzel, K.R., C.M. Barry, and K.A. Caldwell, *Friendships in Middle School: Influences on Motivation and School Adjustment*. Journal of educational psychology, 2004. **96**(2): p. 195.
72. Wentzel, K.R., *Relations of social goal pursuit to social acceptance, classroom behavior, and perceived social support*. Journal of Educational Psychology, 1994. **86**(2): p. 173.
73. Wentzel, K.R., *Social relationships and motivation in middle school: The role of parents, teachers, and peers*. Journal of educational psychology, 1998. **90**(2): p. 202.
74. Van Ryzin, M.J., A.A. Gravely, and C.J. Roseth, *Autonomy, belongingness, and engagement in school as contributors to adolescent psychological well-being*. Journal of youth and adolescence, 2009. **38**(1): p. 1-12.
75. Vieno, A., et al., *Social support, sense of community in school, and self-efficacy as resources during early adolescence: an integrative model*. American Journal of Community Psychology, 2007. **39**(1-2): p. 177-190.
76. Bond, L., et al., *Social and school connectedness in early secondary school as predictors of late teenage substance use, mental health, and academic outcomes*. Journal of Adolescent Health, 2007. **40**(4): p. 357. e9-357. e18.
77. Rasmussen, M., et al., *School connectedness and daily smoking among boys and girls: the influence of parental smoking norms*. The European Journal of Public Health, 2005. **15**(6): p. 607-612.
78. Johnson, S.L., *Improving the school environment to reduce school violence: A review of the literature*. Journal of school health, 2009. **79**(10): p. 451-465.
79. Shackleton, N., et al., *Systematic review of reviews of observational studies of school-level effects on sexual health, violence and substance use*. Health & place, 2016. **39**: p. 168-176.
80. Bonell, C., et al., *The effects on student health of interventions modifying the school environment: systematic review*. Journal of epidemiology and community health, 2013. **67**(8): p. 677-681.
81. Markham, W.A., et al., *Does school ethos explain the relationship between value-added education and teenage substance use? A cohort study*. Social science & medicine, 2012. **75**(1): p. 69-76.
82. Tobler, A.L., et al., *Preventing the link between SES and high-risk behaviors: "Value-added" education, drug use and delinquency in high-risk, urban schools*. Prevention Science, 2011. **12**(2): p. 211-221.
83. Bonell, C., et al., *The effects of the school environment on student health: a systematic review of multi-level studies*. Health & place, 2013. **21**: p. 180-191.

84. Bond, L., et al., *The Gatehouse Project: can a multilevel school intervention affect emotional wellbeing and health risk behaviours?* Journal of epidemiology and community health, 2004. **58**(12): p. 997-1003.
85. Bond, L., et al., *Building capacity for system-level change in schools: lessons from the Gatehouse Project.* Health Education & Behavior, 2001. **28**(3): p. 368-383.
86. Bond, L., et al., *Long-Term Impact of the Gatehouse Project on Cannabis Use of 16-Year-Olds in Australia.* Journal of School Health, 2004. **74**(1): p. 23-29.
87. Patton, G.C., et al., *Promoting social inclusion in schools: a group-randomized trial of effects on student health risk behavior and well-being.* American journal of public health, 2006. **96**(9): p. 1582-1587.
88. Hale, D.R. and R.M. Viner, *The correlates and course of multiple health risk behaviour in adolescence.* BMC public health, 2016. **16**(1): p. 1.
89. Moffitt, T.E., *Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy.* Psychological review, 1993. **100**(4): p. 674.
90. Larson, R.W., et al., *Changes in adolescents' daily interactions with their families from ages 10 to 18: Disengagement and transformation.* Developmental Psychology, 1996. **32**(4): p. 744.
91. Steer, A., *learning behavior: lessons learned a review of behavior standards and practices in out schools*, S.a.F. Department for Children, Editor. 2009: London.
92. Moore, S.E., et al., *Consequences of bullying victimization in childhood and adolescence: A systematic review and meta-analysis.* World journal of psychiatry, 2017. **7**(1): p. 60.
93. Hair, E.C., et al., *Risky behaviors in late adolescence: co-occurrence, predictors, and consequences.* Journal of Adolescent Health, 2009. **45**(3): p. 253-261.
94. Fisher, H.L., et al., *Bullying victimisation and risk of self harm in early adolescence: longitudinal cohort study.* bmj, 2012. **344**: p. e2683.
95. Winsper, C., et al., *Involvement in bullying and suicide-related behavior at 11 years: a prospective birth cohort study.* Journal of the American Academy of Child & Adolescent Psychiatry, 2012. **51**(3): p. 271-282. e3.
96. Cosma, A., et al., *Trends in bullying victimization in Scottish adolescents 1994–2014: changing associations with mental well-being.* International journal of public health, 2017. **62**(6): p. 639-646.
97. Moore, S.E., et al., *Adolescent peer aggression and its association with mental health and substance use in an Australian cohort.* Journal of adolescence, 2014. **37**(1): p. 11-21.
98. Calvete, E., I. Orue, and M. Gámez-Guadix, *Cyberbullying victimization and depression in adolescents: The mediating role of body image and cognitive schemas in a one-year prospective study.* European Journal on Criminal Policy and Research, 2016. **22**(2): p. 271-284.
99. Cole, D.A., et al., *Longitudinal and incremental relation of cybervictimization to negative self-cognitions and depressive symptoms in young adolescents.* Journal of abnormal child psychology, 2016. **44**(7): p. 1321-1332.
100. Espinoza, G., *Daily cybervictimization among Latino adolescents: Links with emotional, physical and school adjustment.* Journal of Applied Developmental Psychology, 2015. **38**: p. 39-48.
101. Bonanno, R.A. and S. Hymel, *Cyber bullying and internalizing difficulties: Above and beyond the impact of traditional forms of bullying.* Journal of youth and adolescence, 2013. **42**(5): p. 685-697.
102. Smokowski, P.R., C.B. Evans, and K.L. Cotter, *The differential impacts of episodic, chronic, and cumulative physical bullying and cyberbullying: The effects of victimization on the school experiences, social support, and mental health of rural adolescents.* Violence and Victims, 2014. **29**(6): p. 1029-1046.

103. Nansel, T.R., et al., *Cross-national consistency in the relationship between bullying behaviors and psychosocial adjustment*. Archives of Pediatrics & Adolescent Medicine, 2004. **158**(8): p. 730-736.
104. Wolke, D., et al., *Bullying and victimization of primary school children in England and Germany: Prevalence and school factors*. British Journal of Psychology, 2001. **92**(4): p. 673-696.
105. Nansel, T.R., et al., *Relationships between bullying and violence among US youth*. Archives of Pediatrics & Adolescent Medicine, 2003. **157**(4): p. 348-353.
106. Radliff, K.M., et al., *Illuminating the relationship between bullying and substance use among middle and high school youth*. Addictive behaviors, 2012. **37**(4): p. 569-572.
107. Tharp-Taylor, S., A. Haviland, and E.J. D'Amico, *Victimization from mental and physical bullying and substance use in early adolescence*. Addictive behaviors, 2009. **34**(6-7): p. 561-567.
108. Fletcher, A., et al., *How might schools influence young people's drug use? Development of theory from qualitative case-study research*. Journal of Adolescent Health, 2009. **45**(2): p. 126-132.
109. Mishna, F., et al., *Risk factors for involvement in cyber bullying: Victims, bullies and bully-victims*. Children and Youth Services Review, 2012. **34**(1): p. 63-70.
110. Demaray, M.K. and C.K. Malecki, *Perceptions of the frequency and importance of social support by students classified as victims, bullies, and bully/victims in an urban middle school*. School Psychology Review, 2003. **32**(3): p. 471-490.
111. Jamal, F., et al., *The social ecology of girls' bullying practices: exploratory research in two London schools*. Sociology of health & illness, 2015. **37**(5): p. 731-744.
112. González-Cabrera, J., et al., *Relationship between cyberbullying roles, cortisol secretion and psychological stress*. Computers in Human Behavior, 2017. **70**: p. 153-160.
113. Zych, I., et al., *School bullying and dating violence in adolescents: A systematic review and meta-analysis*. Trauma, Violence, & Abuse, 2021. **22**(2): p. 397-412.
114. Boxer, P., et al., *Exposure to low-level aggression in schools associations with aggressive behaviour, future expectation, and perceived safety*. Violence and victims, 2003. **18**(6): p. 691-705.
115. Kochenderfer, B.J. and G.W. Ladd, *Peer victimization: Cause or consequence of school maladjustment?* Child development, 1996. **67**(4): p. 1305-1317.
116. Nansel, T.R., et al., *Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment*. Jama, 2001. **285**(16): p. 2094-2100.
117. Glew, G.M., et al., *Bullying, psychosocial adjustment, and academic performance in elementary school*. Archives of pediatrics & adolescent medicine, 2005. **159**(11): p. 1026-1031.
118. Yang, J., X. Wang, and L. Lei, *Perceived school climate and adolescents' bullying perpetration: A moderated mediation model of moral disengagement and peers' defending*. Children and youth services review, 2020. **109**: p. 104716.
119. Price, L.H., et al., *Telomeres and early-life stress: an overview*. Biological psychiatry, 2013. **73**(1): p. 15-23.
120. Vreeman, R.C. and A.E. Carroll, *A systematic review of school-based interventions to prevent bullying*. Archives of Pediatrics & Adolescent Medicine, 2007. **161**(1): p. 78-88.
121. Copeland, W.E., et al., *Adult psychiatric outcomes of bullying and being bullied by peers in childhood and adolescence*. JAMA psychiatry, 2013. **70**(4): p. 419-426.
122. Lereya, S.T., et al., *Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries*. The Lancet Psychiatry, 2015. **2**(6): p. 524-531.
123. Beinar, S., et al., *Youth at Risk?: A National Survey of Risk Factors, Protective Factors and Problem Behaviour Among Young People in England, Scotland and Wales*. 2002: Communities that Care London.

124. Jantzer, V., et al., *The cost incurred by victims of bullying from a societal perspective: estimates based on a German online survey of adolescents*. European child & adolescent psychiatry, 2019. **28**(4): p. 585-594.
125. Takizawa, R., B. Maughan, and L. Arseneault, *Adult health outcomes of childhood bullying victimization: evidence from a five-decade longitudinal British birth cohort*. American journal of psychiatry, 2014. **171**(7): p. 777-784.
126. DCSF, *Your child, your schools, our future: building a 21st century schools system*. 2009, London: TSO.
127. Health, D.o., *Healthy Lives, Brighter Futures—The Strategy for Children's and Younger People's Health*. 2009, DH London.
128. Durlak, J.A., et al., *The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions*. Child development, 2011. **82**(1): p. 405-432.
129. Bonell, C., et al., *Systematic review of the effects of schools and school environment interventions on health: evidence mapping and synthesis*. Public Health Research, 2013. **1**(1).
130. Shackleton, N., et al., *School-Based Interventions Going Beyond Health Education to Promote Adolescent Health: Systematic Review of Reviews*. Journal of Adolescent Health, 2016. **58**(4): p. 382-396.
131. Smith, J.D., et al., *The effectiveness of whole-school antibullying programs: A synthesis of evaluation research*. School psychology review, 2004. **33**(4): p. 547.
132. Ttofi, M.M. and D.P. Farrington, *Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review*. Journal of Experimental Criminology, 2011. **7**(1): p. 27-56.
133. Wells, J., J. Barlow, and S. Stewart-Brown, *A systematic review of universal approaches to mental health promotion in schools*. Health Education, 2003. **103**(4): p. 197-220.
134. Lipsey, M.W. and D.B. Wilson, *Practical meta-analysis*. 2001: Sage Publications, Inc.
135. Olweus, D. and S.P. Limber, *The Olweus Bullying Prevention Program: Implementation and Evaluation over Two Decades*, in *Handbook of bullying in schools*. 2009, Routledge. p. 387-412.
136. Ng, E.D., J.Y.X. Chua, and S. Shorey, *The effectiveness of educational interventions on traditional bullying and cyberbullying among adolescents: A systematic review and meta-analysis*. Trauma, Violence, & Abuse, 2020: p. 1524838020933867.
137. Gaffney, H., D.P. Farrington, and M.M. Ttofi, *Examining the effectiveness of school-bullying intervention programs globally: A meta-analysis*. International Journal of Bullying Prevention, 2019. **1**(1): p. 14-31.
138. Taylor, R.D., et al., *Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects*. Child development, 2017. **88**(4): p. 1156-1171.
139. Silva, J.L.d., et al., *Results from interventions addressing social skills to reduce school bullying: a systematic review with meta-analysis*. Trends in Psychology, 2018. **26**: p. 509-522.
140. Sutton, J., P.K. Smith, and J. Swettenham, *Bullying and 'theory of mind': A critique of the 'social skills deficit' view of anti-social behaviour*. Social development, 1999. **8**(1): p. 117-127.
141. Fonagy, P., et al., *A cluster randomized controlled trial of child-focused psychiatric consultation and a school systems-focused intervention to reduce aggression*. Journal of Child Psychology and Psychiatry, 2009. **50**(5): p. 607-616.
142. Wigelsworth, M., N. Humphrey, and A. Lendrum, *Evaluation of a school-wide preventive intervention for adolescents: The secondary social and emotional aspects of learning (SEAL) programme*. School Mental Health, 2013. **5**(2): p. 96-109.
143. Brooks, F., *The link between pupil health and wellbeing and attainment: a briefing for head teachers, governors and staff in education settings: November 2014*. 2014.

144. Blank, L., et al., *Promoting well-being by changing behaviour: A systematic review and narrative synthesis of the effectiveness of whole secondary school behavioural interventions*. *Mental Health Review Journal*, 2010. **15**(2): p. 43-53.
145. Hallam, S., J. Rhamie, and J. Shaw, *Evaluation of the primary behaviour and attendance pilot*. 2006: Department for Education and Skills London,, UK.
146. Challen, A., et al., *UK resilience programme evaluation*. 2011.
147. Van Ness, D.W. and K.H. Strong, *Restoring justice: An introduction to restorative justice*. 2014: Routledge.
148. Latimer, J., C. Dowden, and D. Muise, *The effectiveness of restorative justice practices: A meta-analysis*. *The prison journal*, 2005. **85**(2): p. 127-144.
149. Morris, A. and G. Maxwell, *Restorative justice for juveniles: Conferencing, mediation and circles*. 2001: Bloomsbury Publishing.
150. Wallis, P., *Understanding restorative justice: How empathy can close the gap created by crime*. 2014: Policy Press.
151. Braithwaite, J. *Restorative justice: Theories and worries*. in *Visiting Experts' Papers: 123rd International Senior Seminar, Resource Material Series*. 2004.
152. Van Dyke, C.J. and M.J. Elias, *How forgiveness, purpose, and religiosity are related to the mental health and well-being of youth: A review of the literature*. *Mental Health, Religion and Culture*, 2007. **10**(4): p. 395-415.
153. Worthington, E.L. and M. Scherer, *Forgiveness is an emotion-focused coping strategy that can reduce health risks and promote health resilience: Theory, review, and hypotheses*. *Psychology & Health*, 2004. **19**(3): p. 385-405.
154. van der Valk, A., *There Are No Bullies: Just Children Who Bully And You Can Help Them*. *The Education Digest*, 2014. **79**(8): p. 39.
155. Kokotsaki, D., C. White, and B. Hopkins, *Capturing change: a review of the implementation of Restorative Approaches and its outcomes within a local authority in North East England*. *Online educational research journal.*, 2014. **5**(5): p. 151.
156. Acosta, J.D., et al., *A cluster-randomized trial of restorative practices: An illustration to spur high-quality research and evaluation*. *Journal of Educational and Psychological Consultation*, 2016. **26**(4): p. 413-430.
157. Hopkins, B., *Just schools. A whole school approach to restorative justice* (London, Jessica Kingsley), 2004.
158. Acosta, J., et al., *Evaluation of a Whole-School Change Intervention: Findings from a Two-Year Cluster-Randomized Trial of the Restorative Practices Intervention*. *Journal of youth and adolescence*, 2019. **48**(5): p. 876-890.
159. Tyler, T.R., *Restorative justice and procedural justice: Dealing with rule breaking*. *Journal of social issues*, 2006. **62**(2): p. 307.
160. Darling-Hammond, S., et al., *Effectiveness of restorative justice in US K-12 schools: A review of quantitative research*. *Contemporary School Psychology*, 2020. **24**: p. 295-308.
161. Karp, D.R. and B. Breslin, *Restorative justice in school communities*. *Youth & Society*, 2001. **33**(2): p. 249-272.
162. Anfara Jr, V.A., K.R. Evans, and J.N. Lester, *Restorative justice in education: What we know so far*. *Middle School Journal*, 2013. **44**(5): p. 57-63.
163. Song, S.Y. and S.M. Swearer, *The cart before the horse: The challenge and promise of restorative justice consultation in schools*. 2016, Taylor & Francis.
164. Sumner, M.D., C.J. Silverman, and M.L. Frampton, *School-based restorative justice as an alternative to zero-tolerance policies: Lessons from West Oakland*. Thelton E. Henderson Center for Social Justice, 2010.
165. Wong, D.S., et al., *Program effectiveness of a restorative whole-school approach for tackling school bullying in Hong Kong*. *International journal of offender therapy and comparative criminology*, 2011. **55**(6): p. 846-862.

166. Mas-Expósito, L., et al., *Implementation of Whole School Restorative Approaches to Promote Positive Youth Development: Review of Relevant Literature and Practice Guidelines*. Education Sciences, 2022. **12**(187).
167. Buckley, S. and G.M. Maxwell, *Respectful schools: Restorative practices in education: A summary report*. 2007: Office of the Children's Commissioner and the Institute of Policy Studies, School of Government, Victoria University, Wellington.
168. Learning, M.D.o.C.a.F., *A three-year evaluation of alternative approaches to suspensions and expulsions: a report to the Legislature*. 1998.
169. Board, Y.J., *National evaluation of the restorative justice in schools programme*. London: Youth Justice Board. Whose Justice, 2004.
170. Skinnis, L. and M. Hough, *An Evaluation of Bristol RAiS*. London: ICPR, King's College London, 2009.
171. Kane, J., et al., *Restorative practices in Scottish schools*. Edinburgh: Scottish Executive, 2007.
172. Moore, G.F., et al., *School composition, school culture and socioeconomic inequalities in young people's health: Multi-level analysis of the Health Behaviour in School-aged Children (HBSC) survey in Wales*. British educational research journal, 2017. **43**(2): p. 310-329.
173. Dahlgren, G. and M. Whitehead, *Policies and strategies to promote social equity in health*. Stockholm: Institute for future studies, 1991.
174. Fletcher, A., et al., *Involving young people in changing their school environment to make it safer: findings from a process evaluation in english secondary schools*. Health Education, 2015. **115**(3/4): p. 322-338.
175. Cross, D., et al., *The Friendly Schools Friendly Families programme: Three-year bullying behaviour outcomes in primary school children*. International Journal of Educational Research, 2012. **53**: p. 394-406.
176. Langford, R., et al., *The WHO Health Promoting School framework for improving the health and well-being of students and their academic achievement*. The Cochrane Library, 2014.
177. Fraguas, D., et al., *Assessment of school anti-bullying interventions: a meta-analysis of randomized clinical trials*. JAMA pediatrics, 2021. **175**(1): p. 44-55.
178. Ponsford, R., et al., *Whole-school interventions promoting student commitment to school to prevent substance use and violence and improve educational attainment (In press)*. 2021.
179. Langford, R., et al., *The World Health Organization's Health Promoting Schools framework: a Cochrane systematic review and meta-analysis*. BMC public health, 2015. **15**(1): p. 130.
180. Gaffney, H., M.M. Ttofi, and D.P. Farrington, *Effectiveness of school-based programs to reduce bullying perpetration and victimization: An updated systematic review and meta-analysis*. Campbell Systematic Reviews, 2021. **17**(2): p. e1143.
181. Riestenberg, N., *In-school behavior intervention grants: A three-year evaluation of alternative approaches to suspensions and expulsions*. Roseville, MN: Minnesota Dept. of Children, Families & Learning, 2001.
182. Shaw, G., *Restorative practices in Australian schools: Changing relationships, changing culture*. Conflict Resolution Quarterly, 2007. **25**(1): p. 127-135.
183. Farrington, D.P. and M.M. Ttofi, *School-based programs to reduce bullying and victimization*. The Campbell Collaboration, 2009. **6**: p. 1-149.
184. Durlak, J.A. and E.P. DuPre, *Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation*. American journal of community psychology, 2008. **41**(3-4): p. 327-350.
185. Smith, P.K., C. Salmivalli, and H. Cowie, *Effectiveness of school-based programs to reduce bullying: A commentary*. Journal of Experimental Criminology, 2012. **8**(4): p. 433-441.
186. Copeland, W.E., et al., *Childhood and adolescent psychiatric disorders as predictors of young adult disorders*. Archives of general psychiatry, 2009. **66**(7): p. 764-772.

187. Segawa, E., et al., *Evaluation of the effects of the Aban Aya Youth Project in reducing violence among African American adolescent males using latent class growth mixture modeling techniques*. Evaluation review, 2005. **29**(2): p. 128-148.
188. Baron, R.M. and D.A. Kenny, *The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations*. Journal of personality and social psychology, 1986. **51**(6): p. 1173.
189. Gardner, F., J. Burton, and I. Klimes, *Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: outcomes and mechanisms of change*. Journal of child psychology and psychiatry, 2006. **47**(11): p. 1123-1132.
190. Garandeau, C.F. and C. Salmivalli, *Can healthier contexts be harmful? A new perspective on the plight of victims of bullying*. Child Development Perspectives, 2019. **13**(3): p. 147-152.
191. Salmivalli, C., *Peer victimization and adjustment in young adulthood: commentary on the special section*. Journal of abnormal child psychology, 2018. **46**(1): p. 67-72.
192. Huitsing, G., et al., *The healthy context paradox: victims' adjustment during an anti-bullying intervention*. Journal of Child and Family Studies, 2019. **28**(9): p. 2499-2509.
193. Melendez-Torres, G.J., et al., *Locating and testing the healthy context paradox: examples from the INCLUSIVE trial (In press)*. BMC Medical Research Methodology, 2022.
194. Patton, G.C., et al., *The Gatehouse Project: a systematic approach to mental health promotion in secondary schools*. Australian & New Zealand Journal of Psychiatry, 2000. **34**(4): p. 586-593.
195. Moore, G.F., et al., *Process evaluation of complex interventions: Medical Research Council guidance*. bmj, 2015. **350**: p. h1258.
196. Rutter, H., et al., *The need for a complex systems model of evidence for public health*. The Lancet, 2017. **390**(10112): p. 2602-2604.
197. Craig, P., et al., *Developing and evaluating complex interventions*. Medical Research Council, UK, 2011.
198. Saarento, S., A.J. Boulton, and C. Salmivalli, *Reducing bullying and victimization: Student-and classroom-level mechanisms of change*. Journal of abnormal child psychology, 2015. **43**(1): p. 61-76.
199. Hanckel, B., et al., *The use of Qualitative Comparative Analysis (QCA) to address causality in complex systems: a systematic review of research on public health interventions*. BMC public health, 2021. **21**(1): p. 1-22.
200. Keshavarz, N., et al., *Schools as social complex adaptive systems: a new way to understand the challenges of introducing the health promoting schools concept*. Social science & medicine, 2010. **70**(10): p. 1467-1474.
201. Popper, K.R., *Objective knowledge: An evolutionary approach*. 1972.
202. Bonell, C., G. Melendez-Torres, and S. Quilley, *The potential role for sociologists in designing RCTs and of RCTs in refining sociological theory: A commentary on Deaton and Cartwright*. Social science & medicine (1982), 2018.
203. Moore, G.F. and R.E. Evans, *What theory, for whom and in which context? Reflections on the application of theory in the development and evaluation of complex population health interventions*. SSM-population health, 2017. **3**: p. 132-135.
204. Merton, R.K., *On sociological theories of the middle range [1949]*. 1949: na.
205. Breuer, E., et al., *Using theory of change to design and evaluate public health interventions: a systematic review*. Implementation Science, 2015. **11**(1): p. 1-17.
206. Shearn, K., et al., *Building realist program theory for large complex and messy interventions*. International Journal of Qualitative Methods, 2017. **16**(1): p. 1609406917741796.
207. Hill, A.B., *The clinical trial*. British Medical Bulletin, 1951. **7**(4): p. 278-282.
208. Services, M.R.C.H. and P.H.R. Board, *A framework for development and evaluation of RCTs for complex interventions to improve health*. 2000: Medical Research Council.

209. Chalmers, I., *Fisher and Bradford Hill: theory and pragmatism?* International journal of epidemiology, 2003. **32**(6): p. 922-924.
210. Evans, J., *Epidemiology in Practice: Randomised Controlled Trials*. Community eye health, 1998. **11**(26): p. 26.
211. Oakley, A., *Who's afraid of the randomized controlled trial? Some dilemmas of the scientific method and 'good' research practice*. 2004: Routledge.
212. O' Cathain, A., E. Murphy, and J. Nicholl, *Three techniques for integrating data in mixed methods studies*. Bmj, 2010. **341**: p. c4587.
213. Yin, R.K., *Mixed methods research: Are the methods genuinely integrated or merely parallel*. Research in the Schools, 2006. **13**(1): p. 41-47.
214. Oakley, A., et al., *Process evaluation in randomised controlled trials of complex interventions*. BMJ: British Medical Journal, 2006. **332**(7538): p. 413.
215. Evans, R., J. Scourfield, and S. Murphy, *Pragmatic, formative process evaluations of complex interventions and why we need more of them*. 2015, BMJ Publishing Group Ltd.
216. Moore, G., et al., *Process evaluation of complex interventions*, Medical Research Council, Editor. 2015.
217. Oakley, A., et al., *Evaluating Processes: A Case Study of a Randomized Controlled Trial of Sex Education*. Evaluation, 2004. **10**(4): p. 440-462.
218. Suchman, E., *Evaluative Research: Principles and Practice in Public Service and Social Action Progr*. 1968: Russell Sage Foundation.
219. Steckler, A.B., L. Linnan, and B. Israel, *Process evaluation for public health interventions and research*. 2002: Jossey-Bass San Francisco, CA.
220. Strange, V., et al., *Integrating process with outcome data in a randomized controlled trial of sex education*. Evaluation, 2006. **12**(3): p. 330-352.
221. Pocock, S.J. and M. Abdalla, *The hope and the hazards of using compliance data in randomized controlled trials*. Statistics in medicine, 1998. **17**(3): p. 303-317.
222. Parry-Langdon, N., et al., *Process evaluation of health promotion interventions*. Policy & Politics, 2003. **31**(2): p. 207-216.
223. Wight, D. and A. Obasi, *Unpacking the black box: the importance of process data to explain outcomes*. 2003.
224. Filstead, W.J., *Qualitative methods: A needed perspective in evaluation research*. Qualitative and quantitative methods in evaluation research, 1979: p. 33-48.
225. Stephenson, J., et al., *The long-term effects of a peer-led sex education programme (RIPPLE): a cluster randomised trial in schools in England*. PLoS Med, 2008. **5**(11): p. e224.
226. Patton, M.Q., *A context and boundaries for a theory-driven approach to validity*. Evaluation and Program Planning, 1989. **12**(4): p. 375-377.
227. Stame, N., *Theory-based evaluation and types of complexity*. Evaluation, 2004. **10**(1): p. 58-76.
228. Perry, C.L., et al., *The Child and Adolescent Trial for Cardiovascular Health (CATCH): intervention, implementation, and feasibility for elementary schools in the United States*. Health Education & Behavior, 1997. **24**(6): p. 716-735.
229. Baranowski, T., C. Anderson, and C. Carmack, *Mediating variable framework in physical activity interventions: How are we doing? How might we do better?* American journal of preventive medicine, 1998. **15**(4): p. 266-297.
230. Grant, A., et al., *Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting*. Trials, 2013. **14**(1): p. 15.
231. Sherman, L.W., J.D. Schmidt, and D.P. Rogan, *Policing domestic violence: Experiments and dilemmas*. 1992: Free Press.
232. Tilley, N. *Realistic evaluation: an overview*. in *founding conference of the Danish Evaluation Society*. 2000.
233. Pawson, R. and N. Tilley, *Realist evaluation*. 2004, London: Sage Publications Google Scholar.

234. Porter, S., T. McConnell, and J. Reid, *The possibility of critical realist randomised controlled trials*. *Trials*, 2017. **18**(1): p. 133.
235. Hinds, K. and K. Dickson, *Realist synthesis: a critique and an alternative*. *Journal of Critical Realism*, 2021. **20**(1): p. 1-17.
236. Porter, S., *The uncritical realism of realist evaluation*. *Evaluation*, 2015. **21**(1): p. 65-82.
237. Bhaskar, R., *The possibility of naturalism: A philosophical critique of the contemporary human sciences*. 2014: Routledge.
238. King, A., *The odd couple: Margaret Archer, Anthony Giddens and British social theory*. *The British journal of sociology*, 2010. **61**: p. 253-260.
239. Archer, M.S., *Realist social theory: The morphogenetic approach*. 1995: Cambridge university press.
240. Giddens, A., *The constitution of society: Outline of the theory of structuration*. 1984: Univ of California Press.
241. Bhaskar, R., *A Realist Theory of Science*. 1975: Verso.
242. Lawson, T., *Economics and reality*. 1997: Routledge.
243. Dalkin, S.M., et al., *What's in a mechanism? Development of a key concept in realist evaluation*. *Implementation Science*, 2015. **10**(1): p. 49.
244. Lemire, S., et al., *What is this thing called a mechanism? Findings from a review of realist evaluations*. *New Directions for Evaluation*, 2020. **2020**(167): p. 73-86.
245. Porter, S., *Realist evaluation: an immanent critique*. *Nursing Philosophy*, 2015. **16**(4): p. 239-251.
246. Dalkin, S.M., et al., *Reducing inequalities in care for patients with non-malignant diseases: Insights from a realist evaluation of an integrated palliative care pathway*. *Palliative medicine*, 2016. **30**(7): p. 690-697.
247. Ford, J.A., et al., *Access to primary care for socio-economically disadvantaged older people in rural areas: exploring realist theory using structural equation modelling in a linked dataset*. *BMC medical research methodology*, 2018. **18**(1): p. 57.
248. Ravn, R., *Testing mechanisms in large-N realistic evaluations*. *Evaluation*, 2019. **25**(2): p. 171-188.
249. Greenhalgh, T., et al., *How do you modernize a health service? A realist evaluation of whole-scale transformation in London*. *The Milbank Quarterly*, 2009. **87**(2): p. 391-416.
250. Hawkins, A.J., *Realist evaluation and randomised controlled trials for testing program theory in complex social systems*. *Evaluation*, 2016. **22**(3): p. 270-285.
251. Mirzoev, T., et al., *Study protocol: realist evaluation of effectiveness and sustainability of a community health workers programme in improving maternal and child health in Nigeria*. *Implementation Science*, 2016. **11**(1): p. 83.
252. Gilmore, B., et al., *Improving the performance of community health workers in humanitarian emergencies: a realist evaluation protocol for the PIECES programme*. *BMJ open*, 2016. **6**(8).
253. Farrell, G., A. Tseloni, and N. Tilley, *Signature dish: Triangulation from data signatures to examine the role of security in falling crime*. *Methodological Innovations*, 2016. **9**: p. 2059799115622754.
254. Porter, S., et al., *A critical realist evaluation of a music therapy intervention in palliative care*. *BMC Palliative Care*, 2017. **16**(1): p. 70.
255. McConnell, T., et al., *Evaluation of the effectiveness of music therapy in improving the quality of life of palliative care patients: a randomised controlled pilot and feasibility study*. *Pilot and feasibility studies*, 2016. **2**(1): p. 1-8.
256. Randell, R., et al., *A realist process evaluation of robot-assisted surgery: integration into routine practice and impacts on communication, collaboration and decision-making*. *Health Services and Delivery Research*, 2017. **5**(20).

257. Rycroft-Malone, J., et al., *A realist process evaluation within the Facilitating Implementation of Research Evidence (FIRE) cluster randomised controlled international trial: an exemplar*. Implementation Science, 2018. **13**(1): p. 138.
258. Byng, R., et al., *Exposing the key functions of a complex intervention for shared care in mental health: case study of a process evaluation*. BMC Health Services Research, 2008. **8**(1): p. 274.
259. Bonell, C., et al., *Are randomised controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services*. Trials, 2018. **19**(1): p. 238.
260. Green, J. and N. Thorogood, *Qualitative methods for health research*. 2013: Sage.
261. Pearce, W. and S. Raman, *The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance*. Policy sciences, 2014. **47**(4): p. 387-402.
262. Rowe, M. and C. Oltmann, *Randomised controlled trials in educational research: Ontological and epistemological limitations*. African Journal of Health Professions Education, 2016. **8**(1): p. 6-8.
263. Greenhalgh, J. and A. Manzano, *Understanding 'context' in realist evaluation and synthesis*. International Journal of Social Research Methodology, 2021: p. 1-13.
264. Hume, D., *1739. A treatise of human nature*. London: John Noon, 1978.
265. Perez, C.C., *Invisible women: Exposing data bias in a world designed for men*. 2019: Random House.
266. Salas, X.R., *The ineffectiveness and unintended consequences of the public health war on obesity*. Canadian Journal of Public Health, 2015. **106**(2): p. E79.
267. Mahoney, J. *Beyond correlational analysis: Recent innovations in theory and method*. in *Sociological forum*. 2001. JSTOR.
268. Chen, H.-T. and P.H. Rossi, *The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science*. Social forces, 1980. **59**(1): p. 106-122.
269. Chen, H.-T. and P.H. Rossi, *Evaluating with sense: The theory-driven approach*. Evaluation review, 1983. **7**(3): p. 283-302.
270. Jamal, F., et al., *The three stages of building and testing mid-level theories in a realist RCT: a theoretical and methodological case-example*. Trials, 2015. **16**(1): p. 1.
271. Bonell, C., et al., *Methods don't make assumptions, researchers do: a response to Marchal et al*. Social science & medicine (1982), 2013. **94**: p. 81-82.
272. Bonell, C., et al., *Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al*. Trials, 2016. **17**(1): p. 478.
273. Harachi, T.W., et al., *Opening the black box: using process evaluation measures to assess implementation and theory building*. American journal of community psychology, 1999. **27**(5): p. 711-731.
274. Chen, H.T., *Theory-driven evaluations*. 1990: Sage.
275. Sayer, A., *Realism and social science*. 1999: Sage.
276. Popper, K., *The Poverty of Historicism*. Economica, 1944. **11**(42): p. 86-103.
277. Hill, A.B., *The environment and disease: association or causation?* 1965, Sage Publications.
278. Bonell, C., et al., *Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial*. The Lancet, 2018. **392**(10163): p. 2452-2464.
279. Fletcher, A., et al., *Realist complex intervention science: Applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions*. Evaluation, 2016. **22**(3): p. 286-303.
280. Banerjee, A.V. and E. Duflo, *The experimental approach to development economics*. Annu. Rev. Econ., 2009. **1**(1): p. 151-178.
281. Van Bavel, J.J., et al., *Contextual sensitivity in scientific reproducibility*. Proceedings of the National Academy of Sciences, 2016. **113**(23): p. 6454-6459.

282. Kerr, N.L., et al., *Addressing replicability concerns via adversarial collaboration: Discovering hidden moderators of the minimal intergroup discrimination effect*. Journal of Experimental Social Psychology, 2018. **78**: p. 66-76.
283. Go, J.C., *Bite-Sized lecture 05: The Critical Realist Theory of Ontological Emergence*, C.R.-A.P. Network, Editor. 2019.
284. Warren, E., et al., *Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention*. Trials, 2020. **21**(1): p. 1-14.
285. Nussbaum, M., *Aristotelian social democracy*, in *Liberalism and the Good*, G.M. Mara and H.S. Richardson, Editors. 1990, Routledge: London. p. 203-251.
286. Erikson, E.H., *Identity: Youth and crisis*. 1968: WW Norton & company.
287. Bernstein, B., *Class, codes and control: Applied studies towards a sociology of language*. Vol. 2. 2003: Psychology Press.
288. Bernstein, B., *Class, codes and control. Vol 3: Toward a theory of educational transmission*. 1975, Londres: Routledge and Kegan Paul.
289. Bond, L., et al., *A comparison of the Gatehouse Bullying Scale and the Peer Relations Questionnaire for students in secondary school*. Journal of School Health, 2007. **77**(2): p. 75-79.
290. Smith, D.J., *School experience and delinquency at ages 13 to 16*. 2006: University of Edinburgh, Centre for Law and Society Edinburgh.
291. Varni, J.W., et al., *The PedsQL™* 4.0 as a pediatric population health measure: feasibility, reliability, and validity*. Ambulatory pediatrics, 2003. **3**(6): p. 329-341.
292. Clarke, A., et al., *Warwick-Edinburgh Mental Well-being Scale (WEMWBS): validated for teenage school students in England and Scotland. A mixed methods assessment*. BMC public health, 2011. **11**(1): p. 487.
293. Goodman, R., *The Strengths and Difficulties Questionnaire: a research note*. Journal of child psychology and psychiatry, 1997. **38**(5): p. 581-586.
294. Bonell, C., et al., *Modifying the secondary school environment to reduce bullying and aggression: the INCLUSIVE cluster RCT*. Public Health Research, 2019. **7**(18): p. 1-164.
295. Jackson II, R.L. and M.A. Hogg, *Encyclopedia of identity*. Vol. 1. 2010: Sage.
296. Kant, I., *Critique of Pure Reason (translated and edited by Paul Guyer & Allen W. Wood)*. 1998.
297. Harrison, V.S., *Internal realism and the problem of religious diversity*. Philosophia, 2006. **34**(3): p. 287-301.
298. Go, J.C., *Bite-size lecture 01: Roy Bhaskar's notion of judgemental rationality*. 2019: Critical Realism-Asia Pacific Network.
299. Stones, R., *Sociological reasoning: Towards a post-modern sociology*. 1996: Macmillan International Higher Education.
300. Moffatt, S., et al., *Using quantitative and qualitative data in health services research—what happens when mixed method findings conflict?*[ISRCTN61522618]. BMC Health Services Research, 2006. **6**(1): p. 28.
301. Maxwell, J.A. and K. Mittapalli, *Realism as a stance for mixed methods research*. Handbook of mixed methods in social & behavioral research, 2010: p. 145-168.
302. Morgan, D.L., *Practical strategies for combining qualitative and quantitative methods: Applications to health research*. Qualitative health research, 1998. **8**(3): p. 362-376.
303. Manzano, A., *The craft of interviewing in realist evaluation*. Evaluation, 2016. **22**(3): p. 342-360.
304. Bonell, C., et al., *Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention*. J Epidemiol Community Health, 2019. **73**(5): p. 455-464.

305. Shackleton, N., et al., *A new measure of unhealthy school environments and its implications for critical assessments of health promotion in schools*. *Critical public health*, 2017. **27**(2): p. 248-262.
306. Melendez-Torres, G., et al., *Moderated mediation analyses to assess intervention mechanisms for impacts on victimisation, psycho-social problems and mental wellbeing: evidence from the INCLUSIVE realist randomized trial*. *Social Science & Medicine*, 2021: p. 113984.
307. Frymier, A.B., G.M. Shulman, and M. Houser, *The development of a learner empowerment measure*. *Communication education*, 1996. **45**(3): p. 181-199.
308. Maslach, C. and S.E. Jackson, *The measurement of experienced burnout*. *Journal of organizational behavior*, 1981. **2**(2): p. 99-113.
309. Farivar, S.S., W.E. Cunningham, and R.D. Hays, *Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V. 1*. *Health and quality of life outcomes*, 2007. **5**(1): p. 1-8.
310. Ltd., Q.I.P., *NVivo (Version 12)*. 2018.
311. Marchal, B., et al., *Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research*. *Evaluation*, 2012. **18**(2): p. 192-212.
312. Jagosh, J., et al., *Critical reflections on realist review: insights from customizing the methodology to the needs of participatory research assessment*. *Research Synthesis Methods*, 2014. **5**(2): p. 131-141.
313. Mukumbang, F.C., et al., *An exploration of group-based HIV/AIDS treatment and care models in Sub-Saharan Africa using a realist evaluation (Intervention-Context-Actor-Mechanism-Outcome) heuristic tool: a systematic review*. *Implementation Science*, 2017. **12**(1): p. 1-20.
314. Charmaz, K., *Constructing grounded theory: A practical guide through qualitative analysis*. 2006: Sage.
315. Oliver, C., *Critical realist grounded theory: A new approach for social work research*. *British Journal of Social Work*, 2012. **42**(2): p. 371-387.
316. Glaser, B.G. and A.L. Strauss, *The discovery of grounded theory: strategies for qualitative theory*. New Brunswick: Aldine Transaction, 1967.
317. Mead, G.H., *Mind, self and society*. Vol. 111. 1934: Chicago University of Chicago Press.
318. Geertz, C., *The interpretation of cultures: Selected essays*. Vol. 5019. 1973: Basic books.
319. Urquhart, C., *The evolving nature of grounded theory method: The case of the information systems discipline*. *The Sage handbook of grounded theory*, 2007: p. 339-359.
320. Charmaz, K., *Constructing grounded theory*. 2014: sage.
321. Glaser, B.G., *Doing grounded theory: Issues and discussions*. 1998: Sociology Press.
322. Schatzman, L., *Dimensional analysis: Outline in preciform*. Unpublished manuscript. University of California, San Francisco, 1980.
323. Schatzman, L., *Dimensional analysis: Notes on an alternative approach to the grounding of theory in qualitative research*. *Social organization and social process: Essays in honor of Anselm Strauss*, 1991: p. 303-314.
324. Melendez-Torres, G.J., *Substance use, situational characteristics and sexual outcomes in men who have sex with men*. 2014, Oxford University, UK.
325. Kools, S., et al., *Dimensional analysis: Broadening the conception of grounded theory*. *Qualitative Health Research*, 1996. **6**(3): p. 312-330.
326. Byng, R., I. Norman, and S. Redfern, *Using realistic evaluation to evaluate a practice-level intervention to improve primary healthcare for patients with long-term mental illness*. *Evaluation*, 2005. **11**(1): p. 69-93.
327. Marchal, B., M. Dedzo, and G. Kegels, *A realist evaluation of the management of a well-performing regional hospital in Ghana*. *BMC health services research*, 2010. **10**(1): p. 1-14.
328. Rycroft-Malone, J., et al., *A realistic evaluation: the case of protocol-based care*. *Implementation science*, 2010. **5**(1): p. 1-14.

329. Wand, T., K. White, and J. Patching, *Applying a realist (ic) framework to the evaluation of a new model of emergency department based mental health nursing practice*. *Nursing inquiry*, 2010. **17**(3): p. 231-239.
330. Wang, R. and J.H. Ware, *Detecting moderator effects using subgroup analyses*. *Prevention Science*, 2013. **14**(2): p. 111-120.
331. Kraemer, H.C., et al., *Mediators and moderators of treatment effects in randomized clinical trials*. *Archives of general psychiatry*, 2002. **59**(10): p. 877-883.
332. Emsley, R., G. Dunn, and I.R. White, *Mediation and moderation of treatment effects in randomised controlled trials of complex interventions*. *Statistical Methods in Medical Research*, 2010. **19**(3): p. 237-270.
333. Befani, B., S. Ledermann, and F. Sager, *Realistic evaluation and QCA: conceptual parallels and an empirical application*. *Evaluation*, 2007. **13**(2): p. 171-192.
334. Schneider, C.Q., *Realists and Idealists in QCA*. *Political Analysis*, 2018. **26**(2): p. 246-254.
335. Cress, D.M. and D.A. Snow, *The outcomes of homeless mobilization: The influence of organization, disruption, political mediation, and framing*. *American Journal of Sociology*, 2000. **105**(4): p. 1063-1104.
336. Thomas, J., A. O'Mara-Eves, and G. Brunton, *Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example*. *Systematic reviews*, 2014. **3**(1): p. 67.
337. Ragin, C.C., *Redesigning social inquiry: Fuzzy sets and beyond*. 2009: University of Chicago Press.
338. Cronqvist, L., *Tosmana*. Tool for Small-N Analysis [Computer Programme], Version, 2011. **1**(2.0).
339. Roig-Tierno, N., T.F. Gonzalez-Cruz, and J. Llopis-Martinez, *An overview of qualitative comparative analysis: A bibliometric analysis*. *Journal of Innovation & Knowledge*, 2017. **2**(1): p. 15-23.
340. Cilliers, P., *Complexity, deconstruction and relativism*. *Theory, culture & society*, 2005. **22**(5): p. 255-267.
341. CONDON, R.E., *Type III error*. *Archives of Surgery*, 1986. **121**(8): p. 877-878.
342. Schwartz, S. and K.M. Carpenter, *The right answer for the wrong question: consequences of type III error for public health research*. *American journal of public health*, 1999. **89**(8): p. 1175-1180.
343. Hawe, P., A. Shiell, and T. Riley, *Complex interventions: how "out of control" can a randomised controlled trial be?* *Penelope*, 2004. **328**: p. 1561-1563.
344. Toledo, L., et al., *Knowledge, attitudes, and experiences of HIV pre-exposure prophylaxis (PrEP) trial participants in Botswana*. *World journal of AIDS*, 2015. **5**(2): p. 10.
345. Meyers, K., et al., *Lessons for patient education around long-acting injectable PrEP: findings from a mixed-method study of phase II trial participants*. *AIDS and Behavior*, 2018. **22**(4): p. 1209-1216.
346. Bonell, C., G.J. Melendez-Torres, and E. Warren, *Methodological reflections using qualitative research to explore causal mechanisms of complex health interventions*. *Evaluation (in press)*, 2022.
347. Pawson, R., *Theorizing the interview*. *British Journal of Sociology*, 1996: p. 295-314.
348. Dey, I., *Grounding grounded theory: Guidelines for qualitative inquiry*. 1999. San Diego, CA: Academic.
349. Sager, F. and C. Andereggen, *Dealing with complex causality in realist synthesis: The promise of qualitative comparative analysis*. *American journal of evaluation*, 2012. **33**(1): p. 60-78.
350. Sager, F. and W. Schenkel, *Evaluation UVP*, F.a.L. Swiss Agency for the Environment, Editor. 2004: Bern.

351. Evans, C.B., M.W. Fraser, and K.L. Cotter, *The effectiveness of school-based bullying prevention programs: A systematic review*. *Aggression and Violent Behavior*, 2014. **19**(5): p. 532-544.
352. Cross, D., et al., *Impact of the Friendly Schools whole-school intervention on transition to secondary school and adolescent bullying behaviour*. *European Journal of Education*, 2018. **53**(4): p. 495-513.
353. Ragin, C.C., *Turning the tables: How case-oriented research challenges*. *Rethinking social inquiry: Diverse tools, shared standards*, 2004. **123**.
354. Ragin, C.C. and H.S. Becker, *What is a case?: exploring the foundations of social inquiry*. 1992: Cambridge university press.
355. Melendez-Torres, G., et al., *Developing and testing intervention theory by incorporating a views synthesis into a qualitative comparative analysis of intervention effectiveness*. *Research synthesis methods*, 2019. **10**(3): p. 389-397.
356. Rihoux, B. and C.C. Ragin, *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. 2008: Sage Publications.
357. Kien, C., et al., *Pathways leading to success and non-success: a process evaluation of a cluster randomized physical activity health promotion program applying fuzzy-set qualitative comparative analysis*. *BMC public health*, 2018. **18**(1): p. 1386.
358. Sawyer, M.G., et al., *School-based prevention of depression: a randomised controlled study of the beyondblue schools research initiative*. *Journal of Child Psychology and Psychiatry*, 2010. **51**(2): p. 199-209.
359. Pratchett, L., et al., *Preventing Support for Violent Extremism through Community Interventions: A Review of the Evidence-Rapid Evidence Assessment Full Final Report*. 2009.
360. Melendez-Torres, G., et al., *Weight management programmes: re-analysis of a systematic review to identify pathways to effectiveness*. *Health Expectations*, 2018. **21**(3): p. 574-584.
361. Marx, A. and A. Dusa, *Crisp-set qualitative comparative analysis (csQCA), contradictions and consistency benchmarks for model specification*. *Methodological innovations online*, 2011. **6**(2): p. 103-148.
362. Bonell, C., et al., *Will it work here? A realist approach to local decisions about implementing interventions evaluated as effective elsewhere*. *J Epidemiol Community Health*, 2020.
363. Short, K., P. Eadie, and L. Kemp, *Influential factor combinations leading to language outcomes following a home visiting intervention: A qualitative comparative analysis (QCA)*. *International Journal of Language & Communication Disorders*, 2020. **55**(6): p. 936-954.
364. Education, D.f., *The prevent duty: Departmental advice for schools and childcare providers*. Department of Education (UK), 2015: p. 1-11.
365. Exline, J.J., et al., *Forgiveness and justice: A research agenda for social and personality psychology*. *Personality and social psychology Review*, 2003. **7**(4): p. 337-348.
366. Bonell, C., et al., *Role theory of schools and adolescent health*. *The Lancet Child & Adolescent Health*, 2019.
367. Ouellet-Morin, I., et al., *Increased serotonin transporter gene (SERT) DNA methylation is associated with bullying victimization and blunted cortisol response to stress in childhood: a longitudinal study of discordant monozygotic twins*. *Psychological medicine*, 2013. **43**(9): p. 1813-1823.
368. Westhorp, G., *Developing complexity-consistent theory in a realist investigation*. *Evaluation*, 2013. **19**(4): p. 364-382.
369. Westhorp, G., *Using complexity-consistent theory for evaluating complex systems*. *Evaluation*, 2012. **18**(4): p. 405-420.
370. Bonell, C., et al., *Why schools should promote students' health and wellbeing*. *Bmj*, 2014. **348**(7958): p. g3078.

