

Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): a pilot randomised controlled trial

Chris Bonell, Adam Fletcher, Natasha Fitzgerald-Yau, Daniel Hale, Elizabeth Allen, Diana Elbourne, Rebecca Jones, Lyndal Bond, Meg Wiggins, Alec Miners, Rosa Legood, Stephen Scott, Deborah Christie and Russell Viner



**National Institute for
Health Research**

Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): a pilot randomised controlled trial

Chris Bonell,¹ Adam Fletcher,² Natasha Fitzgerald-Yau,³ Daniel Hale,³ Elizabeth Allen,⁴ Diana Elbourne,⁴ Rebecca Jones,⁴ Lyndal Bond,⁵ Meg Wiggins,⁶ Alec Miners,⁷ Rosa Legood,⁷ Stephen Scott,⁸ Deborah Christie³ and Russell Viner^{3*}

¹Institute of Education, University College London, London, UK

²Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement (DECIPHER), Cardiff University, Cardiff, UK

³Institute of Child Health, University College London, London, UK

⁴Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

⁵Medical Research Council/Chief Scientist Office Social and Public Health Sciences, Glasgow, UK

⁶Social Science Research Unit, Institute of Education, University College London, London, UK

⁷Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

⁸Department of Child and Adolescent Psychiatry, Kings College London, London, UK

*Corresponding author

Declared competing interests of authors: Deborah Christie has received funding from the pharmaceutical industry for consultancy and lecturing.

Published July 2015

DOI: 10.3310/hta19530

This report should be referenced as follows:

Bonell C, Fletcher A, Fitzgerald-Yau N, Hale D, Allen E, Elbourne D, *et al*. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): a pilot randomised controlled trial. *Health Technol Assess* 2015;**19**(53).

Health Technology Assessment is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.116

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nhredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: <http://www.nets.nihr.ac.uk/programmes/hta>

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 09/05/05. The contractual start date was in July 2011. The draft report began editorial review in March 2013 and was accepted for publication in June 2013. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

© Queen's Printer and Controller of HMSO 2015. This work was produced by Bonell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Editor-in-Chief of *Health Technology Assessment* and NIHR Journals Library

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the HTA Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andree Le May Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Professor Aileen Clarke Professor of Public Health and Health Services Research, Warwick Medical School, University of Warwick, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Peter Davidson Director of NETSCC, HTA, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Professor Elaine McColl Director, Newcastle Clinical Trials Unit, Institute of Health and Society, Newcastle University, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Health Sciences Research, Faculty of Education, University of Winchester, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Institute of Child Health, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: nihredit@southampton.ac.uk

Abstract

Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): a pilot randomised controlled trial

Chris Bonell,¹ Adam Fletcher,² Natasha Fitzgerald-Yau,³ Daniel Hale,³ Elizabeth Allen,⁴ Diana Elbourne,⁴ Rebecca Jones,⁴ Lyndal Bond,⁵ Meg Wiggins,⁶ Alec Miners,⁷ Rosa Legood,⁷ Stephen Scott,⁸ Deborah Christie³ and Russell Viner^{3*}

¹Institute of Education, University College London, London, UK

²Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement (DECIPHer), Cardiff University, Cardiff, UK

³Institute of Child Health, University College London, London, UK

⁴Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

⁵Medical Research Council/Chief Scientist Office Social and Public Health Sciences, Glasgow, UK

⁶Social Science Research Unit, Institute of Education, University College London, London, UK

⁷Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

⁸Department of Child and Adolescent Psychiatry, Kings College London, London, UK

*Corresponding author r.viner@ucl.ac.uk

Background: Youth bullying and other aggressive behaviours are a major public health concern owing to their impact on adolescent physical and mental health and well-being. Whole-school restorative approaches have been identified as a promising method of addressing aggressive behaviour but there have been no randomised trials undertaken to examine their effects.

Aim: To examine the feasibility and acceptability of implementing and trialling the INCLUSIVE (initiating change locally in bullying and aggression through the school environment) intervention in English secondary schools.

Design: Cluster randomised controlled pilot trial in eight schools (1 : 1 computer-generated random allocation post baseline by a statistician blind to the identity of clusters) and process evaluation.

Setting: Secondary schools in England (purposively sampled to ensure diversity).

Participants: Year 8 students (aged 12–13 years), teachers, other school staff and intervention providers.

Intervention: Whole-school restorative approach to address bullying and aggression, involving the following standard processes: school action group formation and external facilitation to review needs assessment data, identify priorities, and plan and monitor school-level actions; staff training in restorative practices; and a new social and emotional skills curriculum.

Comparison group: Standard practice.

Main outcome measures: (1) The primary outcome of interest was the feasibility and acceptability of delivering and trialling the intervention according to prespecified criteria; (2) process data were analysed to explore participants' experiences of implementing and trialling the intervention and how these varied according to school context; and (3) indicative primary outcomes (aggressive behaviour measures), secondary outcomes, intermediate outcomes and economic evaluation methods were piloted.

Data sources: Students ($n = 1144$ baseline; $n = 1114$ follow-up) and teachers ($n = 387$ baseline; $n = 336$ follow-up) were surveyed at the start and end of the 2011–12 academic year (baseline September 2011; follow-up June–July 2012). A total of 1017 students surveyed at baseline remained in the study at follow-up (89%). Other quantitative data were collected via intervention provider checklists ($n = 4$) and action group surveys ($n = 44$); qualitative data were collected via interviews ($n = 34$), focus groups ($n = 20$) and observations of action group meetings ($n = 16$).

Results: (1) All prespecified feasibility and acceptability criteria were met. (2) Qualitative data indicated that all intervention components and the trial design were feasible and acceptable to students and staff, including in more disadvantaged school contexts. Qualitative data also suggested that student participation may be a core component in improving relationships and engagement across the school. The later-than-planned project start (July) and the timing of the baseline surveys (September), which needed to be completed pre allocation, caused delays in launching the intervention, staff training and other intervention outputs. (3) Three pilot primary outcomes were examined (completion rate at follow-up range: 91.7–94.2%) and the Gatehouse Bullying Scale and the Edinburgh Study of Youth Transitions and Crime school misbehaviour subscale were acceptable, discriminating and reliable measures of bullying and aggression in this context. Our pilot economic analyses support the use of the Child Health Utility 9D scale with this population and the feasibility of cost–utility analysis, although this should be supplemented with a cost–consequence analysis. There was no evidence of harm.

Conclusions: It is feasible and acceptable to implement and trial the INCLUSIVE intervention in English secondary schools, although a longer lead-in time is required to enable timely intervention outputs to occur. A Phase III cluster randomised controlled trial is required to examine the effectiveness and cost-effectiveness over a 3-year period of implementation for reducing aggressive behaviours, promoting mental health and well-being, and reducing health inequalities.

Trial registration: Current Controlled Trials ISRCTN88527078.

Funding: The National Institute for Health Research Health Technology Assessment programme (research), the Paul Hamlyn Foundation, the Big Lottery Fund and the Coutts Charitable Trust (intervention). The report will be published in full in *Health Technology Assessment*; Vol. 19, No. 53. See the NIHR Journals Library website for further project information.

Contents

List of tables	xiii
List of figures	xv
List of boxes	xvii
List of abbreviations	xix
Plain English summary	xxi
Scientific summary	xxiii
Chapter 1 Introduction	1
Aggressive behaviours among young people: a public health priority	1
Policy responses to aggression and bullying among young people	2
Effective school-based interventions to address bullying and aggression	2
Restorative approaches	3
Health Technology Assessment commission	4
INCLUSIVE intervention	6
Involvement of young people	6
Study aim, objectives and research questions	8
Chapter 2 Pilot trial design	11
Overview of study design	11
Sampling and recruitment	11
Comparison group	13
Health Technology Assessment progression criteria	13
<i>Data sources</i>	13
<i>Data analysis methods</i>	14
Process evaluation: participants' experiences	15
<i>Qualitative process data</i>	15
<i>Quantitative process data</i>	16
<i>Data analysis</i>	16
Piloting of outcome measures	16
<i>Primary outcome measure development</i>	17
<i>Pilot secondary outcome measures</i>	18
<i>Pilot intermediate outcome measures</i>	19
<i>Analysis plan for pilot outcome measures</i>	19
Economic evaluation	20
<i>Validation of the Child Health Utility 9D questionnaire</i>	20
<i>Developing a framework for economic evaluation</i>	21
Trial registration, governance and ethics	22

Chapter 3 Undertaking the trial	23
Recruitment of schools	23
Baseline student and staff surveys	23
Randomisation process	24
Follow-up student and staff surveys	24
Intervention provider checklists and other documentary evidence	25
Qualitative data collection	25
Action group member survey	26
Public involvement by young people during the intervention	26
Chapter 4 Results	27
Description of pilot trial sample	27
<i>Flow of participants in the pilot trial</i>	27
<i>Student survey response rates</i>	28
<i>School characteristics</i>	28
<i>Student characteristics</i>	28
Health Technology Assessment progression criteria assessment	31
<i>Intervention feasibility and acceptability</i>	31
<i>Trial feasibility and acceptability</i>	34
Process evaluation: participants' experiences	35
<i>Intervention feasibility and acceptability: key themes</i>	35
<i>Implementation of intervention components and contextual variation</i>	42
<i>Quantitative process measures</i>	54
<i>Research design and methods acceptability: key themes</i>	56
Development and piloting of indicative primary outcome measures	60
<i>Development of primary outcome measures</i>	60
<i>Data on pilot primary outcome measures</i>	68
Pilot secondary outcomes	70
Pilot intermediate outcome measures	70
Economic evaluation	70
<i>Validation of the Child Health Utility 9D questionnaire</i>	70
Chapter 5 Discussion	75
Key findings	75
Strengths	76
Limitations	77
Comparison with the literature	77
Developing a framework for economic evaluation for a full trial	78
<i>Choice of outcome measure and analytic methods</i>	78
<i>Which health effects should be included?</i>	79
<i>A public sector cost perspective focused on promoting equality</i>	79
Chapter 6 Conclusion and recommendations for further research	81
Progression to a Phase III trial	81
Intervention design and delivery	81
<i>Whole-school approach</i>	81
<i>Needs assessment survey</i>	82
<i>Action groups</i>	82
<i>External facilitators</i>	82
<i>Staff training in restorative practices</i>	82
<i>Social and emotional skills curriculum</i>	83
<i>New intervention partnerships</i>	83

Trial design and methods	83
<i>Primary outcomes</i>	83
<i>Secondary outcomes</i>	83
<i>Sample of schools</i>	83
<i>Surveys</i>	84
<i>Quantitative data on intervention fidelity</i>	84
<i>Qualitative data</i>	84
<i>Economic evaluation</i>	85
Scalability and generalisability	85
Acknowledgements	87
References	89
Appendix 1 Trial Steering Committee membership	97
Appendix 2 Pilot secondary outcome analyses	99
Appendix 3 Pilot intermediate outcome analyses	103
Appendix 4 Teacher survey response rates	105

List of tables

TABLE 1 School characteristics	29
TABLE 2 Sociodemographic characteristics of students at baseline	30
TABLE 3 Intervention feasibility and acceptability criteria (RQ1 and RQ2)	32
TABLE 4 Action group member survey: respondent characteristics	33
TABLE 5 Action group member survey: participation and experiences	34
TABLE 6 Student experiences of restorative practices and PSHE	55
TABLE 7 Teacher experiences of restorative practices, behaviour management and PSHE	56
TABLE 8 Missing data for pilot primary outcomes by sex	60
TABLE 9 Pilot primary outcome response rates by sex at baseline and follow-up	61
TABLE 10 Pilot primary outcome overall scores at baseline	67
TABLE 11 Reliability statistics for pilot primary outcomes	68
TABLE 12 Pilot primary outcomes [mean (SD)] at baseline by arm and overall	69
TABLE 13 Pilot primary outcomes [mean (SD)] at baseline by arm stratified by sex	69
TABLE 14 Pilot primary outcomes [mean (SD)] at follow-up by arm and overall	70
TABLE 15 Baseline CHU-9D responses from all participants by domain	71
TABLE 16 Correlations (<i>p</i> -value) between the CHU-9D utility score and other outcome measures	72
TABLE 17 Assessing convergent validity (<i>p</i> -value) (statistical tests are two tailed)	73
TABLE 18 Sensitivity of the CHU-9D utility scores to changes in other outcome measures	73
TABLE 19 Pilot secondary outcomes at baseline	100
TABLE 20 Pilot secondary outcomes at follow-up	101
TABLE 21 Pilot intermediate measures at baseline	103
TABLE 22 Pilot intermediate measures at follow-up	103
TABLE 23 Teacher response rates at baseline and follow-up	105

List of figures

FIGURE 1 The INCLUSIVE logic model	5
FIGURE 2 Sampling and matching criteria for schools ($n = 8$)	12
FIGURE 3 Flow chart of recruitment and participation of schools and students	27
FIGURE 4 Student survey response rates by school at baseline and follow-up	28
FIGURE 5 Schools according to sampling and matching criteria ($n = 8$)	28
FIGURE 6 Distribution of CHU-9D utility scores	72

List of boxes

BOX 1 The INCLUSIVE intervention inputs	7
BOX 2 Research questions addressing the agreed criteria for progression to a full trial	13
BOX 3 Case study of student-led action at Goldstone Park	48

List of abbreviations

AAYP	Aban Aya Youth Project	NIHR	National Institute for Health Research
BBSCQ	Beyond Blue School Climate Questionnaire	PedsQL	Paediatric Quality of Life Inventory
CCA	cost–consequence analysis	PSHE	personal, social and health education
CHU-9D	Child Health Utility 9D	QALY	quality-adjusted life-year
CUA	cost–utility analysis	QoL	quality of life
EQ-5D	European Quality of Life-5 Dimensions	RCT	randomised controlled trial
ESYTC	Edinburgh Study of Youth Transitions and Crime	RQ	research question
FSM	free school meals	SAE	stamped addressed envelope
GBS	Gatehouse Bullying Scale	SDQ	Strengths and Difficulties Questionnaire
HSE	Healthy School Ethos	SMT	senior management team
HTA	Health Technology Assessment	SRD	self-reported delinquency
ICC	intraclass correlation	SWEMWBS	Short Warwick–Edinburgh Mental Well-Being Scale
LSHTM	London School of Hygiene and Tropical Medicine	TSC	Trial Steering Committee
MRC	Medical Research Council	YRG	young researchers group
NCB	National Children’s Bureau		
NICE	National Institute for Health and Care Excellence		

Plain English summary

We examined 'INCLUSIVE' (initiating change locally in bullying and aggression through the school environment), a programme for secondary schools to reduce bullying and aggression among students. The programme involves several elements. First, staff and students work together to improve how the school is run, ensuring students have a say. Second, school staff manage discipline differently, focusing on repairing relationships rather than just punishing bad behaviour. Third, students are taught social and emotional skills, helping them to get on better together. In this study, INCLUSIVE was delivered only in four schools for 1 year, so we could not examine its impact. Instead, using questionnaires, interviews and focus groups with staff and students, we examined whether INCLUSIVE was feasible and acceptable. We found that it was. Schools improved their policies and let students have a say in decisions, altered how they managed discipline and taught students social and emotional skills. Local policy changes included new school rules that students felt were fair (and so were less likely to break) and timetables giving students more time with their form tutors (so problems with other students could be resolved more quickly). In terms of discipline, schools liked the greater focus on repairing relationships and thought this would reduce bullying. Staff and students also liked teaching and learning about emotional skills and thought this would also contribute to reducing bullying and aggression. We conclude that a larger study, which builds on the successful methods we used in this pilot study, to examine the impact of INCLUSIVE on bullying and aggression would be valuable.

Scientific summary

Background

Bullying and other aggressive behaviours among young people are a major public health concern because of their short- and long-term physical and mental health harms, and associated economic costs to the NHS and other areas of government. School-based interventions that combine changes to the school environment with the promotion of social and emotional skills have been identified as a promising approach for reducing rates of aggressive behaviours and other problem behaviours in trials carried out in the USA and Australia. A restorative approach to behaviour management in schools has also been found to be appropriate for improving relationships and supporting institutional-level cultural change to foster a more positive school ethos. However, no randomised trials have been undertaken to examine the effects of interventions that adopt a whole-school restorative approach.

Intervention

The INCLUSIVE (initiating change locally in bullying and aggression through the school environment) intervention is a whole-school restorative approach to behaviour change which aims to reduce bullying and aggression (primary outcome) and promote mental and emotional well-being and other (secondary) outcomes. It is informed by Markham and Aveyard's theory of human functioning and school organisation (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**:1209–20) and evidence from systematic reviews and trials in the USA and Australia. INCLUSIVE is a school-led intervention that combines changes to the school environment with the promotion of social and emotional skills and restorative practices through the formation of a school action group, involving students alongside staff (supported by an external facilitator), to review needs assessment data, determine priorities, and develop and implement an action plan for changing the school environment to improve relationships at the school and reduce aggression; whole-school staff training in restorative practices; and a new social and emotional skills curriculum for year 8 students (age 12–13 years). The intervention combines prespecified inputs, processes and outputs with the capacity for tailoring some elements to institutional needs and local ownership via decision-making by staff and students in the action group, which is informed by detailed needs assessment data reports. There is an emerging consensus that approaches that combine fidelity of core processes and components with some local adaptation are an appropriate means of delivering complex public health interventions.

Study aim, objectives and research questions

The aim of the pilot trial was to examine the feasibility and acceptability of implementing and trialling this intervention in English secondary schools. The study had three objectives.

The first objective was to assess whether prespecified feasibility and acceptability criteria were met, which were agreed with the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) co-ordinating centre and deemed necessary conditions for progressing to a Phase III trial. In order to meet this objective we collected and analysed data to address the following research questions (RQs):

- RQ1: was it feasible to implement the intervention in (at least) three out of four intervention schools? (This was assessed according to the following implementation criteria: the needs assessment survey had a $\geq 80\%$ response rate; the action group met six or more times during the course of the school year and was always quorate; the action group reviewed and revised school policies; whole-school actions were a collaborative process involving staff and students from across the school; peer mediation

- and/or 'buddying' schemes were reviewed and enhanced; ≥ 20 staff completed restorative practice training; restorative practices were used; and the student curriculum was delivered to year 8 students.)
- RQ2: was the intervention acceptable to a majority of school senior management team (SMT) members and a majority of action group members?
 - RQ3: did randomisation occur and was this acceptable to the school SMT?
 - RQ4: did (at least) three out of four schools from each of the intervention and comparison arms accept randomisation and continue to participate in the study?
 - RQ5: were the student survey response rates acceptable in (at least) three out of four comparison schools?

The second objective was to explore students', school staffs' and facilitators' experiences of implementing and trialling the INCLUSIVE intervention and how these varied across the different school contexts in order to refine the intervention and trial methods. In order to meet this objective, we collected and analysed data to address the following RQs:

- RQ6: what are students', school staff members' and intervention facilitators' experiences of the intervention, particularly in terms of whether it is feasible and acceptable?
- RQ7: how successfully was each component implemented and did this vary according to school context?
- RQ8: how acceptable were the research design and data collection methods to students and staff?

The third objective was to pilot and field test potential primary, secondary and intermediate outcome measures and economic methods prior to a Phase III trial. It was not an objective of the pilot study to assess intervention effects, and nor was it powered to do so, but data were collected and analysed to address the following RQs:

- RQ9: which of the pilot indicative primary outcome measures of aggressive behaviour performs best according to completion rate, discrimination and reliability statistics?
- RQ10: was it feasible and acceptable to collect data at baseline and follow-up on pilot indicative primary (aggressive behaviour), secondary (quality of life, psychological distress, psychological well-being, health-risk behaviours, NHS use, contact with the police, truancy and school exclusion) and intermediate outcome measures (students' perception of the school environment and connection to the school)?
- RQ11: is it feasible to measure year 8 students' health utility status using the Child Health Utility 9D (CHU-9D) measure and to embed an economic evaluation within a Phase III trial?

Methods

A cluster randomised controlled pilot trial (Medical Research Council Phase II exploratory trial) was undertaken in eight secondary schools in London and south-east England with allocation to the INCLUSIVE intervention (four schools) or continuation of normal practice (four schools). In order to assess the feasibility and acceptability of delivering and trialling the intervention according to prespecified criteria (*objective 1*), we collected a range of quantitative and qualitative data via pilot surveys of year 8 students (aged 12–13 years); teachers and action group members; intervention provider checklists; semistructured interviews with school SMT members; observations of action group meetings; and documentary evidence (e.g. policy documents, etc.). To explore participants' experiences of implementing and trialling the INCLUSIVE intervention (*objective 2*), semistructured interviews with schools' SMT members, action group members and intervention providers, as well as focus groups and surveys with students and teachers, were also employed. Although the primary objective of the study was to examine feasibility and acceptability, and this trial did not aim to (and therefore was not powered to) study intervention effects, we also collected data on indicative primary, secondary and intermediate outcome measures through student baseline and follow-up surveys in order to pilot these measures and explore potential economic analysis methods (*objective 3*).

Results

The INCLUSIVE intervention was feasible to implement and acceptable to school staff and students at four secondary schools with varying levels of deprivation and a range of Ofsted ratings of school effectiveness.

The first objective of the pilot trial was to assess whether the prespecified NIHR HTA feasibility and acceptability criteria were met. All these criteria were met in full. At all four intervention schools, the needs assessment survey had an > 80% response rate (range 91–97%); an action group was formed, which met at least six times during the year and revised school policies relating to aggression and its determinants, including through whole-school actions involving staff and students from across the school; > 20 staff completed restorative practice training (range 30–75); peer mediation and/or 'buddying' schemes were reviewed and enhanced in line with restorative principles, and restorative practices were introduced (e.g. 'circle time', 'checking in/out', etc.); and the student curriculum was delivered to all year 8 students (range 7–12 hours). The intervention was acceptable to school SMT members and student and staff action group members. The intervention was also highly rated by both student and staff action group members who were surveyed (e.g. > 93% of action group members surveyed felt it was a good way to ensure students contributed to decision-making at their school). Furthermore, trial methods were acceptable to all SMT members, randomisation occurred as planned and no schools dropped out of the study, with follow-up student survey response rates of 91–94% in the intervention schools and 87–96% in comparison schools.

The second objective was to explore students', school staffs' and facilitators' experiences of implementing and trialling the INCLUSIVE intervention in order to optimise the intervention and refine evaluation methods prior to a subsequent Phase III trial. Qualitative data indicated that all intervention components were acceptable. Students consistently reported that the intervention inputs (i.e. the needs assessment survey, action group, external facilitator, staff training and curriculum components) and restorative practices were appropriate and acceptable to them. The intervention's emphasis on 'having a say' and 'having respect' was a strong source of acceptability among students. Changes enacted included new school rules that students felt were fair (and so were less likely to break), and new timetables giving students more time with their form tutors (so problems with other students could be resolved more quickly). Qualitative data also suggested student participation may be a core component in improving relationships and engagement across schools.

Among school staff, the intervention tapped into a strong desire for locally adaptable processes for improving staff–student relationships and addressing aggressive behaviour. SMT members, particularly head teachers, reported that the intervention was a good 'fit' with schools' 'core business' of improving attainment and the national policy agenda focused on improving student behaviour. Although staff reported that some 'restorative-type' approaches were already being used in their school, this intervention was highly attractive because it could provide a new framework, process and resources for embedding restorative practices more consistently and more widely across the whole school. The extent to which the intervention balanced fidelity of standard processes and key components with scope for local participation and adaptation of some elements was seen as a strength. This intervention could therefore work across different school contexts, and our evaluation suggested that feasibility and acceptability were particularly strong in more disadvantaged and challenging schools.

The later-than-planned project start (July) and the timing of the baseline surveys (September), which needed to be completed pre allocation, caused delays in launching the intervention, scheduling and delivering staff training, and other intervention outputs. School staff and intervention providers consistently reported that a longer lead-in was needed, as well as improved co-ordination of activities and a longer duration of intervention (our intention always being that the intervention would be 3 years in a subsequent Phase III trial). School leaders noted that outputs did not begin until the second term and restorative conferencing did not begin to be implemented until the summer term owing to delays in staff training. However, our qualitative data consistently reported that, rather than reflecting any intrinsic

limitations of the intervention, these delays were caused by the delayed start date and the timing of the baseline surveys in this study. There was a consensus among staff and facilitators that major impacts on behaviour were likely but only in the subsequent school year, once policies and practices had been changed, and that a 2-year intervention was the minimum needed to expect a change in the school environment. Nonetheless, teachers at intervention schools were significantly more likely to report, at follow-up, being well supported with behaviour management by both senior members of staff and others implementing consistent behaviour management techniques. Staff reported that training could be further improved through a more comprehensive pre-training audit to identify schools' needs, greater use of interactive training methods and more 'realistic' examples from similar secondary schools.

The third objective was to pilot and field test potential primary, secondary and intermediate outcome measures and economic methods prior to a Phase III trial. This involved examining measures' completion, discrimination and reliability. The Gatehouse Bullying Scale and the Edinburgh Study of Youth Transitions and Crime school misbehaviour subscale were acceptable, discriminating and reliable measures of bullying and aggression. We found significant limitations with using the Aban Aya Youth Project questions to assess youth violence in the UK context owing to limited discrimination, and the potential for 'floor effects' and limited reliability. Completion rate at follow-up was high for all three pilot outcome measures (range 91.7–94.2%). Our pilot economic analyses support the use of the CHU-9D scale with this population and the feasibility of cost–utility analysis, although this should be supplemented with a cost–consequence analysis. The study did not aim to detect intervention effects, and lacked both the statistical power and the intervention duration to be able to do so. There was no evidence of harm.

Conclusion and recommendations

This 1-year pilot is the first evaluation of a whole-school restorative intervention that aims to initiate change locally in the school environment to address bullying and aggression. The INCLUSIVE intervention is feasible and acceptable to implement and trial in British secondary schools. The intervention has a strong potential to address health inequalities because of its feasibility and acceptability in more challenging schools with more disadvantaged students. It is also a potentially scalable and sustainable intervention owing to the method of combining fidelity of processes and core outputs with adaptation of some elements to local needs. A Phase III cluster trial is required to examine the effectiveness and cost-effectiveness of the INCLUSIVE intervention over a 3-year period for reducing aggressive behaviours, promoting mental health and well-being, and reducing health inequalities. Findings from the pilot will be used to improve elements of the delivery of the intervention, particularly regarding training and timescales. The intervention's relative effectiveness in different contexts and with different population subgroups should be examined in an adequately powered study involving a wide range of schools.

Trial registration

This trial is registered as ISRCTN88527078.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research (research), Paul Hamlyn Foundation, Big Lottery Fund and Coutts Charitable Trust (intervention).

Chapter 1 Introduction

Aggressive behaviours among young people: a public health priority

The prevalence, harms and costs of aggressive behaviours, such as bullying and violence, among young people make addressing them a public health priority.¹⁻⁴ The World Health Organization considers bullying to be a major adolescent health problem, defining it as the intentional and repetitive use of physical or psychological force against another individual or group.⁵ This includes verbal and relational aggression that aims to harm the victim or their social relations, such as through spreading rumours or purposely excluding them.^{6,7} The prevalence of bullying among British youth is above the European average,⁸ with approximately 25% of young people stating that they have been subjected to serious peer bullying.⁹

Being a victim of peer bullying is associated with an increased risk of physical health problems;¹⁰ engaging in health risk behaviours such as substance use;¹¹⁻¹³ long-term emotional, behavioural and mental health problems;¹⁴⁻¹⁶ self-harm and suicide;¹⁷ and poorer educational attainment.^{18,19} Students who experience physical, verbal and relational bullying on a regular basis tend to experience the most adverse health outcomes.²⁰ There is also emerging evidence suggesting that childhood exposure to bullying and aggression may also influence lifelong health through biological mechanisms, for example through reduced telomere length.²¹ The perpetrators of peer bullying are also at a greater risk of a range of adverse emotional and mental health outcomes, including depression and anxiety.^{8,22}

Although bullying is different from youth violence, which involves the intention of *physical force*, bullying is often a precursor to more serious violent behaviours commonly reported by British youth. For example, one UK study of 14,000 students found that 1 in 10 young people aged 11–12 years reported carrying a weapon and 8% of this age group admitted they had attacked another with the intention to hurt them seriously.²³ By age 15–16 years, 24% of students report that they have carried a weapon and 19% report attacking someone with the intention of hurting them seriously.²³ Interpersonal violence can cause physical injury and disability, and is also associated with long-term emotional and mental health problems. There are also links between aggression and antisocial behaviours in youth and violent crime in adulthood.^{24,25} Although all bullying does not necessarily involve or lead to violence, and youth violence may emerge in isolation from bullying or victimisation, these different aggressive behaviours share common determinants which make common universal approaches to address aggressive behaviours appropriate.^{14,22,23,25}

Although a universal problem, the prevalence of aggressive behaviours varies significantly between schools, both in the UK^{22,26} and internationally.^{27,28} For example, among 11- to 14-year-olds in London, the frequency of reporting being bullied at school has been found to vary from 21% to 47% across different schools, and the frequency of reporting being recently involved in physical violence from 22% to 52%.²⁶ There are also marked social gradients in bullying and violence among young people, with both family deprivation and school-level deprivation increasing the risk of experiencing bullying.²⁹ As well as leading to further health inequalities, the economic costs to society as a whole of youth aggression, bullying and violence are extremely high. For example, the total cost of crime attributable to conduct problems in childhood has been estimated at about £60B a year in England and Wales.³⁰

Policy responses to aggression and bullying among young people

Reducing aggression, bullying and violence in British schools has been a consistent priority within recent public health and education policies.^{31–33} Schools are at the ‘epicentre’ of peer bullying, as the site where it most often begins, and where children and young people are most concerned about bullying and victimisation.³⁴ The Home Office has found that assaults against 10- to 15-year-olds are more likely to happen at secondary school than anywhere else, with 95% of victims likely to know the perpetrator.³⁵ There is also increasing educational concern regarding low-level provoking and aggressive behaviours in secondary schools, which are educationally disruptive and emotionally harmful and can lead to more overt physical aggression over time.^{36–38}

In 2009, the Steer Review concluded that schools’ approaches to discipline, behaviour management and bullying prevention vary widely and have little or no evidence base, and further resources and research are urgently needed to combat aggressive behaviours and other conduct problems.³⁸ The award of National Healthy Schools status, and its equivalents in Wales, Scotland and Northern Ireland, had previously required that schools foster a ‘supportive’ and ‘positive’ school environment, although there has been no evidence-based guidance on how to achieve this³⁹ and it is no longer a statutory requirement in England. Furthermore, the national schools inspectorate, Ofsted, is no longer required to inspect schools on how they promote students’ health and well-being.⁴⁰ There is, therefore, an urgent need to determine which interventions are effective in addressing bullying and aggression in schools, and to scale up such interventions across local and national school networks.

The current British government has stated its commitment to greater intersectoral, joined-up action to promote youth development in its ‘Positive for Youth’ strategy, including through work in schools and reducing bullying and aggression among adolescents.⁴¹ There is also an increasing recognition that bullying and aggression can be serious barriers to academic attainment, and therefore schools must address these if they are to fulfil their core mission.¹⁸ Most recently, the cross-government response to the 2011 riots included a commitment to assess existing school-based interventions focused on youth violence and ensure schools know how to access the most effective methods and approaches.⁴²

Effective school-based interventions to address bullying and aggression

A number of reviews identify school-based interventions to address bullying and aggression that are effective. The effects of curriculum-only interventions are patchy and often not sustained.^{43–49} Although several reviews find insufficient studies to assess the effects of whole-school interventions, two report on the particular effectiveness of such approaches. First, Vreeman and Carroll’s review concluded that whole-school interventions addressing different levels of school organisation were more often effective in reducing victimisation and bullying than curriculum interventions.⁴⁶ The authors suggest this might be because these interventions address bullying as a systemic problem meriting an ‘environmental solution’. Smith and colleague’s review draws similar conclusions and suggests that whole-school interventions are likely to be universal in reach and a cost-effective and non-stigmatising approach to preventing bullying.⁴³

This is in keeping with other evidence from the UK and internationally, which shows that schools promote health most effectively when they are not treated merely as sites for health education but also as physical and social environments which can actively support healthy behaviours and outcomes.^{50,51} These new school environment interventions thus take a ‘socioecological’⁵² or ‘structural’⁵³ approach to promoting health, whereby behaviours are understood to be influenced not only by individual characteristics, but also the wider social context. A recent National Institute for Health Research (NIHR)-funded systematic review of the health effects of the school environment found evidence from observational and experimental studies that modifying the way in which schools manage their ‘core business’ of teaching, pastoral care and

discipline can promote student health and potentially reduce health inequalities across a range of outcomes, including reductions in violence and other aggressive behaviours, as well as improvements in various measures of mental health and physical activity and reductions in alcohol, tobacco and drug use.⁵¹

Two recent studies report on randomised controlled trial (RCT) studies of interventions that aim to enable schools to improve aspects of their core business in order to reduce aggression and other health risk behaviours. First, the Aban Aya Youth Project (AAYP) is a multicomponent intervention that enables schools to modify their social environment as well as deliver a social skills curriculum. This approach is designed to increase social inclusion by 'rebuilding the village' within schools serving disadvantaged, African-American communities. To promote whole-school institutional change at each school, teacher training was provided and an action group was established (comprising both staff and students) to review policies and prioritise actions needed to foster a more inclusive school climate. For boys, the intervention was associated with significant reductions in the growth of violence (47%) and aggressive behaviour (59%).⁵⁴ The intervention also brought benefits in terms of reductions in sexual risk behaviours and drug use, as well as provoking behaviour and school delinquency. Second, the Gatehouse Project in Australia also aimed to reduce health problems through changing the school climate and promoting security, positive regard and communication among students and school staff. As with the AAYP, an action group was convened in each school, facilitated by an external 'critical friend' and informed by data from a student survey, alongside a social and emotional skills curriculum. A cluster RCT found consistent reductions in a composite measure of health risk behaviours, which included violence and antisocial behaviour.^{55,56}

School environment interventions that impact on a range of health risk behaviours including aggression are likely to be one of the most efficient ways of addressing multiple health harms in adolescence because of their potential for modifying population-level risk, as well as their reach and sustainability.⁵¹ Multiple risk behaviours in adolescence are subject to socioeconomic stratification, and strongly associated with poor health outcomes, social exclusion, educational failure and poor mental health in adult life.⁵⁷ A recent King's Fund report on the *Clustering of Unhealthy Behaviours Over Time* emphasised the association of multiple risk behaviours with mortality and health across the life-course, and the policy importance of reducing multiple risk behaviours among young people through new interventions that address their common determinants.⁵⁸

Restorative approaches

The Steer Review called for English schools to consider adopting more 'restorative' approaches to prevent bullying and other aggressive behaviour, and to minimise the harms associated with such problems.³⁸ These approaches have been developed and used widely in the criminal justice system. Their central tenet is to repair the harms caused to relationships and communities rather than merely assign blame and enact punishment. Such approaches have now been adapted for use in schools and can operate at a whole-school level, informing changes to disciplinary policies, behaviour management practices and how staff communicate with students in order to improve relationships, reduce conflict and repair harm. An example of such restorative practices currently employed in schools is the use of circle time to develop and maintain good communication and relationships.⁵⁹ Restorative 'conferencing' can also be used in schools to deal with more serious incidents.⁵⁹

Restorative approaches have been evaluated only using non-random designs, although such studies do suggest that the restorative approach is a promising one in the UK⁶⁰⁻⁶² and internationally.⁶³⁻⁶⁵ For example, in England and Wales, the Youth Justice Board evaluated the use of restorative approaches at 20 secondary schools and six primary schools, and reported significant improvements regarding students' attitudes to bullying and reduced offending and victimisation in schools that adopted a whole-school approach to restorative practice. The study concluded that, when embedded in institutional cultural change, restorative principles and practices have the greatest potential. Similarly, a 2-year evaluation of

restorative practices in Scottish schools found that the schools that reported the greatest success were those that focused on addressing the school culture and promoting supportive relationships throughout the whole school community.⁶¹ A study carried out in Bristol, England, also concluded that when restorative approaches are implemented on a whole-school basis they can transform the institutional climate.⁶²

Restorative approaches thus appear to have the potential to complement school environment interventions such as the AAYP and the Gatehouse Project. They thus offer a highly promising way forward for reducing aggressive behaviours among British youth. A recent Cochrane review found no randomised trials of interventions employing restorative approaches to reduce bullying in schools and recommended that this should be a priority for future research.⁴⁹ If trialled and found to be effective, such a universal school-based approach could be scaled up to reach very large numbers of young people and deliver significant population-level health improvements. The Medical Research Council (MRC) guidance on complex interventions recommends that interventions are piloted to assess feasibility and acceptability, when these are uncertain, prior to Phase III randomised trials.⁶⁶ This is the case with health improvement interventions in British secondary schools, which can be particularly challenging because of schools' increasingly narrow focus on educational attainment and an inspection framework which largely ignores student well-being.^{67,68}

Health Technology Assessment commission

To address the public health problems and key research gaps outlined above, the Health Technology Assessment (HTA) research programme at the NIHR commissioned a phased evaluation of a universal school-based intervention to reduce aggression and bullying among British secondary school students in 2011.

The INCLUSIVE (initiating change locally in bullying and aggression through the school environment) intervention is a new school environment intervention that is informed by AAYP⁵⁴ and the Gatehouse Project,⁵⁵ as well as the Healthy School Ethos (HSE) project that was piloted in two English schools in 2007–8.⁶⁹ These interventions offer a systematic and scalable process by which schools are enabled not only to deliver a health promotion curriculum but also to modify how they manage their 'core business' of teaching, pastoral care and discipline in order to encourage a more health-promoting school environment. We modified these previous interventions and combined them with a restorative approach to school management, teaching practices and conflict resolution (*Figure 1*). The intervention logic model is informed by Markham and Aveyard's theory of human functioning and school organisation.⁷⁰ Although primarily focused on reducing aggression and bullying, our intervention is also hypothesised to promote student health more generally, including through reduced alcohol consumption and drug and tobacco use, reduced sexual risk behaviour and improved psychological well-being.

The HTA programme provided funding for 20 months for a pilot trial to assess the feasibility and acceptability of the INCLUSIVE intervention and trial methods over one academic year. Criteria were agreed for progression to a full trial, with further funding for a Phase III trial of a 3-year intervention being dependent on a new funding application. Research costs were provided by the HTA programme. As NHS support costs for the intervention were not available, because the intervention was being undertaken in schools, intervention funding was provided by the Paul Hamlyn Foundation, the Big Lottery Fund and the Coutts Charitable Trust. The research was funded for 20 months from July 2011 until February 2013. The project was planned to begin in April 2011; however, funding was not confirmed by the HTA programme and by our intervention funders until March 2011. Subsequent contracting issues meant that the official start of the research project was delayed until July 2011.

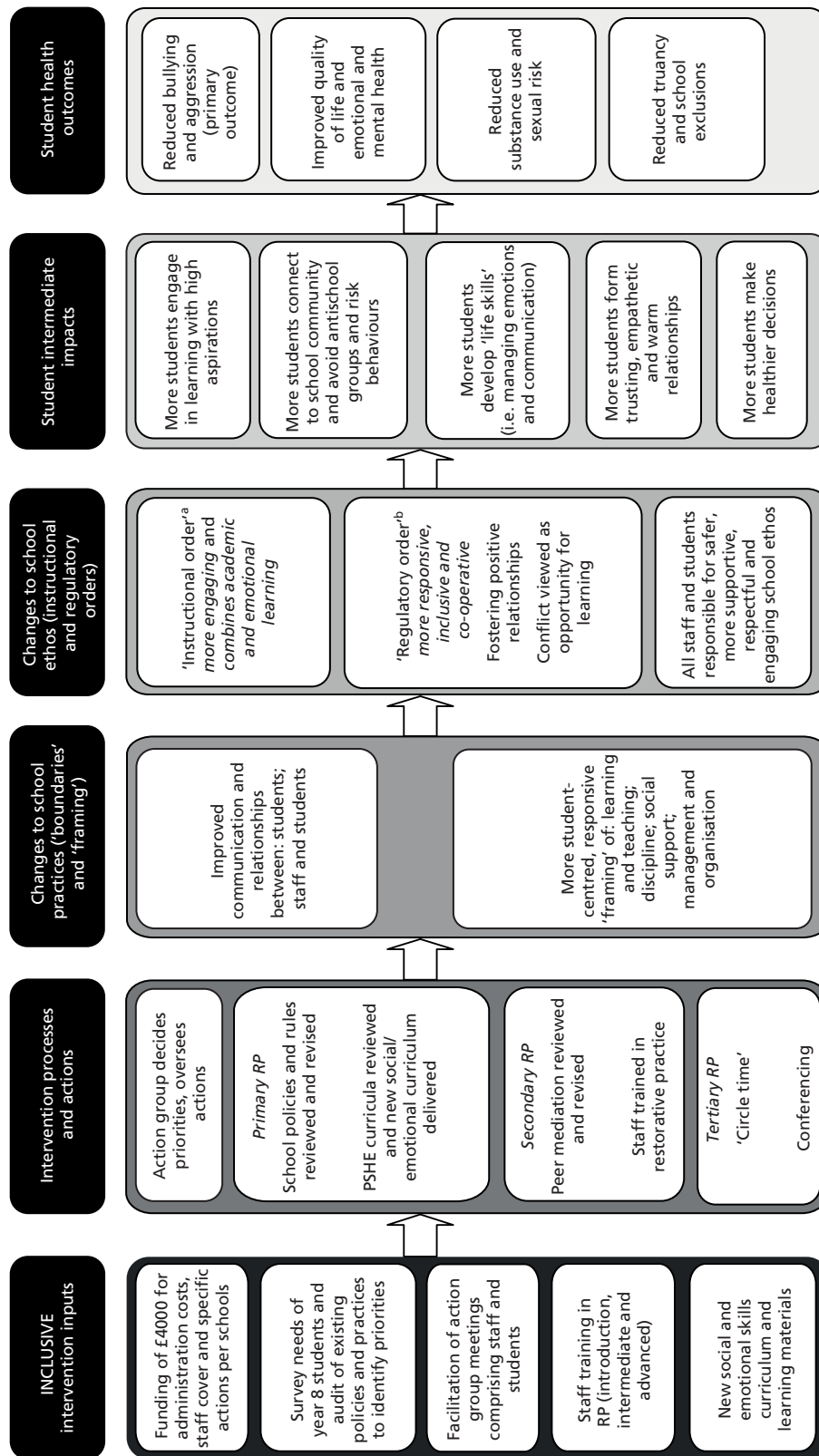


FIGURE 1 The INCLUSIVE logic model. PSHE, personal, social and health education; RP, restorative practice. a, i.e. learning and teaching in school; b, i.e. discipline, social support and sense of community in school.

INCLUSIVE intervention

The INCLUSIVE intervention involves the following inputs: £4000 of funding per school; a needs assessment survey; an external expert facilitator; staff training in restorative practices; and a social and emotional skills curriculum. (Box 1 has more details of each of these intervention inputs.) These inputs were intended to enable schools to convene an action group comprising students alongside staff and support that group in the processes of reviewing and revising school rules and policies relating to discipline, behaviour management and staff–student communication; reviewing and enhancing the school’s existing peer mediation and ‘buddying’ schemes via which they recruit and train students to assist their peers in resolving conflicts within the school environment before they escalate; implementing restorative practices; identifying local priorities; and delivering the six-module social and emotional skills curriculum. The action group comprised (at a minimum) six students and six staff, including at least one senior management team (SMT) member, and one member of each of the teaching, pastoral and support staff. Membership of specialist health staff, such as the school nurse and/or local child and adolescent mental health services staff, was desirable but optional. The action group was intended to meet at least six times over the school year (i.e. approximately once every half-term).

The intervention was specifically designed to enable local tailoring, informed by the needs assessment report and other local data sources. These locally adaptable actions occurred within a standardised overall process with various core standardised intervention elements, such as the whole-school staff training in restorative practices; review and revision of school rules and policies; and a new social and emotional skills curriculum for year 8 students (aged 12–13 years). This balance of standardisation and flexibility, which is common practice in complex interventions,^{66,71} enables a balance between fidelity of the core components with local adaptation,⁷² thus allowing schools to build on their current good practice, and also encouraging students and staff to develop ownership of the work, which may be a key factor in intervention effects.^{69,73} The facilitator worked with schools to ensure all members of the action group were supported in identifying and undertaking locally determined actions to improve the school environment.⁶⁹

This study examined the feasibility and acceptability of planning and delivering the intervention over 1 academic year only, as this was deemed sufficient time to pilot a new whole-school restorative approach. It was anticipated that a subsequent Phase III trial of effectiveness would involve a 3-year intervention. The first 2 years would involve an externally facilitated intervention (i.e. schools being provided with additional funding, facilitation and training, etc.) and the third year would involve schools continuing with the intervention unassisted to assess its sustainability in the absence of facilitation.

Involvement of young people

Young people from the National Children’s Bureau’s (NCB) young researchers group (YRG) were involved both in the development of the intervention and in the research protocol for the pilot study. This is a group of approximately 15 young people who receive research training from the NCB. This group includes young people with medical conditions, who have a range of experiences of health services, as well as young people from a diversity of social backgrounds. Two meetings with the NCB young researchers were held: one on 16 July 2011 and the other on 3 November 2011.

The first meeting (July 2011) was held before the intervention commenced and focused on the design of the intervention and baseline survey, in particular which aspects of the school environment might influence bullying and how we might assess this, as well as seeking the YRG’s advice more generally on how to engage young people in the research and ensure high survey completion rates. The YRG provided important insights into which aspects of their school environment they felt might influence levels of aggression. The group had a range of suggestions which fell into four broad categories: staff, sports/extracurricular, school policies and common areas/break times. The group consistently reported that staff should also be available to discuss private matters and respect students’ views. These issues were fed into

BOX 1 The INCLUSIVE intervention inputs

1. *Funding*: each school action group received £4000 to cover schools' administrative costs, provide cover for staff involvement and fund specific actions to support change (e.g. training and equipment for peer mediators, convening student 'away days' to revise school rules, etc.). This funding was provided in addition to the costs of the external facilitator, whose services were provided to intervention schools free of charge as part of this intervention. Allocation of the £4000 of funding was determined by the school action group, with financial responsibility taken by a SMT member of the action group.
2. *Needs assessment*: we used our baseline survey of students in year 8 to examine students' views of the school environment and their experiences of aggression and bullying (overall and by sex) and various known risk factors for aggression and bullying. Detailed reports were produced using these student survey data that presented tailored information for each intervention school on students' reports of the following: school engagement and 'connectedness'; perceptions of safety/risks; relationships, social support and social skills; and teaching and learning about social and emotional skills in personal, social and health education (PSHE) lessons. Data were 'benchmarked' against the average across all four intervention schools and these reports were used by school action groups to determine local priorities and inform decision-making regarding their actions to improve the school environment.
3. *Facilitator*: schools were supported by an expert facilitator (a freelance education consultant with previous secondary school leadership experience). This individual co-ordinated the intervention in each school and held preliminary meetings with the SMT, school staff and students to build interest, commitment and trust. The facilitator also supported the action group to ensure broad participation, critical reflection and effective delivery; organised the staff training (see below); assisted schools in integrating a restorative approach across existing school policies and practices; and worked with schools to adapt, and integrate the curriculum (see below) into, the school timetable and existing lesson plans; and provided school leaders with ongoing, informal support and feedback.
4. *Staff training in restorative practices*: the facilitator provided an introduction to restorative approaches (approximately 30–60 minutes) to all staff in the school, and a further half-day of training were delivered by a specialist training provider and offered to all staff (a minimum of 20 staff in each school were required to attend). This initial half-day of training focused on introducing them to restorative practices, such as circle time. Circle time is a technique that can be used in schools to support the development of young people's social and emotional literacy and communication skills to promote positive relationships. It typically involves students and teachers sitting together in a circle formation in order to share their ideas and feelings and to discuss other social, emotional or curricular issues. The central underlying principle of circle time is that all participants are treated as equals and valued for their unique contribution. An enhanced 3-day training course in restorative practices for secondary schools was also provided by the specialist training provider for 5–10 staff at each school, including training in formal 'conferencing' to deal with more serious incidents, such as violent incidents or serious bullying within school, through bringing together students, parents and staff. The aim of these school-based restorative conferences is to provide a safe, inclusive environment in which all individuals involved in a serious aggressive incident feel able to constructively and fully participate in the process of resolution to repair any harm which has occurred to relationships, as well as to identify appropriate forms of punishment. Through the use of a safe, inclusive conference environment facilitated by trained staff both the victim(s) and perpetrator(s) can participate.
5. *Curriculum*: students in year 8 received 5–10 hours' teaching on restorative practices, relationships, and social and emotional skills based on the Gatehouse Project curriculum. Our curriculum was devised as a set of learning modules which schools could address in locally decided ways, using either our teaching materials and methods or their own existing approaches if these aligned with our curriculum. The following modules were covered: establishing respectful relationships in the classroom and the wider school; managing emotions; understanding and building trusting relationships; exploring others' needs and avoiding conflict; and maintaining and repairing relationships. Informed by the needs assessment data, schools tailored the curriculum to their own needs and could deliver modules as 'stand-alone' lessons, for example within PSHE, and/or integrated into various subject lessons (e.g. English).

the intervention design, particularly through the needs assessment survey design and through facilitators, who worked with school action groups to focus on changing the school climate, and informed the baseline survey.

The second meeting (November 2011) coincided with the formation and initial meetings of the school action groups and, therefore, focused on exploring the young researchers' views on how to encourage students to participate in these action groups (see *Public involvement by young people during the intervention* for further information).

Study aim, objectives and research questions

The long-term goal of the investigator team is to examine the effectiveness and cost-effectiveness of the INCLUSIVE intervention for reducing bullying and other aggressive behaviours among young people in the UK. In line with current guidance concerning the evaluation of complex public health interventions,⁶⁶ the aim of this pilot trial was to examine the feasibility and acceptability of implementing and trialling this intervention in English secondary schools. We chose to do so in a challenging purposive sample of schools rated by the Ofsted school inspectorate as either satisfactory or good/outstanding, and with moderate or high rates of student entitlement to free school meals (FSM). The study had three objectives.

The first objective was to assess whether prespecified feasibility and acceptability criteria were met, which were agreed with the NIHR HTA programme co-ordinating centre and deemed necessary conditions for progressing to a Phase III trial. In order to meet this objective, we collected and analysed data to address the following research questions (RQs):

- RQ1: was it feasible to implement the intervention in (at least) three out of four intervention schools? (This was assessed according to the following implementation criteria: the needs assessment survey had an $\geq 80\%$ response rate; the action group met six or more times during the course of the school year and was always quorate; the action group reviewed and revised school policies; whole-school actions were a collaborative process involving staff and students from across the school; peer mediation and/or 'buddying' schemes were reviewed and enhanced; ≥ 20 staff completed restorative practice training; restorative practices were used; and the student curriculum was delivered to year 8 students.)
- RQ2: was the intervention acceptable to a majority of school SMT members and a majority of action group members?
- RQ3: did randomisation occur and was this acceptable to the school SMT?
- RQ4: did (at least) three out of four schools from each of the intervention and comparison arms accept randomisation and continue to participate in the study?
- RQ5: were the student survey response rates acceptable in (at least) three out of four comparison schools?

The second objective was to explore students', school staff members' and facilitators' experiences of implementing and trialling the INCLUSIVE intervention and how these varied across the different school contexts in order to refine the intervention/methods. In order to meet this objective, we collected and analysed data to address the following RQs:

- RQ6: what are students', school staff members' and intervention facilitators' experiences of the intervention, particularly in terms of whether it is feasible and acceptable?
- RQ7: how successfully was each component implemented and did this vary according to school context?
- RQ8: how acceptable were the research design and data collection methods to students and staff?

The third objective was to pilot and field test potential primary, secondary and intermediate outcome measures and economic methods prior to a Phase III trial. In order to meet this objective, we collected and analysed data to address the following RQs:

- RQ9: which of the pilot indicative primary outcome measures of aggressive behaviour performs best according to completion rate, discrimination and reliability statistics?
- RQ10: was it feasible and acceptable to collect data at baseline and follow-up on pilot indicative primary (aggressive behaviour), secondary (quality of life, psychological distress, psychological well-being, health-risk behaviours, NHS use, contact with the police, truancy and school exclusion) and intermediate outcome measures (students' perception of the school environment and connection to the school)?
- RQ11: is it feasible to measure year 8 students' health utility status using the Child Health Utility 9D (CHU-9D) measure and to embed an economic evaluation within a Phase III trial?

Chapter 2 Pilot trial design

In this chapter of the report we describe the study design and methods used to assess the HTA 'progression criteria' (objective 1), explore participants' experiences of the process of implementing and trialling the intervention (objective 2) and examine pilot trial outcomes and develop a framework for economic evaluation (objective 3), as well as provide details regarding trial registration, governance and ethics. The process of undertaking the trial and the methods of data collection are described in detail in *Chapter 3*.

Overview of study design

The intervention was piloted during the 2011–12 academic year (September 2011–July 2012). The study had a 1 : 1 allocation cluster RCT design, as recommended in the MRC complex intervention guidance for Phase II exploratory trials.⁶⁶ Following baseline surveys, four clusters (schools) were randomly allocated to the intervention arm and four to the comparison arm. The intervention was intended principally to augment, rather than to replace, existing activities (e.g. training, curricula, etc.) in intervention schools. However, it was intended to change existing non-restorative disciplinary school policies and practices when restorative approaches were deemed more appropriate. Comparison schools continued with their usual practices.

Quantitative and qualitative process evaluation data were collected and analysed to assess whether prespecified feasibility and acceptability criteria were met (see section *Health Technology Assessment progression criteria* below for more details). We also collected qualitative and quantitative process data to explore students', school staff members' and facilitators' experiences of implementing and trialling the INCLUSIVE intervention and how these varied across the different school contexts in order to refine the intervention and trial methods (see *Process evaluation: participants' experiences* below for more details). Although not intended to (and therefore not powered to) examine intervention effects, this study did provide the opportunity to pilot our primary, secondary and intermediate indicative outcome measures, and develop methods for a future economic evaluation (see sections *Piloting of outcome measures* and *Economic evaluation* below for more details).

Sampling and recruitment

Schools eligible to participate were mixed-sex, state secondary schools (including academies) in London and south-east England judged by the national schools inspectorate (Ofsted) as being 'satisfactory' or better and in which $\geq 6\%$ of students are eligible for FSM. Independent schools, single-sex schools and schools with unsatisfactory Ofsted ratings and/or schools in which $< 6\%$ of students were eligible for FSM (which represent the least economically deprived 15% of British schools) were not eligible for inclusion. We excluded schools rated as unsatisfactory because we judged that such schools, being subject to special measures interventions, would not be able to implement our intervention. We excluded schools in which $< 6\%$ of students were eligible for FSM because we aimed to test the feasibility and acceptability of the intervention in more challenging school contexts.

Eight schools were identified and recruited by the intervention facilitators between July and September 2011 (see *Recruitment of schools* for more detail). *Figure 2* explains the final sampling and matching criteria for schools in the study. The initial aim (at the funding application stage) was to identify a mix of schools with Ofsted ratings ranging from satisfactory to outstanding and above or below *the national average* for the number of students eligible for FSM (approximately 15% of secondary-school students). However, FSM eligibility is considerably higher than the national average in London (approximately 23%).⁴⁰ As the majority of schools were recruited from London boroughs and nearby counties, we revised the final

		Ofsted-rating of school effectiveness	
		'Satisfactory'	'Good'/'outstanding'
Eligibility for FSM in 2010–11	>6% and <23% of all students	Two schools	Two schools
	≥ 23% of all students	Two schools	Two schools

FIGURE 2 Sampling and matching criteria for schools ($n=8$).

sampling and matching criteria so that schools were divided by Ofsted rating and between those with *greater than or less than 23% student eligibility for FSM*, as a benchmark of greater/lesser economic deprivation in this region.

Recruiting and matching pairs of schools according to their FSM eligibility and Ofsted rating prior to randomisation was useful primarily because it allowed the intervention, trial methods and outcome measures to be piloted in a diverse and particularly challenging range of school contexts. It also enabled us to pilot our ability to use such techniques to increase baseline equivalence between arms within a Phase III trial, but we recognised that even with matching we were unlikely to achieve baseline equivalence in a small pilot study.

Although the intervention aimed to address the whole school, only students in year 8 (aged 12 and 13 years) at each participating school were recruited as the analytical sample, which participated in baseline and follow-up surveys. No other student inclusion or exclusion criteria operated. The sample size for the student baseline and follow-up surveys was estimated to be approximately 1200 students (approximately 150 year 8 students per school). No power calculation was performed, as this was a pilot study and the research objective was to evaluate the feasibility and acceptability of the intervention and trial methods, and not to estimate intervention effects. All teaching staff at the participating schools were also invited to participate in baseline and follow-up surveys. Non-teaching school staff were not eligible for inclusion in these surveys. Action team members were also invited to participate in surveys post intervention which focused on their experiences of the intervention. Although the surveys of teachers and action team members were not included in the funding application, they were considered useful to pilot in preparation for a Phase III trial and were therefore included in our trial protocol.

Subgroups of year 8 students and school staff were also recruited to take part in focus groups and individual interviews. At each of the four intervention schools, four groups of year 8 students and one group of school staff were recruited to participate in focus group discussions. Students were sampled purposively, and grouped with similar peers, according to their sex and level of school engagement (as reported by staff), to ensure a range of views and school experiences. School staff members were purposively sampled according to their sex, experience and role at the school to ensure a range of staff was represented. A maximum of 10 participants were recruited per focus group by researchers working with a member of the SMT. Intervention action group members (three to seven per school) and SMT staff (one or two per school) were also recruited to take part in semistructured interviews at the end of the school year. All INCLUSIVE facilitators working in schools ($n=3$), one of whom co-ordinated the intervention, and the staff training provider ($n=1$) were also recruited to take part in individual semistructured interviews after the intervention had been delivered.

Comparison group

Schools allocated to the comparison group continued with standard practice and received no additional input. Our experience from previous school trials was that retaining those schools allocated to the comparison group can be challenging.⁷⁴ We provided a payment of £500 at the end of the study as participation compensation to comparison schools (to cover administrative costs and/or provide cover for staff involvement in organising data collection). In addition, we offered feedback of survey data to comparison schools after completion of the study, if requested, because we know that schools highly value data that can be used to monitor and change policy and practice.⁷⁵

Health Technology Assessment progression criteria

Data collection during the pilot trial focused on assessing acceptability and feasibility and allowing us to judge progress against the agreed criteria for progression to a Phase III trial. The first objective was specifically to assess whether the criteria deemed necessary in order to progress to a Phase III trial were met (Box 2). These were agreed by the investigator team, the HTA co-ordinating centre and the Trial Steering Committee (TSC) prior to commencing the pilot trial, being considered to represent evidence of feasibility and acceptability of the intervention and trial methodology.

Data sources

In order to answer RQ1 (see Box 2), multiple sources of quantitative and qualitative data were collected. To assess needs assessment response rates, student baseline survey response rates at each intervention

BOX 2 Research questions addressing the agreed criteria for progression to a full trial

RQ1: was it feasible to implement the intervention in (at least) three out of four intervention schools? This was assessed according to the following implementation criteria:

1. The needs assessment survey had a $\geq 80\%$ response rate.
2. The action group met six or more times during the course of the school year and was always quorate (i.e. minimum of two-thirds of members present).
3. The action group reviewed and revised school policies (e.g. relating to discipline, behaviour management and staff–student communication).
4. Whole-school actions (e.g. rewriting school rules) were a collaborative process involving staff and students from across the school.
5. Peer mediation and/or ‘buddying’ schemes were reviewed and enhanced if necessary.
6. ≥ 20 staff completed restorative practice training.
7. Restorative practices (e.g. circle time, restorative conferencing, etc.) were used.
8. The student curriculum was delivered to year 8 students.

RQ2: was the intervention acceptable to the majority of school SMT members and a majority of action group members?

RQ3: did randomisation occur and was this acceptable to schools’ SMTs?

RQ4: did (at least) three out of four schools from each of the intervention and comparison arms accept randomisation and continue to participate in the study?

RQ5: were the student survey response rates acceptable at (at least) three out of four comparison schools?

school were examined. To examine the fidelity of action group implementation, documentary evidence was collected (intervention facilitators' checklists, action group meeting minutes and school policies), and supplemented by observations of action group meetings and interviews with action group members at each school to cross-check the validity of the documentary evidence and provide additional data. To examine the reach of staff training and the uptake of restorative practices, documentary evidence was collected from training providers' checklists and supplemented by observations of staff training sessions and focus groups with school staff to cross-check the validity of the documentary evidence and provide additional data. To examine the delivery of the student curriculum, documentary evidence from intervention facilitators' checklists was collected and supplemented with focus groups with school staff and students to cross-check the validity of the documentary evidence and provide additional data.

In order to answer RQ2 (see *Box 2*), semistructured interviews with school SMT members at each participating school were undertaken to explore their views on the intervention's acceptability and action group members at each intervention school were surveyed to examine their views on acceptability. A subsample of action group members was also interviewed to explore participants' views in more depth and cross-check the validity of other data sources. To answer RQ3 (see *Box 2*), the acceptability of randomisation was explored in the semistructured interviews with school SMT members at each participating school. In order to answer RQ4 and RQ5 (see *Box 2*), the retention of schools was assessed, as were response rates to baseline and follow-up student surveys in intervention and comparison schools.

Data analysis methods

Interviews and focus groups were transcribed verbatim and qualitative data entered into NVivo 10 software version 10 (QSR International Pty Ltd, Burlington, MA, USA) to aid data management and analysis. Documentary evidence and records of observations were also uploaded to support cross-checking and data triangulation. Codes were applied to transcripts to identify key themes and how these inter-relate in order to develop an analytical framework. Each transcript was coded to indicate the type of participant, the school and the date, allowing analytical themes to be explored in relation to different groups' experiences and processes to be compared across schools. Quantitative data from the action group member survey were entered into and analysed using Stata 12 (StataCorp, College Station, TX, USA).

In order to assess whether the intervention was delivered as planned in at least three out of the four pilot schools (RQ1), the following analyses took place to examine each of the specific implementation criteria systematically:

- Student needs assessment data reports were examined at each intervention school to assess whether the response rate was greater or lower than 80% of all year 8 students at the school at baseline.
- Intervention facilitators' checklists and action group meeting minutes were analysed to assess whether each intervention school action group met six or more times (with a minimum of two-thirds of members present), and were cross-checked against and supplemented by data from observations of action group meetings and thematic content analysis of action group member interviews.
- Intervention facilitators' checklists and school policy documents were analysed to assess whether each action group reviewed and revised school policies (e.g. relating to discipline, behaviour management and staff–student communication) and cross-checked with supplementary data from observations of action group meetings and thematic content analysis of action group member interviews.
- Action group meeting minutes and thematic content analysis of qualitative data from interviews with action group members were examined to assess whether whole-school actions (e.g. rewriting school rules) were a collaborative process involving staff and students from across the school.
- Action group meeting minutes and thematic content analysis of qualitative data from interviews with action group members were examined to assess whether peer mediation and/or 'buddying' schemes were reviewed and, if necessary, enhanced.
- Documentary evidence from training providers' checklists was examined to assess the attendance of staff restorative practices training, and cross-checked with supplementary data from observations of staff training sessions and thematic content analysis of data collected via focus groups with school staff.

- Focus groups of school staff examined the uptake of restorative practices (e.g. circle time, restorative conferencing, etc.) in intervention schools.
- Intervention facilitators' checklists were examined to check the student curriculum was delivered to year 8 students in all intervention schools, and cross-checked via thematic content analysis of data collected via focus groups with school staff and students.

Semistructured interviews with school SMT and action group members were subjected to thematic content analysis and coded according to views on acceptability, and supplemented with analyses of the action group member survey data in order to examine how many action group members reported acceptability (RQ2). We monitored whether randomisation occurred, and thematic content analysis of qualitative data from semistructured interviews with school SMT members was also used to assess the acceptability of randomisation (RQ3). We also monitored how many schools from each of the intervention and comparison arms accepted randomisation and continued to participate in the study at follow-up (RQ4) and student response rates at follow-up at each comparison school (RQ5).

Process evaluation: participants' experiences

In addition to examining intervention delivery according to prespecified criteria (objective 1), a second objective was to explore students', school staff members' and facilitators' experiences of implementing and trialling the INCLUSIVE intervention, and how these varied in different school contexts in order to refine the intervention and trial methods. To answer RQ6–8, multiple sources of qualitative data collected via interviews, focus groups and observations and quantitative data from student and staff surveys were analysed. The qualitative process data allowed us to explore students', school staff members' and facilitators' experiences in depth and see how these varied in different school contexts. In addition, surveys of students and teachers provided an overview of their experiences of intervention activities.

Qualitative process data

Semistructured interviews with a range of stakeholders were undertaken to explore in greater depth the processes of planning, delivering and receiving the intervention and undertaking the trial: semistructured interviews with school SMT members at each intervention school were undertaken to explore their experiences of planning, delivering and trialling the intervention, and how successfully each component was implemented; semistructured interviews with school SMT members at comparison schools provided insights regarding the feasibility and acceptability of our trial methods; and semistructured interviews with intervention facilitators and the training provider explored their experiences of planning and delivering the intervention, and how and why these may have varied across different school contexts. Further individual interviews with student and staff action group members at each intervention school explored in-depth how successfully action groups were implemented and their experiences of barriers and facilitators to implementation.

Focus group discussions, each with a maximum of 10 year 8 students, were undertaken at each of the four intervention schools, to explore the acceptability of the intervention to students in year 8, how benefits (or harms) may occur to students' health via such an intervention and their experiences of the evaluation methods. A focus group comprising a maximum of six school staff members at each intervention school also explored how successfully each component was implemented, why fidelity may vary across schools and the acceptability of the intervention, as well as staff views on the trial methods. Unstructured observations of action group meetings, staff training and the school environments were also recorded to provide further contextual data on potential barriers and facilitators to implementation. Further details on the qualitative data collection process and the sample of participants recruited are provided in *Chapter 3, Qualitative data collection*.

Quantitative process data

Follow-up survey data also provided an overview of year 8 students' and teachers' experiences of intervention activities. Neither survey measured these factors at baseline, so it was not possible to adjust effect estimates for baseline differences in school practice. Items on students' experiences of restorative practices ('Do teachers at your school ever use circle time?') and participation in decision-making ('At this school do students have a say in writing the rules?') were included in the student follow-up survey. Eight items on students' experiences of PSHE were also included at follow-up, which were adapted from the MRC Social and Public Health Sciences Unit's Teenage Health in Schools item measures using a four-point response scale (Totally agree/I agree a bit with responses/I do not really agree/No, totally disagree). These eight items were as follows: PSHE helps me feel more confident; PSHE helps me understand other people's feelings and problems; in PSHE I talk about how I feel; in PSHE students can be honest about how they feel; in PSHE we talk about how our words and actions affect other people; in PSHE we talk about strategies for working with others; and in PSHE we talk about how to be a good friend.

The teacher follow-up survey included two items on teachers' use of restorative practices ('Do teachers at your school ever use circle time to discuss how students feel about school?' and 'Do staff at this school ever use restorative conferences to deal with disputes and repair relationships?'), two measures of behaviour management practices ('How well are teachers supported with behaviour management at this school by senior members of staff?' and 'How well are teachers supported with behaviour management at this school by all staff implementing consistent techniques across the school?') and one question on PSHE ('Do you think that PSHE lessons at this school help to promote students' social and emotional well-being?').

Data analysis

Qualitative data collected via interviews and focus groups were transcribed verbatim and entered into NVivo software to aid data management and analysis. Thematic content analysis of qualitative data was undertaken⁷⁶ by two researchers (AF, the trial manager, and CB, the process evaluation manager for this study) via the following stages of analysis. First, each transcript was coded by AF to indicate the type of participant (e.g. female, student, etc.), the school and the date and then checked by the second researcher (CB). Second, one researcher (AF) coded all the interview and focus group transcripts to identify key themes, using an inductive framework based on our RQs as higher-level codes and memos to record inter-relationships across questions (e.g. to explore how implementation varied across different school contexts). Third, the second researcher (CB) cross-checked these codes. Finally, the two researchers compared their analyses, refined the coding framework and generated additional codes and memos in discussion, before coding the data a second time deductively to explore other analytical themes grounded in the data (e.g. unexpected barriers or facilitators to implementation).

Analyses of survey data examined how many students and teachers reported evidence of successful implementation at follow-up (see above for a full list of items). All quantitative data were analysed in Stata 12 and adjusted for clustering by school and, when possible, appropriate confounders: the analyses of students' reports adjusted for sex, ethnicity and housing tenure at baseline, and the analyses of teachers' responses adjusted for sex, ethnicity and teaching role at baseline. Adjustment for baseline differences in school practices was not possible, as these were not measured.

Piloting of outcome measures

The final objective of this study was to pilot and field test our indicative primary, secondary and intermediate outcome measures and economic methods prior to a Phase III trial (RQ9–11). The methods for piloting our economic measures and analyses are described in the section *Economic evaluation*. In this section, we provide details of the primary, secondary and intermediate outcome measures piloted in this study and describe the methods of statistical analysis used for assessing the performance of the

indicative primary outcome measures (bullying and aggressive behaviour) and changes in our indicative primary outcomes at follow-up. Outcomes were measured in baseline and follow-up questionnaire surveys of year 8 students conducted in September 2011 and June/July 2012, respectively (described in more detail in *Chapter 3*).

Primary outcome measure development

In order to develop primary outcome measures of bullying and aggressive behaviour for a Phase III trial, we examined the properties (response rates, discrimination and reliability) of two variables pre-hypothesised as primary indicative outcome measures: (1) the bullying victimisation scale used in the Gatehouse Project trial⁷⁸ and (2) the AAYP violence subscale for measuring aggressive behaviours used in the AAYP trial.⁵⁴ We also examined one additional pilot primary outcome: (3) the Edinburgh Study of Youth Transitions and Crime (ESYTC) school misbehaviour subscale.⁷⁷ Outcomes (1) and (2), respectively, assess student self-reports regarding bullying victimisation and perpetration of aggression within the last 3 months and are, therefore, potentially able to assess changes within a school year. Although they have been used previously in intervention studies, these were in Australia and the USA, respectively, and thus may not be appropriate to the UK. Outcome (3) might be more relevant to the UK context but has not previously been used in intervention research.

The Gatehouse Bullying Scale (GBS)⁷⁸ consists of 12 items that assess overt and covert types of bullying victimisation. Students were asked whether they had been teased or called names, had rumours spread about them, been deliberately left out of things and/or recently been physically threatened or hurt. Each of the four types of bullying was defined as frequent if it occurred on most days, and upsetting if the student reported being quite upset. The resulting bullying score was on a scale from 0 to 3, where 0 indicated that the student had not been subjected to any of the four types of bullying, and scores of 1–3 denoted increasing intensity (frequency and level of upset) of one or more of the four types of bullying.

Four items from the AAYP violence scale⁵⁴ were piloted. Questions concerning students' self-reported perpetration of violent behaviours over the previous 3 months requested information regarding the frequency of the most common overt direct aggressive behaviours, including verbal aggression (threatening to beat up, cut or stab) and physical aggression (cutting, stabbing). The four items were scored on a scale of 0–3 (0 = never; 1 = yes but not in the last 3 months; 2 = once recently; 3 = more than once recently). The resulting violence score ranged from 0 to 12; higher average scores represented higher levels of violence.

The ESYTC school misbehaviour subscale⁷⁹ measured several domains of violence and aggression at school. The 'school misbehaviour' scale piloted at baseline included 10 items relating to aggression towards students and teachers, which were coded according to four response categories: hardly ever or never; less than once a week; at least once a week; and most days. A further three items were added and piloted at follow-up regarding physical aggression in more detail using these response categories. The total score was a summed frequency of school misbehaviour with high scores indicating higher levels of self-reported school misbehaviour.

Consultation with young people on measures of bullying and aggression

We consulted with young people from the NCB YRG at a meeting held before the surveys were undertaken to explore young people's views concerning how to define bullying and aggression within schools and assess whether the aggression items designed for INCLUSIVE were appropriate and acceptable (see *Chapter 1, Involvement of young people*). It was important to assess young people's views on what constitutes aggressive behaviour both to inform the aims of the intervention and to inform our measures. The group identified a wide range of behaviours that could be perceived as bullying and aggression, including physical violence, verbal abuse, 'jokes', cyber bullying, pictures, drawings and websites, rumours, exclusion, peer pressure and theft. They concluded that bullying should be defined based on how often it occurs, arguing that repetitive and relentless behaviour is more likely to be experienced as aggressive.

We then asked the young researchers for their opinions on the scales we were testing as potential primary outcomes for a Phase III RCT. The group identified some concerns with both questionnaires. There were concerns about false reporting, which might be improved by ensuring confidentiality. Some items on the AAYP violence scale were thought of as possibly too 'extreme', leading to concerns that no British students would report these (e.g. stabbing someone). In contrast, other questions might assess experiences reported by almost everyone (e.g. using bad language). Some items were considered not to be well defined (e.g. verbal bullying), which may lead to variability in student responses. However, in general, the young researchers felt that the measures were likely to reflect students' experiences of aggression and bullying and were appropriate to pilot.

Pilot secondary outcome measures

We also piloted and field tested pre-hypothesised secondary outcomes using the baseline and follow-up surveys of year 8 students. These pilot secondary outcome measures are described in detail below.

Quality of life

This was measured using two instruments. First, the 30-item Paediatric Quality of Life Inventory (PedsQL) version 4.0⁸⁰ was used to assess all-round quality of life (QoL). The PedsQL has been shown to be a reliable and valid measure of QoL in normative adolescent populations.⁸⁰ It consists of 30 items representing five functional domains: physical, emotional, social, school and well-being. Items are rated on a series of five-point Likert scales ranging from 0, 'never', to 4, 'almost always'. The PedsQL yields a total QoL score, two summary scores for 'physical health' and 'psychosocial health' and three subscale scores for 'emotional', 'social' and 'school' functioning. For the total QoL score, items are reverse-scored and linearly transformed to a scale of 0–100 (i.e. 0 = 100, 1 = 75, 2 = 50, 3 = 25 and 4 = 0); higher scores represented better QoL. Second, the CHU-9D measure⁸¹ was used to assess health-related QoL and piloted and examined prior to use in a subsequent economic analysis; the validation of the CHU-9D is described below in detail (see *Validation of the Child Health Utility 9D questionnaire*).

Psychological function

Psychological distress was measured using the child self-completion Strengths and Difficulties Questionnaire (SDQ), a validated measure of a range of behavioural and emotional problems in children and adolescents.⁸² The SDQ consists of 25 items which assess difficulties in four domains (emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems) and an additional measure of pro-social behaviour. Items are rated on a scale of 0 (not true) to 2 (certainly true). A 'total difficulties' score is calculated by the summation of all scales, apart from pro-social behaviour, a high score being indicative of a higher level of emotional and behavioural difficulty. In this study, the total SDQ score and the emotional, conduct, hyperactivity and peer problem scores were treated as continuous variables.

Psychological well-being

Psychological well-being was measured using the Short Warwick–Edinburgh Mental Well-Being Scale (SWEMWBS), which consists of seven items designed to capture a broad concept of positive mental well-being, including psychological functioning, cognitive-evaluative dimensions and affective-emotional aspects.⁸³ Items were rated on a five-point scale: none of the time (score = 1), rarely (2), some of the time (3), often (4), all of the time (5). The responses are scored and aggregated to form a 'well-being index' (total score), which can range from a minimum of 7 (those who answered 'rarely' on every statement) to a maximum of 35 (those who answered 'all of the time' to all statements). Higher scores represent improved mental well-being. The SWEMWBS was piloted in this study because it potentially provides a brief (seven-item) measure of well-being, although it has not yet been validated with year 8 students in English schools.

Health risk behaviours

Measures of two key adolescent health risk behaviours were piloted as secondary outcome variables. Self-reported use of tobacco and consumption of alcohol were examined using the HSE pilot trial measures.⁶⁹

Students were asked if they had smoked in the past month (30 days) and if they had drunk alcohol, defined as 'more than a sip', in the past month (30 days).

Note that drug use and contraceptive use would also be secondary 'health risk behaviour' outcomes in a Phase III trial in which students would be followed up until the end of year 10 (aged 14 or 15 years), but they were deemed not appropriate for the pilot trial surveys of year 8 students (aged 12 or 13 years).

NHS use

Self-reported use of NHS services (primary care, accident and emergency, other) in the past year was scored as a dichotomous variable (i.e. those who reported that they had used NHS services in the last year were given a value of 1).

Contact with the police

Contact with the police was assessed using the Young People's Development Programme evaluation measure.⁸⁴ Young people were asked if they had been stopped, told off or picked up by the police in the last 12 months. This was a dichotomous response variable (ever/never had contact).

Truancy and school exclusion

Self-reported truancy and school exclusion were measured using items from the ESYTC survey.⁷⁸ Self-reported truancy was a dichotomous variable ('ever bunked/skipped school? yes/no'), as was self-reported school exclusion, which was assessed according to whether they had ever been permanently or temporarily excluded from school.

Pilot intermediate outcome measures

To examine the logic model (see *Figure 1*) via which the intervention is hypothesised to impact on students' health and well-being, we piloted the following intermediate outcomes examining year 8 students' perception of the school environment and antischool actions.

Student perceptions of school climate

The Beyond Blue School Climate Questionnaire (BBSCQ) scale was used to measure students' perceptions of the school climate. The scale was originally developed in Australia,⁸⁵ using items selected from the Quality of School Life,⁸⁶ Patterns of Adaptive Learning⁸⁷ and Psychological Sense of School Membership questionnaires.⁸⁸ It consists of 28 items, which produce an overall score and also assess four key domains of school climate (subscale): supportive teacher relationships, sense of belonging, participative school environment and student commitment to academic values. Each item was coded 1–4 on a four-point scale. Responses ranged from 'Yes, totally agree' to 'No, totally disagree'. A total BBSQ score was calculated by the summation of all scales, reversing scores when necessary. The total BBSQ score and subscale scores were treated as continuous variables; higher scores represent a positive report of school climate.

Antischool actions

Students' reports of antischool actions were derived from the ESYTC self-reported delinquency (SRD) subscale.⁷⁹ The total SRD score was a summed frequency based on six items, with high scores indicating higher levels of self-reported school delinquency.

Analysis plan for pilot outcome measures

All quantitative data analyses were carried out in Stata 12. In order to examine which of the pilot primary outcome measures performed best, according to their completion rate, discrimination and reliability (RQ9), and inform our choice of outcomes for a Phase III trial, multiple analyses were undertaken. First, the completion rate and prevalence for each outcome measure (total score and for each individual item) was examined overall and by sex. Second, mean scores, standard deviations and response distributions were calculated for all the pilot primary outcome measures at baseline to examine their discrimination and the potential for 'floor effects' (i.e. the scales may not be sensitive to low levels of reported bullying) with

year 8 students (aged 12–13 years). Third, the intraclass correlation (ICC) for each score was calculated and used as an indication of the stability of the measure over time. Fourth, Cronbach's alpha statistics were calculated at baseline and follow-up to provide an indication of each scale's internal consistency, a measure of the extent to which items within the scale are measuring the same latent construct.

The pilot study did not aim to, and was not powered to, determine intervention effects. Nonetheless, we did carry out analyses of outcomes, purely in order to pilot our primary outcome measures and analysis methods. To estimate intervention effects on the indicative pilot primary outcome measures, multilevel modelling was carried out according to the principle of intention to treat. Appropriate multivariate regression models were used with a random effect at the school level to allow for clustering, fitting baseline measures of outcomes when possible. Cluster-adjusted mean differences were computed for each of the pilot primary outcomes, with odds ratios used as a measure of effect for analyses of dichotomous data. Results of unadjusted and adjusted analyses are presented with a 95% confidence interval. 'Unadjusted' analyses adjusted only for baseline measures of the outcome in question. Adjusted analyses additionally adjusted for other pre-hypothesised potential confounders (sex, ethnicity and housing tenure) as covariates. This strategy of adjustment was highly conservative, the study's lack of statistical power rendering any more comprehensive approach to adjustment liable to generate highly unstable estimates of effect. All analyses were complete case analyses, as we had no reason to believe data were not missing at random.

We should stress that these quantitative analyses offer no guide to the intervention's effectiveness. The lack of statistical power means that no value can be placed on the cross-sectional point estimates or on estimates of differences between arms in longitudinal analyses. Only 95% confidence intervals offer any indication of potential effect sizes. Furthermore, as expected, the random allocation of only eight units did not generate baseline equivalence and there were marked differences by arm in several measures of students' sociodemographic characteristics, such as parental employment and family structure (see *Chapter 4, Description of pilot trial sample*) and prior experiences of bullying and aggression estimates (see *Chapter 4, Development and piloting of indicative primary outcome measures*), both of which suggested that the intervention group was notably more disadvantaged at baseline. Thus, the estimates of effects are subject not only to random error but also to confounding, which our conservative approach to adjustment could not control. Estimates of effect should therefore not be interpreted as reflecting the potential effectiveness of the intervention.

Economic evaluation

It is important that studies evaluating complex public health interventions include an economic component.^{89,90} The aim of the economic component of this pilot study was not to perform an economic evaluation of the intervention per se, but rather to collect and collate evidence regarding the appropriate design of an economic evaluation in a Phase III trial. To do this, the analysis was divided into two tasks. The first task was to assess the feasibility and desirability of using the recently developed CHU-9D to measure changes in health. The second task was to define an appropriate economic evaluation framework based on the CHU-9D results, broader pilot trial data and consideration of the wider literature.

Validation of the Child Health Utility 9D questionnaire

We chose to use the CHU-9D rather than any other outcome measure for three reasons. First, although it was known at the time of the study's conception that the European Quality of Life-5 Dimensions (EQ-5D) health questionnaire was being developed for use with children and young people (the EQ-5D-Y), it was not clear how this process would evolve. Second, discussions regarding the face validity of the EQ-5D-Y and the CHU-9D among the investigation team strongly supported the use of the latter because of the items it contains: they were considered to be much more relevant given the target age group. Third, it was noted that the CHU-9D classification system has been specifically developed using children's input, whereas the EQ-5D-Y simply involved changes to the wording of the original system in order to make it more

'child friendly'. We discussed the appropriateness of including both questionnaires in the student survey in order to compare their results; however, it was agreed that the questionnaire was already approaching a maximum length and that concerns over the face validity of the EQ-5D-Y outweighed any potential benefits of including it in the pilot.

The feasibility and desirability of using the CHU-9D in a subsequent Phase III trial were assessed by determining its acceptability, construct validity and sensitivity. Acceptability was examined by analysing response rates, and basic psychometric properties were initially assessed by analysing the distribution of the calculated utility scores and exploring the potential for floor and ceiling effects. Construct validity was assessed using two approaches. First, the discriminant validity of the CHU-9D (i.e. its ability to differentiate between individuals in different states of 'health') was assessed by correlating baseline survey CHU-9D utility scores with the baseline PedsQL scores. Second, the convergent validity of the CHU-9D was examined by assessing the relationship between the CHU-9D subscales and subscales on other questionnaires that were believed to be measuring similar items. For example, it was hypothesised that the 'sad' CHU-9D scale would be positively related to the PedsQL psychosocial summary score, and that the CHU-9D 'worried' and 'sad' subscales would correlate with the SDQ and the SWEMWBS. As the data were either non-normally distributed or ordinal, Spearman's rank correlation coefficients were calculated in all instances (all statistical tests were two-tailed). Finally, sensitivity was assessed by calculating the change in utility score between the baseline and follow-up surveys, and correlating this change in score with the corresponding change in the PedsQL total score and the three pilot primary outcome measures, again using Spearman's rank test. Correlations and tests for trend were also performed to assess the relationship between the CHU-9D utility score, the PedsQL total score and these outcomes.

Developing a framework for economic evaluation

At their most basic level, economic evaluations of cost-effectiveness compare the costs and consequences of a relevant set of alternative health-care programmes or interventions. It implicitly follows that, in order to maximise outcomes given a limited budget, resources should be allocated towards interventions that are considered to be cost-effective and away from those that are not. However, an economic evaluation of INCLUSIVE is likely to be complex for several reasons. First, no multiattribute health 'utility' status classification systems have been validated for use among young people in schools or for 'antibullying' interventions. Second, costs and benefits are likely to extend beyond traditional health boundaries and into a number of other sectors and policy domains (e.g. education and the criminal justice system). Third, it is possible that the intervention will not only impact on students but also potentially provide benefits for school staff. Fourth, a RCT will not capture the potential longer-term health incremental benefits and costs arising later in the life-course as a result of the intervention; the value of undertaking complementary decision modelling, which this would require, is unclear.

There are various different forms of economic evaluation. They are similar in the way that they measure and value costs but different in terms of how they measure and value health outcomes. For example, in a cost-benefit analysis, outcomes are expressed in monetary terms, whereas, in a cost-utility analysis (CUA), they are typically expressed as quality-adjusted life-years (QALYs), where QALYs are calculated as the time spent in a state of health adjusted for the 'quality' or 'utility' of this time. These techniques contrast with a cost-effectiveness analysis approach, for which outcomes are expressed in 'natural' units, such as 'life-years' or perhaps 'episodes of bullying'. The most appropriate form of economic evaluation to use typically depends on the level of efficiency that the evaluation is trying to address (e.g. within a health-care sector such as the NHS or at a broader public health level) and the type of intervention outcomes that are anticipated. For example, QALYs reflect changes only in 'health'; they do not capture potential wider non-health benefits that might be valued by society, such as improved educational attainment.

A number of guidelines exist for performing economic evaluations on which to base a framework for INCLUSIVE. However, the practical approach we took, given that a Phase III RCT would be UK based and consist of a relatively 'complex public health intervention', was to follow the methodology outlined by the National Institute for Health and Care Excellence's (NICE's) public health guidance.⁹¹ More specifically,

key design elements were abstracted from its methods guide, such as 'the categories of cost to include' (i.e. the cost perspective) and 'the appropriate time horizon for the analysis', in order to debate the appropriate design response. The overall analysis framework in which this choice rests was of equal importance; for example, the choice of time horizon, whether decision modelling is needed and which costs to include.

To define an appropriate economic evaluation framework for a subsequent Phase III trial, pilot trial data and the wider literature were assessed, including the plausibility and desirability of different frameworks given the intervention objectives and school contexts. Note that several design issues were taken as given: for example, that the control arm in the Phase III trial should consist of 'routine practice' and that it will be designed in a manner that will minimise the chance of 'contamination' from the intervention arm. To guide us, we identified a recent systematic review that summarised and critiqued the literature with respect to 'universal' interventions that aimed to promote emotional and social well-being in secondary schools.⁹²

Trial registration, governance and ethics

This study was funded by the HTA research programme (HTA project: 09/05/05) at the NIHR, which convened annual monitoring meetings. The trial protocol was registered with Current Controlled Trials (ISRCTN88527078, www.controlled-trials.com/ISRCTN88527078/). The study was overseen by a TSC, which included an independent chair, two other independent members, an investigator representative of each institution involved in the research and a representative from HTA. The TSC met 6-monthly throughout the study (three times in total) to examine the methods proposed and monitor data for quality and completeness. With the agreement of the TSC and NIHR HTA co-ordinating centre, a separate Data Monitoring and Ethics Committee was *not* established because this was a pilot trial with no interim analysis. Membership of the TSC is shown in *Appendix 1*.

Ethics committee approval was given by the London School of Hygiene and Tropical Medicine (LSHTM) Research Ethics Committee (application reference number: 5954) and details of the research, including possible benefits and risks, were provided to schools through written information and personal meetings. The head teacher and the chair of governors at all participating schools signed a participation agreement at the start of the study. Written consent was sought from young people at all stages of data collection using age-appropriate information sheets together with oral explanation by researchers. Parents who did not wish their child to participate were able to opt out. Written consent was also sought from teachers, other school staff and the facilitators prior to their participation in any data collection, using age-appropriate information sheets for young people. All participants had the opportunity to ask questions and to opt out at any point. All information remained confidential within the research team and neither schools nor individuals were identified at any point. In this report, the names of all schools and individuals involved in the intervention have been replaced with pseudonyms. The principal investigator (RV) and the trial manager (AF) were also trained in Good Clinical Practice prior to the study commencing.

Chapter 3 Undertaking the trial

The primary foci of this pilot trial were feasibility and acceptability.⁶⁶ Collection and analysis of process data were therefore central to this study in order to assess whether it would be appropriate to continue to a Phase III trial and to explore participants' experiences. In order to assess the feasibility and acceptability of implementing and trialling the INCLUSIVE intervention in English secondary schools, a diverse range of schools was purposively recruited and the cluster RCT design was piloted at these schools. Both quantitative and qualitative data were collected via surveys of year 8 students (aged 12 or 13 years), school teaching staff and action group members; semistructured interviews; focus groups with staff and students at the intervention schools; and other methods such as the collection of documentary evidence and recording observations. This section describes in detail the process of recruiting schools, randomisation and these methods of data collection.

Recruitment of schools

The pilot study was planned to begin in April 2011 but began in July 2011 because of delays in the contracting process. This made recruitment extremely challenging. The final school was not recruited until mid-September 2011. Eight mixed-sex state secondary schools in London ($n = 6$) and south-east England ($n = 2$) were recruited by the lead education consultant, who co-ordinated the intervention, using a convenience sampling frame of 60 schools known to her through existing professional networks. In addition to the short lead-in time, recruitment was also more challenging than would be the case in a Phase III trial because of the purposive criteria we set ourselves (see *Chapter 2, Sampling and recruitment*).

Schools were contacted by letter initially, and then by phone. Owing to delays in contracting and the study start date, school baseline surveys were not completed until the end of September 2011, and randomisation and intervention initiation could not begin until the end of October 2011. As discussed below, this presented challenges for implementation in terms of scheduling meetings and organising staff training in schools because the calendar of training and meetings for the academic year had by then already been determined. Given this, it is a strong reflection of the intervention's feasibility and acceptability that all elements of the intervention were, nonetheless, delivered within 1 year and all progression criteria met (detailed in *Chapter 4, Health Technology Assessment progression criteria assessment*).

Baseline student and staff surveys

Once they had agreed to participate in the study, and each school's head teacher and chair of governors had signed their letter of agreement, schools provided the research team with a list of all students who were in year 8 and all the teaching staff (full or part time) employed at the school at the start of the 2011–12 academic year. Anonymous school and participant codes for all students and teachers were used to set up a password-protected database linking names with anonymous identity numbers. These identity numbers were printed on questionnaires prior to the surveys. All year 8 students' parents were sent a letter and information sheet at the start of the study, which allowed them to 'opt out' of their child participating.

School surveys were completed in September 2011. All baseline data were collected from year 8 students prior to randomisation. Surveys were completed on the school site with trained researchers and fieldworkers in attendance over one school timetable period (approximately 1 hour). Age-appropriate information sheets and consent forms were provided, along with a verbal explanation by researchers/fieldworkers and students. Paper-based questionnaires were provided to those students who agreed in writing to participate and these were completed confidentially in classroom conditions. One or two of the research team and/or fieldworkers were always present to supervise questionnaire completion, and a

teacher was also present during the student surveys to ensure silence while students completed their questionnaires, but remained at the front of the classroom and unable to see the questionnaires. The researchers/fieldworkers helped students with questions they did not understand. To obtain data from those absent on the day of the survey, a consent form and questionnaire with an anonymous identity number were left at the school in a stamped addressed envelope (SAE) for them to return directly to the trial manager by post.

Teaching staff at each school were also invited to participate in a baseline survey at the start of the academic year (September 2011). One school requested not to participate in the teacher survey after the start of the study, and therefore teacher surveys were undertaken only at seven out of the eight schools. The school that opted out did so because of relatively exceptional, external factors: they found out during the autumn term that they would be subject to a merger with another school and management changes, and therefore it was not appropriate or feasible to survey staff at the school during the 2011–12 academic year. At all other schools, teacher surveys took place on the school site during staff meetings or briefings. Teachers were provided with information about the study and written consent forms, with researchers providing further information and answering any questions when necessary. Paper-based questionnaires were given to those staff who provided written consent to participate and these were completed confidentially. As with the student survey, to obtain data from those not present on the day of the survey, a consent form and questionnaire with an anonymous identity number were left at the school in a SAE for them to return to the trial manager.

Randomisation process

Randomisation was not possible until October 2011, by which time all student baseline surveys had been completed. Randomisation was undertaken remotely in the offices of the Clinical Trials Unit at the LSHTM. Within each 'matched pair' (see *Figure 2*), schools were randomly allocated using simple random number tables (intervention if $n > 0.5$; control arm if $n \leq 0.5$) with no restriction. Schools were then informed of their allocation by the trial manager. As with most social intervention trials, schools, students, teachers and other staff could not be 'blinded' to their allocation status. However, data-input staff were 'blinded' to each school's status throughout the study and analysis of follow-up quantitative data was undertaken 'blind' to allocation.

Follow-up student and staff surveys

Schools provided the research team with an updated list of all year 8 students on the school roll at the start of the summer term of 2012 in order for the research team to update its database and allocate new identity numbers to students who had joined the school since the start of the academic year. All enrolled students were included in the follow-up survey, rather than only those who took part in the baseline survey, in order to assess student turnover within 1 school year prior a larger Phase III study, and to maximise the overall analytic sample to support our assessment of the reliability, validity and other properties associated with our pilot outcomes. The list of teachers was not updated because of the low turnover of school staff within an academic year, and therefore the very few teachers who had joined the school since the start of the academic year were not surveyed at follow-up. New year 8 students' parents were sent a letter and information sheet prior to follow-up, which allowed them to 'opt out' of their child participating.

Using the same procedures as at baseline (described above in *Baseline student and staff surveys*), follow-up surveys of year 8 students and teachers took place at each school at the end of the academic year (June–July 2012). The same team of trained researchers and fieldworkers supervised the consent process and questionnaire completion, ensuring that participants received and completed the questionnaire that had their anonymous identity number printed on it, to track them over time.

It was not always possible for fieldworkers to be blind to allocation status at follow-up, both because several had been involved in the process evaluation and because informal conversations with staff and students during fieldwork tended to reveal to them which arm the school had been allocated to.

Those students and staff absent on the day of the survey were again asked to return their questionnaires by post using a SAE because this method worked well at baseline. However, students' home and personal (e.g. mobile phone) contact details were not collected at baseline, which meant we could not follow-up those students who had left the school at follow-up. Student follow-up surveys took place at all eight schools participating in the study. The one school that did not participate in the teacher survey at baseline also did not participate in the teacher survey at follow-up.

Intervention provider checklists and other documentary evidence

Intervention facilitators ($n = 3$) and the training provider ($n = 1$) recorded intervention activities at each school using a structured, electronic checklist. These were used to assess the measures of feasibility and acceptability set out in the NIHR progression criteria and to facilitate recall during interviews if necessary. Each intervention facilitator kept a structured checklist to collect data systematically on the following at each school: meeting dates, action group membership, key information (e.g. dates, data sources, etc.) during the process of assessing needs and planning actions, school actions and curriculum implementation. Minutes of all action group meetings, and copies of original and revised school rules/policies, were also collected at each intervention school. To examine the reach of staff training, the staff training providers also kept a checklist to record the dates of all training events, and the location and numbers of staff present.

Qualitative data collection

Semistructured interviews with school SMT members took place during the summer term (April–July) of 2012 on the school site in a private meeting room, lasting 50–70 minutes, and using detailed topic guides. These topic guides included questions and prompts regarding our a priori progression criteria and how context might influence the process of implementation.^{69,93} SMT members were interviewed at seven out of the eight schools during the summer term by members of the research team (four intervention schools; three comparison schools), either individually or in pairs (e.g. the head teacher and deputy head teacher were interviewed together at some schools). Action group members were also interviewed individually or in pairs at all the intervention schools during the summer term of 2012 on the school site in a private meeting room. These interviews lasted 30–60 minutes, drawing on a priori topic guides to structure the interviews. Each intervention facilitator ($n = 3$), one of whom also co-ordinated the intervention, participated in semistructured interviews after the intervention in August and September 2012 to explore in depth their experiences of planning and delivering the intervention, as did the training provider ($n = 1$).

Focus groups with year 8 students and school staff in intervention schools were also undertaken during the summer term of 2012. All focus groups took place on the school site, in a private meeting room, and lasted 60–80 minutes. The two researchers present used a semistructured topic guide and participatory techniques, such as mapping and ranking exercises, to promote discussion among all participants. A total of 16 focus groups were held with year 8 students: four focus groups were carried out at each intervention school, for which students were purposively recruited and grouped according to their sex and level of school engagement, as reported by teachers. The size of these focus groups varied between 5 and 10 students (mean size = seven students; 112 students took part across all four schools). A total of four staff focus groups took place: one at each intervention school. Four to six members of staff were recruited to participate at each school, including teaching and non-teaching staff (e.g. teaching assistants, special educational needs co-ordinators, etc.).

All interviews and focus groups were audio recorded and transcribed verbatim. In addition, observations of the action group meetings and staff training and informal observations of the school environment were recorded in fieldwork diaries to provide further contextual data on potential barriers and facilitators to implementation, and to reflect on the logic model, inform intervention development and explore trial methods. The observations of action group meetings and training events allowed us to cross-check data recorded in intervention providers' checklists (described above in *Intervention provider checklists and other documentary evidence*). A total of 16 action group meetings were observed – this ranged from three to five at each intervention schools – and at least one action group meeting was observed at each school in the autumn and spring terms. Field notes (> 100 total) were also recorded based on informal observations at all participating schools.

Action group member survey

In addition to the semistructured interviews with action group members, all student and staff action group members at each intervention school were invited to participate in a brief self-completion survey at the end of the intervention in July 2012. The aim of this cross-sectional survey was to examine the level of participation and participants' experiences in order to assess the acceptability of the intervention to action group members as rigorously as possible (objective 1, RQ2). Anonymous questionnaires and consent forms were distributed at the end of the final action group meeting at each school. Action group members who were not present at the final meeting were asked to return their questionnaires by post using a SAE.

Public involvement by young people during the intervention

Our patient and public involvement group, the NCB YRG, held a second meeting at the end of the first term of the intervention. As this coincided with the formation and first meetings of the school action groups, we asked the young researchers to advise us on how to encourage students to participate in these action groups.

The YRG had several suggestions for encouraging students to join action groups, including incentives (e.g. gift certificates) and the importance of emphasising such participation in job and university applications. However, it was also felt that emphasising the intrinsic value of the project could serve as a strong motivation. Furthermore, the young people felt that providing feedback on progress, making the long-term schedule clear and ensuring student input is listened to and acted on would help encourage continued participation. The ratio of students to teachers was also suggested to be important. These suggestions were fed back to the intervention facilitators, who liaised with school staff running the action groups to support their development based on this feedback.

Chapter 4 Results

We first describe the study sample, in terms of both the schools and the students who participated, and then present key findings related to the primary objective, which was to assess the prespecified progression criteria considered to represent evidence of feasibility and acceptability of the intervention and trial methodology. We then present students', school staff members' and facilitators' experiences of the intervention and trial process to explore acceptability and feasibility in more depth and optimise the intervention and refine methods prior to a Phase III trial. We report quantitative data on the indicative primary outcomes in order to assess which would be used in a Phase III trial. Pilot secondary and intermediate outcome data are presented in full in *Appendices 1* and *2*. In the final part of this chapter, data examining the feasibility and desirability of using the CHU-9D scale in a subsequent economic evaluation are presented, and the suggested framework for an economic evaluation is discussed in *Chapter 5*.

Description of pilot trial sample

Flow of participants in the pilot trial

Figure 3 describes the flow of schools and students during the pilot trial. The total baseline student survey sample of year 8 students (aged 12 or 13 years) was 1144. At follow-up, 1114 year 8 students provided data, although there were 97 'new' students who did not complete the baseline survey at these schools

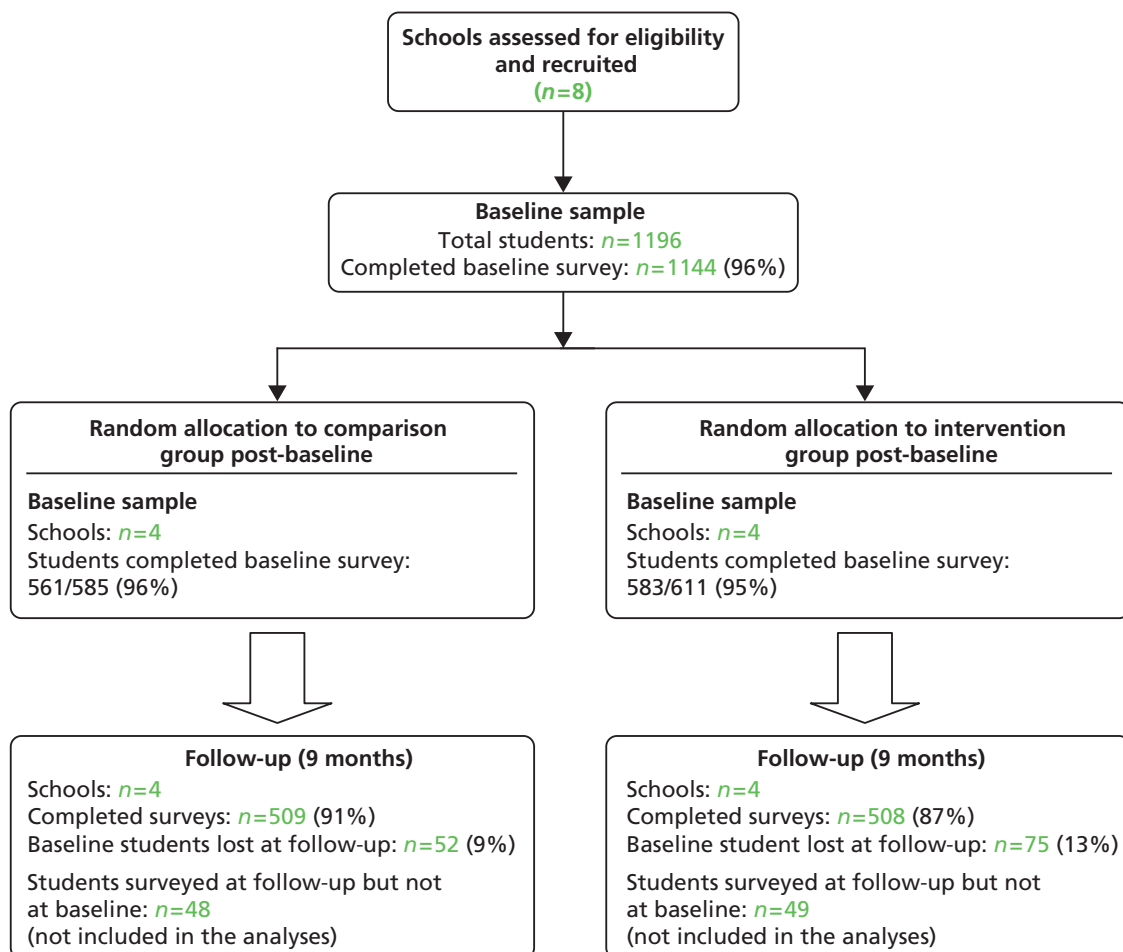


FIGURE 3 Flow chart of recruitment and participation of schools and students.

(because they had joined the school since the start of the year, were away from school during the baseline survey or refused to participate at baseline but not at follow-up). A total of 1017 students remained in the study from the baseline survey to follow-up, which represented 91% of students in the comparison group ($n = 509$) and 87% of students in the intervention group ($n = 508$).

Student survey response rates

The baseline student survey response rate ranged from 83% to 100%, with an average response rate of 96%. Follow-up response rates were similar with a 93% average response rate (Figure 4).

School characteristics

Eight mixed-sex state secondary schools in London and south-east England were recruited and matched according to the purposive sampling criteria used to ensure diversity according to the Ofsted rating of school effectiveness and overall rate of eligibility for FSM, the government’s standard measure of deprivation (Figure 5). Table 1 describes schools’ characteristics.

Student characteristics

Table 2 shows the sociodemographic characteristics of students at baseline by comparison and intervention group, and overall. Imbalances between the groups were evident at baseline across multiple measures of social and economic disadvantage, such as family structure, parental employment and housing tenure.

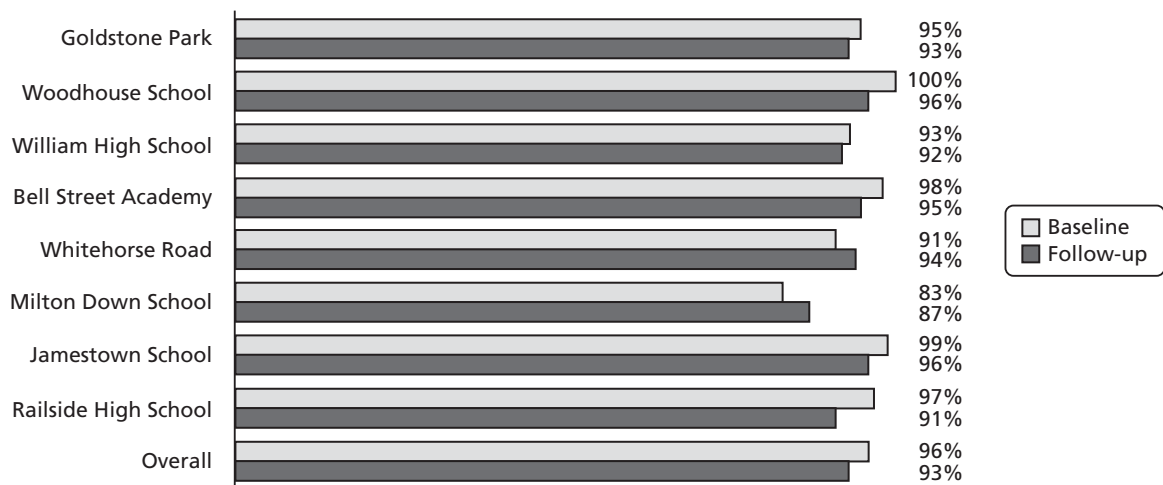


FIGURE 4 Student survey response rates by school at baseline and follow-up.

		Ofsted-rating of school effectiveness	
		'Satisfactory'	'Good'/'outstanding'
Eligibility for FSM in 2010–11	>6% and <23% of all students	(1a) Goldstone Park (intervention)	(2a) Williamson High School (intervention)
		(1b) Woodhouse School (comparison)	(2b) Bell St. Academy (comparison)
	≥ 23% of all students	(3a) Whitehorse Road (intervention)	(4a) Jamestown School (comparison)
		(3b) Milton Down School (comparison)	(4b) Railside High School (intervention)

FIGURE 5 Schools according to sampling and matching criteria ($n = 8$).

TABLE 1 School characteristics

2011 DfE data	Comparison schools				Intervention schools				National average
	(1b) Woodhouse School	(2b) Bell Street Academy	(3b) Rush Croft	(4a) Jamestown School	(1a) Goldstone Park	(2a) Williamson High School	(3a) Whitehorse Road	(4b) Railside High School	
Ofsted inspection rating	Satisfactory	Outstanding	Satisfactory	Good	Satisfactory	Good	Satisfactory	Outstanding	N/A
Students eligible for FSM	< 23%	< 23%	> 23%	> 23%	< 23%	< 23%	> 23%	> 23%	14.6%
Age range (years)	11–18	11–18	11–18	11–18	11–16	11–18	11–18	11–16	N/A
Sex	Mixed	Mixed	Mixed	Mixed	Mixed	Mixed	Mixed	Mixed	N/A
School type	Voluntary aided	Sponsored academy	Community	Voluntary aided	Community	Community	Voluntary aided	Foundation	N/A
Total number of students (approximately)	1100	500	800	800	850	900	750	1000	N/A
Students with SEN statement or on School Action Plus (approximately)	10%	8%	10%	8%	10%	17%	10%	12%	N/A
Students for whom English is an additional language (approximately)	35%	15%	50%	64%	16%	22%	63%	25%	12.3%
Students achieving at least five A*-C GCSEs (or equivalent) including English and mathematics	70–79%	N/A	40–49%	40–45%	60–69%	50–59%	50–59%	50–59%	58.9%
Students achieving at least five A*-C GCSEs (or equivalent)	> 90%	N/A	60–69%	70–79%	80–89%	80–89%	80–89%	80–89%	79.5%
Unauthorised absence	1–2%	1–2%	> 5%	1–2%	1–2%	2–4%	< 1%	2–4%	1.41%
Number of teachers (FTE approximately)	80+	< 40	60	60	55	60	60	60	N/A
Number of teaching assistants (FTE approximately)	10	20	20	20	< 10	30	20	20	N/A
Pupil-teacher ratio	14:1	14:1	14:1	15:1	17:1	15:1	14:1	17:1	N/A

DfE, Department for Education; FTE, full-time equivalent; GCSE, General Certificate of Secondary Education; N/A, not applicable; SEN, special educational needs.

TABLE 2 Sociodemographic characteristics of students at baseline

Category	Comparison [<i>n</i> (%)]	Intervention [<i>n</i> (%)]	Overall [<i>n</i> (%)]
Age (years), mean (SD)	12.11 (0.32)	12.12 (0.44)	12.11 (0.39)
Sex			
Male	299 (54.3)	309 (54.2)	608 (54.2)
Female	252 (45.7)	261 (45.8)	513 (45.8)
Ethnicity			
White British	216 (39.1)	282 (49.5)	498 (44.3)
Asian/Asian British	81 (14.6)	87 (15.3)	168 (15.0)
Black/black British	104 (18.8)	103 (18.1)	207 (18.4)
Chinese/Chinese British	5 (0.9)	4 (0.7)	9 (0.8)
Mixed ethnicity	55 (9.9)	46 (8.1)	101 (9.0)
Other	92 (16.6)	48 (8.4)	140 (12.5)
Religion			
None	58 (10.5)	222 (38.8)	280 (24.9)
Christian	409 (74.0)	195 (34.1)	604 (53.7)
Jewish	0 (0.0)	0 (0.0)	0 (0.0)
Muslim/Islam	42 (7.6)	93 (16.3)	135 (12.0)
Hindu	15 (2.7)	12 (2.1)	27 (2.4)
Sikh	1 (0.2)	1 (0.2)	2 (0.2)
Do not know	12 (2.2)	36 (6.3)	48 (4.3)
Other	16 (2.9)	13 (2.3)	29 (2.6)
Family structure			
Two parents	380 (68.0)	346 (60.5)	726 (64.2)
Single mother	133 (23.8)	149 (26.0)	282 (24.9)
Single father	7 (1.3)	19 (3.3)	26 (2.3)
Reconstituted	27 (4.8)	50 (8.7)	77 (6.8)
Other	12 (2.1)	8 (1.4)	20 (1.8)
Parental employment			
No	48 (8.7)	81 (14.5)	129 (11.6)
Yes	415 (75.6)	364 (65.1)	779 (70.3)
Do not know	86 (15.7)	114 (20.4)	200 (18.1)
Housing tenure			
Social rented	77 (14.2)	104 (18.4)	181 (16.4)
Private rented	89 (16.5)	75 (13.3)	164 (14.8)
Private owned	226 (41.8)	206 (36.5)	432 (39.1)
Other	20 (3.7)	6 (1.1)	26 (2.4)
Do not know	129 (23.8)	174 (30.8)	303 (27.4)
Family affluence			
FAS mean (SD)	5.83 (1.85)	5.47 (1.83)	5.65 (1.85)

FAS, Family Affluence Scale; SD, standard deviation.

The intervention group was consistently more disadvantaged. Students also differed in terms of their ethnicity and religion between the comparison and intervention groups. Such imbalances are not unexpected given the small number of units randomised, albeit with crude matching on two measures.

Overall, 54.2% of students were male and 45.8% were female, which varied little across the trial arms (< 1%). The average age was 12.1 years at baseline, which also did not vary across the groups. The majority of students were white British (44.3%), followed by black and black British (18.4%), Asian/Asian British (15%), other (12.5%), mixed ethnicity (9%) and Chinese (0.8%). The proportion of students from black and other minority ethnic groups was higher in the comparison group, and a much higher proportion of students in the comparison schools categorised themselves as belonging to another ethnic group not described in the survey (16.6% 'other' in the comparison group compared with 8.4% in the intervention group). The most common religion was Christianity (53.7%), although the proportion of students reporting being Christian varied markedly between the intervention (34%) and comparison groups (74%). More than twice as many students reported being Muslim in the intervention group (16.3% vs. 7.6%).

All four measures of socioeconomic disadvantage indicated that students in the intervention group were more socially and economically disadvantaged than those in the comparison group. Fewer students in the intervention group lived with both parents at baseline, and a higher proportion of students in the comparison schools reported that at least one parent in their household was in full-time or part-time employment (75.6% vs. 65.1%). In addition, more students from the intervention schools lived in socially rented accommodation (18.4%) than students in the comparison schools (14.2%), and the Family Affluence Scale, which assesses car ownership, children having their own bedroom, the number of computers at home and the number of holidays in the past 12 months, suggested a higher level of family affluence in the comparison group.

Health Technology Assessment progression criteria assessment

The first objective was to assess whether the criteria deemed necessary in order to progress to a Phase III trial were met (see *Chapter 2, Box 2*). These were agreed by the investigator team, the HTA co-ordinating centre and the TSC prior to the pilot trial commencing, and these were considered to represent evidence of feasibility and acceptability of the intervention and trial methodology.

Intervention feasibility and acceptability

The intervention implementation criteria (RQ1 and RQ2) were all met in full, as evidenced by *Table 3*. This was a major success given the problems arising from the late initiation of the project and the consequent delays in the recruitment of schools, baseline surveys and allocation. Substantially more than 80% of students completed the needs assessment survey at all intervention schools (range 91–97%). At least six action group meetings were held in each intervention school during the pilot year (range 6–7), and the size (range 13–25) and composition of action groups provided further evidence of their feasibility and acceptability. A range of SMT, other staff and student members attended action group meetings at all schools, as outlined in *Table 3*.

TABLE 3 Intervention feasibility and acceptability criteria (RQ1 and RQ2)

Activities/measures	Goldstone Park	Williamson High School	Whitehorse Road	Railside High School
RQ1 Was it feasible to implement the intervention in (at least) three out of four intervention schools? This was assessed using the following measures				
Needs assessment data				
Response rate	95%	93%	91%	97%
Action group				
Meetings held	7	6	6	6
Total number of members	22	20	13	25
SMT members	1	1	1	1
Subject teachers	7	3	2	8
Support staff	5	4	2	3
Administrative staff	1	1	1	2
Students	7	9	6	9
Other	1	2	1	2
Actions taken				
Review of policies	✓	✓	✓	✓
Rewriting rules	✓	✓	✓	✓
Revised peer mediation	✓	✓	✓	✓
Staff restorative practices training				
Date of half-day training session	3 January 2012	9 January 2012	24 April 2012	3 January 2012
Attendees at half-day session (n)	83 (75 staff members; 2 year 8 student action group members)	32 (30 staff members; 2 year 8 student action group members)	35 (all staff members)	68 (65 staff members; 3 year 8 student action group members)
Dates of enhanced sessions	26 March 2012, 9 May 2012, 2 July 2012	14 June 2012, 21 June 2012, 28 June 2012	30 April 2012, 17 May 2012, 15 June 2012	26 March 2012, 9 May 2012, 2 July 2012
Attendees at enhanced sessions (n)	9 (including deputy head)	12 (including the head)	8 (including deputy head)	10 (including the head)
Restorative practices				
Circle time implemented	✓	✓	✓	✓
Restorative conferences	✓	✓	✓	x
Curriculum				
Hours provided to year 8 students	7	10.5–12	10	10
Units delivered	2.1, 2.2, 3.1–3.3, 6.1, 6.2, 6.5, 6.6	1.1–1.6, 3.1–3.5, 5.1–5.4	2.1, 2.2, 4.1–4.4, 5.1–5.4	2.1, 2.2, 3.1, 6.1, 6.6
Other year groups targeted	Year 7	Years 7 and 9	None	None

TABLE 3 Intervention feasibility and acceptability criteria (RQ1 and RQ2) (*continued*)

Activities/measures	Goldstone Park	Williamson High School	Whitehorse Road	Railside High School
RQ2 Was the intervention acceptable to a majority of school SMT members and a majority of action group members?				
SMT interview(s) indicated intervention acceptable?	✓	✓	✓	✓
Action group interviews indicated intervention acceptable?	✓	✓	✓	✓

A survey of action group members was undertaken to examine the level of participation and participants' experiences in order to assess the acceptability of the intervention to action members as rigorously as possible. *Table 4* reports respondents' characteristics and *Table 5* reports their level of participation and experiences. The great majority of action group members found that the groups were representative of the students and staff from the school, that the needs assessment was useful, that the external facilitator helped the group function and that the group ensured decided actions were implemented (see *Table 4*). Over 93% of action group members who responded to the action group survey reported that the intervention was a good way to ensure students contribute to decision-making at their school. As noted in *Table 3*, all school action groups reviewed and revised school policies (e.g. relating to discipline, bullying, pastoral care, peer mediation, etc.) as planned, and there was evidence of whole-school collaborative actions involving staff and students from across the school. This generally occurred only in the summer term because of delays with the intervention, which in turn caused delays in recruitment and baseline surveys, which needed to be completed prior to allocation. In addition, at the four intervention schools, peer mediation and/or 'buddying' schemes were reviewed and revised if necessary.

In total, > 20 staff completed restorative practices training at each school as planned. The number of staff attending the half-day training ranged from 30 to 65 (see *Table 3*); the number of staff attending the enhanced 3-day restorative practices training ranged from 8 to 12. All schools had implemented restorative practices, such as circle time, by the end of the intervention, although this generally was not implemented until the third (summer) term (i.e. towards the end of the intervention). The enhanced training in restorative conferencing took place near the end of the year at all intervention schools (see *Table 3* dates). This slippage in training limited the extent to which restorative conferencing approaches could be implemented in the study period: three schools had begun to use this approach to address more serious incidents in the summer term, while one school had not, but planned to do so the following year (Railside High School).

TABLE 4 Action group member survey: respondent characteristics

Position	Male (n)	Female (n)	Missing sex (n)	Overall (n)
Student	6	8	0	14
Subject teacher	0	6	0	6
Head of year	0	1	1	2
Head of department	3	4	0	7
School senior management	1	2	0	3
Other	3	9	0	12
<i>Total</i>	<i>13</i>	<i>30</i>	<i>1</i>	<i>44</i>

TABLE 5 Action group member survey: participation and experiences

Item	Response	n (%)
How many INCLUSIVE project action group meetings have you attended?	None	0 (0.0)
	One or two	3 (6.8)
	Three or more	41 (93.2)
Did the action group include students from a range of different backgrounds?	A very good range/quite a good range	35 (81.4)
	Not a very good range/not a good range at all	8 (18.6)
Did the action group involve a range of different staff from across the school?	A very good range/quite a good range	42 (95.5)
	Not a very good range/not a good range at all	2 (4.5)
Was the needs assessment report useful in helping the action group decide what actions to take?	Very useful/somewhat useful	40 (97.6)
	Not very useful/not useful at all	1 (2.4)
Was the external facilitator useful in ensuring that all action group members could have their say?	Very useful/somewhat useful	43 (100.0)
	Not very useful/not useful at all	0 (0.0)
Was the external facilitator useful in helping the action group decide what actions to take?	Very useful/somewhat useful	43 (100.0)
	Not very useful/not useful at all	0 (0.0)
Was the external facilitator useful in helping to ensure that actions were actually implemented?	Very useful/somewhat useful	39 (95.1)
	Not very useful/not useful at all	2 (4.9)
Was there someone from this school on the action group who co-ordinated it and showed leadership?	Definitely true/partly true	44 (100.0)
	Not true	0 (0.0)
Do you think the INCLUSIVE project was a good way to ensure that students contribute to decision-making at this school?	Very good/quite good	41 (93.2)
	Not very good/not good at all	3 (6.8)
Overall, do you think the action group made good decisions about what actions to take?	Very good/quite good	43 (100.0)
	Not very good/not good at all	0 (0.0)
Do you think the action group made sure that these actions were implemented?	Yes	34 (79.1)
	No	2 (4.7)
	Not sure	7 (16.3)

The student curriculum was delivered to year 8 students (hours ranged from 7 to 12). Our curriculum was devised as a set of learning modules that schools could address in locally decided ways, using either our teaching materials and methods or their own existing approaches, if these aligned with our curriculum. This reflected our tailored approach whereby our curriculum was intended to complement and, when necessary, extend and deepen social and emotional curricula already delivered in schools, rather than 'reinventing the wheel'. Schools varied in terms of which of our teaching materials they used (see *Table 3*). At all schools the intervention team agreed with school managers' decisions regarding their choice of social and emotional learning materials.

Finally, all school SMT and action group members interviewed reported that the intervention was feasible and acceptable (RQ2) and their views are reported in more detail below (*Process evaluation: participants' experiences*).

Trial feasibility and acceptability

The acceptability of randomisation (RQ3) was explored in the semistructured interviews with school SMT members at each participating school, who consistently reported acceptability. Their views are reported in more detail below (*Process evaluation: participants' experiences*). Randomisation occurred as planned, and

was acceptable to all schools. The retention of schools (RQ4) was assessed, as were response rates by school (RQ5) at follow-up. No schools dropped out of the study, and student survey responses rates were 91–94% at follow-up in the intervention schools and 87–96% in comparison schools.

Process evaluation: participants' experiences

Our second objective was to explore students', school staff members' and facilitators' experiences of implementing and trialling the INCLUSIVE intervention and how these varied across the different school contexts in order to refine the intervention and trial methods. In order to meet this objective, we collected and analysed data to address the following questions: (RQ6) what are students', school staff members' and facilitators' experiences of the intervention, particularly in terms of whether it is feasible and acceptable?; (RQ7) how successfully was each component implemented and how did this vary according to school context?; and (RQ8) how acceptable were the research design and data collection methods to students and staff?

In this section, RQ6 is addressed first, drawing on qualitative data from both interviews and focus groups to provide an overview of key themes. To address RQ7, these qualitative data are also drawn on to examine the implementation of each intervention component in detail and explore contextual variations; quantitative process items included in the follow-up student and staff surveys are also reported. Finally, to address RQ8, participants' experiences of trial recruitment, randomisation and survey methods are reported, in order to refine our methods prior to a Phase III trial.

Intervention feasibility and acceptability: key themes

Students and staff consistently reported that all the intervention 'inputs' (needs assessment survey, action group, external facilitator, staff training and curriculum components) and restorative practices were appropriate and acceptable to them. Participants' views on each intervention component are described in detail in the next section (*Implementation of intervention components and contextual variation*). In this section, the acceptability of the intervention is reported as described by students, staff and intervention providers, as well as some challenges to implementation regarding intervention co-ordination and project timing.

Student approval: 'having a say', 'having respect'

Key cross-cutting themes in student and staff accounts were that the focus on students 'having a say' and greater 'respect' for students' views were important sources of acceptability and might, as intended in our logic model, be a key component in terms of promoting student commitment to school and thereby reduce aggression and bullying. For example, year 8 students consistently reported that the needs assessment survey and the action group had given them a greater 'voice'. Furthermore, the action group was seen as a new 'experience' to be more 'involved' in 'having a say' at school. This provided a strong motivation for joining action groups. For example, one student explained:

I've never been involved in things like that before, so I thought it would be a new, good experience.
Action group student, year 8 female, Goldstone Park

Specific practices arising from the intervention, such as circle time and restorative conferencing, were also seen as acceptable and ensured that students across the school felt they had more 'voice'. Students frequently reported that they considered such practices appropriate because they addressed both the tendency of some school staff not to 'listen' to students and key problems in schools such as unhappiness and fighting:

A PE teacher, he did it [circle time] in PE, before the lesson we all sit in a circle on the benches and we'll just speak, like how we're feeling and what's made us feel like that.
Male student, year 8, Whitehorse Road

There are fights, there are situations that like need restorative justice conferences.

Action group student, year 8 male, Williamson High School

Changing the school environment in order to make it 'calmer' and more 'respectful' was also considered a very appropriate aim and also motivated participation of students in the project:

Yeah, I thought that because [with restorative approaches] you're saying things in a more calmer way they'd want to listen more and take in what you're saying [. . .] It is needed in this school!

Action group student, year 8 male, Williamson High School

This perception of a 'lack of respect' from teachers and other staff at their school was an overarching theme among students' accounts of their school life and a key source of frustration, disengagement and conflict. For example, one student explained:

We're the students, yeah, it's 'you're excluded' – teachers make things sounds worse than they actually are [. . .] and then you have to respect them all the time.

Female student, year 8, Whitehorse Road

A female student, who reported poor relationships with some staff at Goldstone Park and had already become quite disengaged from school during year 7, reported that her motivation for joining the action group was partly a result of such frustration:

Teachers think we have to respect them, they don't give us respect back.

Action group student, year 8 female, Goldstone Park

Students also often reported that there were 'only one or two' teachers or other staff members to whom they felt comfortable talking. Overall, students felt that the action group and restorative practices 'could work', as per our logic model, for improving student–teacher relationships in this context, which was a strong source of acceptability to them. A student member of the action group at Williamson High School explained that 'slowing down' the disciplinary process to listen to students was needed to improve relationships:

As teachers obviously you have to be strict but there are times, there are like limits like instead of [using] on-calls [internal exclusions so much]. If I've sworn, if I've been totally rude to the teacher, my behaviour's bad, yeah, but for doing a little thing straight away, bang, on-call, that's, I just think it's ridiculous, that's why I think it would work on the teachers, tell them to sort of you know, slow it down!

Action group student, year 8, Williamson High School

Staff approval: an 'attractive' locally led process of 'positive' behaviour change

School staff members were consistently supportive of a new whole-school restorative approach for addressing bullying and aggressive behaviours in English secondary schools, which provided a strong source of acceptability and had prompted interest in participating in this pilot trial. One head teacher explained:

I guess what attracted us was the idea that, you know, we could get something out of it [. . .] I mean, we've been trying to develop a kind of, more restorative approach to secure behaviour over a period of time [and] so this fitted in with the way we want to work.

Head teacher, Whitehorse Road School

School SMT members consistently reported that they were already aware of some restorative approaches being used either in their school or elsewhere, although, if they were being used at their school, they consistently reported that such approaches had been used inconsistently to date and not employed in a

whole-school manner. The intervention was therefore highly attractive because it could provide a new framework, process and resources for 'embedding' restorative practices more coherently and consistently across the school:

The reason I wanted us to be a pilot was because I was very interested in the concept of the restorative justice and I had tried to incorporate that to some extent in our systems [. . .] I wanted it to be embedded across [this] school.

Head teacher, Rainside High School

The intervention was also attractive to the new head teacher at Goldstone Park, who opted to participate in the pilot trial in his first year at the school because he felt it addressed what the school 'needed to do' most urgently in terms of students' behaviour:

[I thought the intervention] sounded really interesting, and would be perfect, and it was just before I started here [I found out]. I knew some of the issues, because I had worked with the SMT on where we needed to go [prior to coming into post], and what we needed to do. [. . . For example] staff not always dealing with students in a particularly positive way, so by that I mean, quite a lot of shouting, a lot of negative feeling about, when they're dealing with behaviour.

Head teacher, Goldstone Park

The adaptability of the intervention, in contrast to overly prescriptive, 'one-size-fits-all' interventions, was also a strong motivating force and a source of acceptability to school managers. They felt that the process of implementation would meet local needs regarding behaviour change and that they could 'make it work' locally for their own specific institutional context. The new head teacher at Goldstone Park explained:

Were there any potential negatives going into it? I don't think so because, you know, with these sort of things you think we're going to make it work for us. So, if problems arise we'll address them and if we need to adapt it for our context and our circumstances we'll do that, so I didn't really have any doubts.

Head teacher, Goldstone Park

The focus on adaption to the local school context avoided schools having to 'reinvent the wheel' and allowed them to build on their existing work. The assistant head teacher at Rainside High School, who co-ordinated the intervention locally, emphasised how the restorative practices that staff were trained in during the intervention both complemented and added value to what they had done in the past, as well as supporting school managers' long-term vision and philosophy:

The checking in, checking out, circle time, the idea of restorative language with staff and students, which again goes along with things we've done in the past in terms of how you all speak to each other. That is an area that we could develop as a school, you know, that every interaction we have, needs to be with that philosophy behind it.

Assistant head teacher, Rainside High School

Other school staff were also supportive of this new structured process to formalise and extend restorative practices across the school, some of which had already been delivered, albeit inconsistently. A senior member of support staff explained:

Selling it to staff in a really positive way is the key: 'I'm not bringing you something new, you're doing this. What I'm bringing you is how we can make it more structured. What I'm bringing you are the skills to make you better at it'. Because there are key questions that you should and shouldn't be asking when using RA [restorative practices] [. . .] That's what the training and the formalisation does.

Pastoral support manager, Goldstone Park

Members of the SMTs and other staff also consistently reported how schools are 'restless' to improve staff–student relationships and address aggressive behaviour more effectively. The intervention was seen as a helpful 'push' and source of community mobilisation within this context. Staff members at intervention schools also reflected on the value of this 'external push', which was provided by the creation of action groups, the provision of an external facilitator and the needs assessment data, which identified clear priorities to inform the process of addressing behavioural problems. For example:

It's the external push which I think is always really important, because if those meetings weren't happening regularly and somebody from outside wasn't coming in to make sure they were happening . . . I'm not saying it wouldn't happen, but I think sometimes that external support, that external push, external pressure makes sure that those things do happen.

Head teacher, Goldstone Park

The education 'market': supporting adoption and implementation

School managers and the intervention team reported that a strength of this intervention was its fit with national education policies and schools' 'core business' of teaching and discipline. In particular, a restorative approach to changing behaviour in schools was seen as highly compatible with institutional priorities and the criteria against which schools are assessed. This was therefore a further source of acceptability, and facilitated both interest in the trial and implementation of the intervention for those schools allocated to deliver it. For example, the desires to promote a safe environment and more supportive teacher–student relationships were consistently reported as institutional priorities in order to raise attainment. The head teacher at Railside High school explained why improving relationships was intrinsic to achieving their priorities:

Our key priorities are raising attainment through quality teaching and learning, and, within that, there are very specific priorities [. . .] [One specific priority was] behaviour, behaviour for learning in the classroom, but also behaviour around the school, that creates a safe environment. [. . . So] if we don't address those we're never going to allow them to achieve their potential academically, so that's why this fits in. For us it's not an interesting project that we tinker in because everything else is running along finely, it's intrinsic to what we're trying to do in terms of raising attainment.

Head teacher, Railside High School

School managers also felt that the intervention was timely regarding the strong focus within the national inspection framework on managing conflict and aggressive behaviours, and this approach appears to be appropriate both for schools developing new school improvement plans, in light of previous Ofsted reports that identified this as a specific priority, and more generally for those schools that were planning and preparing for future inspections. A member of the SMT team at Whitehorse explained how this project would 'hopefully come out' positively for them during an Ofsted inspection:

We are due to have an Ofsted inspection next year [. . .] we would like to think that when they speak to students, for example, they will be able to talk about this, whether it's 'I've been involved' or 'I've noticed there's been a change', or 'we studied these units in personal development' so hopefully it will come out. But also more directly . . . I'll probably have a conversation with them [the Ofsted inspectors] about behaviour for learning and one of the things I will talk about is that we have, you know, we spent this year looking at our behaviour policy in relation to this project and you know, we decided to go forward with more restorative practice.

Deputy head, Whitehorse Road

Goldstone Park was the subject of an Ofsted inspection towards the end of the 2011–12 academic year. Inspectors praised the school's new approach to aggressive behaviours and its initial successes, which

helped build further whole-school support for the intervention and a restorative approach, following this 'glowing' report:

I think we've had some success. [. . .] Do they all really believe it yet? I think that's to do with results and outcomes isn't it? I think we've started to get that because of the Ofsted report, because of all the positive feedback on behaviour, because of our survey that we did with students and parents, came out very positively [. . .] This really tells us, or gives us the evidence that things are changing. Compared to the previous Ofsted before, this says the students' behaviour is excellent.

Head teacher, Goldstone Park

Attendance and exclusion rates are also key 'metrics' for head teachers, especially in schools with a high proportion of students from lower socioeconomic groups or higher rates of students with special educational needs, which was seen as a further incentive to adopt new restorative approaches to aggressive behaviour. One head teacher explained:

One of the measures [we] use is attendance and we've gone up from 92.8% to 93.4%. [and] we have challenging children in our school, and we have a high number of fixed term exclusions [. . .] You have to learn different strategies, and while we had lots of different strategies which we could offer, this seemed to be a new strategy which seemed to be much more organised which we could use.

Head teacher, Williamson High School

Anecdotal evidence from Williamson High School suggested that the approach may have contributed to improved relationships and behaviour at the school, as a reduction in the number of internal ('on call') and fixed-term (external) exclusions was observed:

There seems to be a notable reduction in on-calls [this year . . .] It's very difficult to say but I [normally] get an on call sheet out on each Thursday, and there are peaks and troughs on that, but it does seem to be, I would say, this is an anecdotal evidence – I haven't got the [final] evidence to back it [up but] I will have in the summer holidays when the year is complete – but it looks to me like a one third reduction in on-calls.

Head teacher, Williamson High School

Implementation challenges: co-ordination, communication and timing

Although school staff reported that the intervention was valued and appropriate, it was most acceptable once all school staff understood how the different components fitted together and how it 'all tied in' with their work and roles.

However, other staff reported 'not knowing enough' and that it was not always clearly communicated to them how it all 'married up'. For example:

I can see how it would help relationships but at the same time I think we don't know enough, like, I don't know enough about what, like I said, the aim is for the school. So I don't know how exactly it fits into the action group.

Action group member, female staff member, Goldstone Park

It was suggested that a clear, more engaging, launch to all school staff as part of the half-day training and information session at the very start of the year would support communication and facilitate a better understanding and interest across the whole school. Improving the intervention manual and resources to include more 'real-life' examples was also suggested as a means to improve communication and support co-ordination:

The one thing schools need is a model, of how it's going to work in the school, in a real life school, so that they can almost touch it, taste it, feel it, and then start implementing it in their own schools.

Head teacher, Railside High School

It was also suggested that ‘front loading’ the enhanced three-day training at the very start of the project would further facilitate whole-school co-ordination and communication via ‘cross-fertilisation’ and the diffusion of information by local ‘champions’:

You always need cross-fertilisation [. . .] the ten of us [who did the enhanced three-day restorative approaches training] could be spread out so that, or we could have given feedback, [been] the champions of it to support it [. . .] She’s obviously got lots of materials, so you don’t want to keep re-inventing the wheel [but] we could have then actually helped lead it and spread the messages so it [had] more ownership to the school really.

Head of year, Rainside High School

Members of the SMTs at intervention schools also reflected on the need for greater clarity regarding what time commitment was required of staff to implement the intervention:

It was very positive. I mean I haven’t got anything to say negative, just like I say it’s, you know, any school is going to [be] just busy really. I suppose all organisations are, but that’s just the way it is. And so the school needs to fit this project into how it works and that can be tricky at times [. . .] but I think we underestimated the commitment and involvement further on.

Deputy head teacher, Whitehorse Road School

The issue of ‘timing’ was a recurring theme in intervention school staff members’ accounts and seen as a challenge to implementing and co-ordinating multiple components across the whole school. There was a consensus that the intervention would benefit from earlier recruitment and randomisation of schools to support forward planning. This, in turn, would support more timely and effective delivery. One facilitator was previously a head teacher and this enabled him to highlight the challenges associated with not being able to organise staff training and meetings in advance of the new school year:

[Head teachers’] planning starts at the end spring, beginning of summer term [for the next academic year]. A lot of the working through [so the head teacher] really needs to have that written into their planning and because it wasn’t written in it made trying to get in space for more training days, in meetings with staff, very, very difficult.

Bryn, action group facilitator

It was acknowledged by school staff and intervention facilitators that a major strength of the intervention was its responsiveness to local needs and priorities, and that it recognises schools ‘start from different places’, as one of the intervention team put it. However, the intervention co-ordinator felt that the intervention should have begun more quickly after randomisation. This would have enabled facilitators to draw on the needs assessment data to engage school SMT and other staff:

The schools didn’t get the data quickly enough, and one of the things I was selling them on was that they were gonna get high quality data that was expensive and they were getting free.

Vanessa, intervention manager

The pilot study thus established that a full trial must begin recruitment no later than midway through the academic year before the academic year in which the intervention would be implemented, with baseline surveys undertaken before the summer holidays, in order to support timely intervention planning and implementation. There was also a consensus that the intervention should extend beyond one school year. School and intervention staff emphasised that few benefits would be expected after only one academic

year because it takes longer than that not just to implement all the components, but to modify the school culture:

This project, it's just over 1 year [. . .] This is all about changing the school ethos. Change school ethos in a year? Well, you're gonna start nibbling at it but [. . .] to do it after 1 year I still am very sceptical about that, whether in fact you can measure any impact.

Bryn, External Facilitator

A recurring theme was that the pilot intervention had been successful at initiating changes, particularly considering the truncated timescales, but that real impacts in the school would manifest only in the following school year. For example, all the intervention schools described how they expected to see greater use of restorative practices the following year, after the pilot trial had finished, which is further evidence that both a longer-term approach and a longer period of follow-up are required than were possible in this pilot trial:

We'll see that more next year, because like I say, the way in which the project has worked out for us, is very much putting things in place for next year, and I think one of the key things, will be restorative language and how well all of our staff are able to adapt to that and use that.

Deputy head teacher, Whitehorse Road

What I wanted to achieve is restorative justice. This is another strategy and we have trained professionals who will now train you on that technique and we will try and use it, and [. . .] then next year it will be embedded into the school and it will be [in] our new behaviour policy, it will be mentioned in our pastoral handbook and it will be mentioned in guidelines for new teachers . . .

Head teacher, Williamson High School

The intervention team's views: getting a 'head start' to build on a 'success'

The training provider, the intervention manager and the other two educational consultants who also facilitated school action groups during the intervention were highly supportive of this whole-school restorative approach, and all reported that they felt it was a 'success' based on their observations. They all reported that the intervention was acceptable and feasible across all four schools, with all participants supporting the notion of a 'universal' approach that could be adapted to different institutional contexts and was responsive to local needs. The intervention team also suggested that a whole-school restorative approach may have a particularly 'natural' place in the most 'challenging' and 'needy' schools with very high proportions of disadvantaged students, which suggests the intervention's potential to address inequalities in adolescent health outcomes:

I think, it's quite interesting that the schools that were arguably most needy implemented it best.

Vanessa, intervention manager

The intervention team were also positive about sustainability. One action group facilitator explained that the intervention was not 'just going to stop' at Railside High School, which was a testament to its feasibility, acceptability and value. It was also seen as addressing an important gap in secondary schools by specifically giving younger students a voice and greater responsibility, which typically does not happen until they enter year 11:

A lot of pupils in primary school have a lot of roles and jobs, and then as soon as they get to secondary school, they don't. They're not given these roles and responsibilities until year eleven. And [the action group students] very much took on board this and got people responsible.

Vanessa, intervention manager

It was suggested that we capitalise on the adaptability of the intervention to the local context as a source of acceptability and motivation by making it 'even more bespoke' and locally adaptable, including through greater integration with school improvement plans:

It has to be more bespoke. [. . .] I mean, essentially, what you have, we should have gone in and asked the schools to talk to us about their School Improvement Plan to start off with, or their School Self-Evaluation [for Ofsted], that's supposed to be their core document, and say, 'What are your priorities?' and consequently, 'Which ones are going forward? Which ones are you resolving? Which ones have you got obstacles? And how do we fit into that?'

Vanessa, intervention manager

Intervention team members also saw the broader policy environment as a major facilitator to implementation, as reducing aggressive behaviour is part of schools' 'core business' and a priority for SMTs, especially (although not only) at the most 'challenging schools'. For example, it could give them an important 'edge' for future Ofsted inspections, as well as address ongoing concerns for all secondary schools regarding 'disaffection':

I mean if you think how important Ofsted and that sort of thing is to them, anything that can give them an edge is really important!

Dawn, action group facilitator

Every secondary school has an issue with disaffected kids from year 9 onwards, every single school. I don't think there's a school in the country that could say they didn't. [. . .] No one's solved disaffection yet.

Vanessa, intervention manager

The greatest testament to the intervention's acceptability among staff and students, and their motivation for a process of whole-school restorative action, was that Whitehorse Road continued to implement all the intervention components during a turbulent year in which staff and students discovered the school was to be closed down:

It's a school that halfway through the intervention they decided, the governors decided, they were going to close the school. They're not taking any new intake from this September, and that [. . .] of course that had an impact on, around the school and it's to their credit that they want to remain in the project and indeed want to continue developing their work with restorative justice, despite that, you know, that distraction.

Bryn, external facilitator

Implementation of intervention components and contextual variation

The following sections are structured to address each of the key intervention inputs and processes in turn (the needs assessment survey, the establishment of a new action group, and external facilitation of the action group, the staff training and student curriculum), how implementation varied among schools and how they should be refined, with specific recommendations outlined for each key component. A case study of Goldstone Park's action group is also included to highlight an example of successful implementation and the school-level changes that can be achieved, even when the intervention is for only 1 year (see Box 3).

Local needs assessment survey: identifying priorities for action

School staff consistently reported that the needs assessment survey allowed them to see the 'big picture' and identify priorities for action, in terms of both the nature of aggressive behaviours at their school and the extent of 'risk' and 'protective' factors for conflict and aggression within the school environment. In

some cases, the needs assessment data provided confirmation of existing concerns and additional motivation to mobilise the school community towards a 'positive' change:

Did we believe the results? Yes probably. It was done with year 8s, which was our most concerning year group at the time. [The survey found that] they were feeling quite negative about the school and I think that's because they were getting lots of negative messages again, and it was because we were enforcing them: 'you are the worst year group in the school, your behaviour is terrible, stop behaving so badly', all from a negative perspective. We're hoping that the second round of the results comes back better and shows improvement, because what we've tried to do, alongside all of this work, is to keep re-emphasising the positive.

Head teacher, Goldstone Park

In other cases, some of the findings were reported to be a 'surprise', which led to new strategies and actions to address problems likely to be associated with student aggressive behaviours. For example, at Railside High School students were reporting high levels of disengagement and low aspirations for the future, which informed the action group plan:

The issue around aspiration and hope, and their futures . . . I think that actually was the biggest surprise, because as a school, that's something we do work on all the time so that, kind of, well that's not good! [. . .] And that's why we planned the horizon day, which was designed around looking at them, making them feel first of all better about themselves, better about school, better about the future, and that's what the whole day was about.

Assistant head teacher, Railside High School

Staff and students in the action groups were also supportive of the use of a large-scale survey. For example, they knew the data were representative of the wide 'range' of year 8 students at their school, and this was a 'useful' source of motivation to address 'aggression levels':

It was quite useful to us because now like, like the aggression level was quite high so now we're trying to think of ideas to like to get it back down.

Action group student, year 8 female, Railside High School

Our facilitators all reported that the needs assessment process at the start of the intervention was highly appropriate, 'powerful' and instructive for identifying priorities:

I think it was useful, I think it's always good to see data [. . .] and I think without the data they probably would have just completely dismissed [some issues] so at least we did have something there, to say, 'Well they said this you know'!

Dawn, action group facilitator

Although schools were keen for an 'external push' (a key source of acceptability, as reported above), some SMT staff did report that the needs assessment felt too 'negative' at times, especially among more established SMT members who had been in their post for several years and might have seen it as a reflection on their many years of work and leadership at the school. Intervention facilitators also picked up on this and felt that schools may 'just see the negatives' if data were not presented appropriately and may become 'too defensive', although this 'defensiveness' also prompted them to mobilise and analyse other data sources.

The educational consultant who co-ordinated the intervention across all four schools and facilitated the action group at Goldstone Park highlighted another potential problem of the needs assessment report relating to how schools (and their data) are now scrutinised by Ofsted inspectors. She felt that not only were some staff not always 'happy' with the issues reported, but also this survey could reflect badly on the

institution during a future Ofsted inspection, which was an additional reason why school staff would be 'uncomfortable' with the needs assessment survey if it was too 'negative':

They [Ofsted] expect to see the school surveys [so] if anything comes out that says their kids feel unsafe, they're very worried about it, not only because they don't want the kids to feel unsafe but if Ofsted were to come in, they could rate them 'Inadequate', so they were very uncomfortable about that, that kind of thing.

Vanessa, intervention manager

The process of assessing local problems and needs appeared to be more acceptable when it was not only informed by the student survey but also involved drawing on other sources of routinely collected data and existing documents (e.g. attendance and exclusion data, parent surveys, auditing incident reports, etc.) as well as other 'softer' sources of information, such as the views of students in the action group and other 'student voice' groups. For example, at Williamson High School the external facilitator reported that the action group audited and analysed its own, routinely collected, statistics on behaviour and exclusions. Other schools established new 'forums' and/or carried out 'focus groups' to gain more in-depth views from students following the needs assessment survey report. The greater use of data and the establishment of a new action group also inspired further large-scale surveys at Goldstone Park, which the head teacher commissioned to monitor and celebrate the 'positive' changes, as well as to continue to identify ongoing challenges:

The survey [at the end of the year] that we did with students and parents, came out very positively [...] We did it after Easter, so April, May time. We did the whole school. And we had a huge response from parents. So I think some of the senior leadership team were a bit shocked [...] but we had a third of parents respond [...] Broadly, the parents feel that the school does care, and that students achieve well in the school, and teachers want them to do well.

Head teacher, Goldstone Park

The only note of caution regarding the use of other sources of data came from the head teacher at Williamson High School. She explained that English secondary schools were increasingly 'data-rich' environments and triangulating multiple 'sources of information' could be problematic.

In addition, although the presentation of the needs assessment data was generally well received, various ways to improve this were suggested. First, more careful and judicious use of 'benchmarking' against the average across intervention schools is required when presenting data to schools, particularly in terms of students' aggressive behaviour and what is 'not good':

[Benchmarking each schools data against the 'average' across the four schools] kind of confused the issue sometimes. [For example, if] sixty percent of year 8 students felt they were being bullied and at school X and you're below the average, that's still not good! [...] I know why the project used that comparison data because if I'd been a head I'd have asked that question, 'Well what about the other schools?' – but that can lead you up some blind alleys.

Bryn, action group facilitator

The data must also be reported clearly and presented in a way that is also accessible to students. Some reported that the data would have made more sense to them if they had been presented differently. For example:

It is and it isn't [useful], because, like the percentages. You don't really pay attention, 'cos normally they're just like in a little space in the page or somewhere and you won't really see them. Like I remember, I got the thing back with the class average [school mean] or whatever it was and we didn't notice them until the teacher had pointed out them.

Action group student, year 8 female, Goldstone Park

Needs assessment survey: recommendations for a Phase III trial

This was an acceptable and powerful 'external input' that helped all the school action groups to identify key priorities and should remain integral to the intervention approach and logic model, although it could be improved through:

- an approach that identifies both the 'positive' and 'negative' features of the school environment, including not only risk but protective factors for aggression and bullying, and school 'assets'
- continuing to compare each school against the average for intervention schools, but ensuring facilitators aid in the interpretation of this, including through also benchmarking against other schools with a similar socioeconomic intake as well as the average overall
- ensuring all reporting is accessible and student centred
- using annual surveys in intervention schools to monitor progress and identify new/ongoing priorities.

Action groups: including students and staff as agents of change

An action group was established at each school and met at least six times during the school year. Each action group included at least six student representatives. Members of schools' SMTs acknowledged that this was an important innovation for their school. For example, one head teacher explained:

The action group has been really important. It's the first time in the school, as far as I understand, that you've had students working with staff. And properly working with staff.

Head teacher, Goldstone Park

School managers who were responsible for implementing the intervention described recruiting a diverse range of action group members and how 'positive' this process had been:

I think in terms of people coming onto the [action group] team, initially it was very positive and we had, in fact, we had too many people probably. I got as diverse a range as possible; there were quite a few students who were keen and involved and they came to the half-day training as well. [. . .] I wouldn't say [it was just] goody-goody high-attainers.

Assistant head teacher, Railside High School

The recruitment and retention of a range of students was seen as central to the success of the action group. Increased student participation in decision-making was central to our logic model, as it was hypothesised to improve staff–student relationships and students' commitment their school, resulting in reduced aggression and bullying. Three out of the four schools chose to recruit year 8 students only, thus allowing them to build on the year 8 needs assessment survey, although limiting representation from across the wider student community. As with the needs assessment process (described above), several staff members and the intervention team felt that broader representation from students across years 7–9 would be appropriate and this could then be 'branded' as a key stage 3 intervention to support its sustainability. An action group member explained why she felt it was particularly important to include the younger students:

Year 7s would have a different perspective to it: they've just come in, they're the ones who, who face all those things. And then they're the ones who will tell you 'when I . . .', the period of transition, 'when I was in year 6 that's what happened, and now I'm in year 7 this is . . . these are my difficulties'. Bullying, all those things, they happen to be affected more because they're the ones who enter into the system

Action group member, English teacher, Railside High School

Students in the action group at Williamson High School, in which some older students were included from year 10 and the sixth form, felt that involving older students was helpful for a 'third perspective' alongside that of younger (key stage 3) students and staff:

They've had more of an experience at the school already, they're sunk in to the school, and to hear our views about how we feel because we've, we're coming up to two years now and so, yeah, it's good to get different views from different years.

Action group student, year 8 male, Williamson High School

It appeared that using multiple methods of recruitment is acceptable to students and may be the most appropriate method for recruiting a diverse range of students. Students reported that they either 'got chosen', 'got asked' and agreed or volunteered after finding out about the group in an assembly, newsletter or via word of mouth from a friend. One school used a drama workshop on bullying to publicise the new project, which was popular with students and could be used to launch the project elsewhere in the future. For instance, Goldstone Park pro-actively 'asked' and encouraged some students who may not have volunteered to join the group to ensure diversity in terms of more and less engaged year 8 students. A pastoral support manager at the school explained why she thought this was appropriate:

It was announced obviously in the assemblies and things like that but we do approach students sometimes and say, 'look you'd be really good on that' because I think especially with your more colourful students, they don't apply, thinking, 'well, I'll never be allowed'. [Then you've] just got to sort of have a chat with them and say, 'you'd be really good at that, why don't you ... ?' – [they say] 'we're not doing that' – 'no, go and apply!'. And they do. And as you can see at those meetings, thoroughly enjoying it [...] And you do need them involved, because if they're on-board they're going to make sure that everybody else is involved, do you know what I mean? All the other students ... because they are known to have some issues and problems from time-to-time, then students – I don't know – respect them more, listen to them more. If they think it's a good idea then it can't be all that bad!

Action group member, pastoral support manager, Goldstone Park

Students in the action groups also reported that it was appropriate for some students to be asked to participate, to ensure diversity, but still allowed to volunteer rather than being coerced:

[The deputy head] he asked me [but] people, like, volunteered to do it and they weren't just [made to ...] random people weren't just told to go! They wanted to be there. So, they contribute instead of just sitting there 'cos they don't want to be there.

Action group student, year 8 male, Goldstone Park

Students and staff also suggested that the requirement for the action group to include management, teaching, support and administrative staff ensured that the group represented 'different perspectives' and could 'paint a picture' of the school to inform the development of an appropriate action plan:

There were governors. There were all different [staff]. There was parents, and there was some non-teaching staff as well, like the ones who do not teach in classes were there – one of them is a lab assistant [...] so these categories, they make up the school staff if you like [they are] going to have a different perspective ... It's going to be a different picture that we can paint.

Action group member, English teacher, Railside High School

The presence of senior management staff was also seen as particularly critical. The intervention team felt a key learning point was that a senior member of the SMT (i.e. head or deputy head) with 'power' to change wider policies was likely to be essential to the composition of an effective school action group. This was not the case at Williamson High School, where a relatively new 'assistant head' represented the SMT and changes to the school policies and ethos were much harder to achieve. The educational consultant who facilitated the action group at Williamson High School explained the consequences of this for the group, comparing it to another school (Whitehorse Road) where significant school policy changes were achieved:

My reservations are to do with the composition of the group. [At Williamson] they had an assistant head teacher who was part of the senior management team but what I think is important [is that] you have somebody on-board who has responsibility for behaviour policy in there [. . .] you really need to have a person who can lead as a power and [they] need to be there and particularly when you're trying to change the ethos because there was a certain tension between the action group and the deputy head [who was not on the action group] who had responsibility for student behaviour [. . .] Whereas at Whitehorse Road they came up with a school policy which incorporated a behaviour policy which incorporated the principles of restorative justice. I spent some time with the head there working on that and they did that well.

Bryn, external facilitator

The weakness of existing student voice groups (e.g. school councils), which were seen as highly unrepresentative and powerless, also provided a strong source of acceptability for the action group and motivation for joining it, both for students and for staff keen to promote student voice. At all four schools, the action group was seen as 'new' and 'different' from the school council:

Most people don't feedback from the school council, and, to be honest there's not really – well I don't see the point in having the school council.

Action group student, year 8 female, Goldstone Park

Like they say, student council, like 'oh yeah we do it every week, we've done a few things'. And it's like 'what have you done?'

Action group student, year 8 male, Railside High School

The topics are good on the action group [compared to the school council . . .] We actually talk about things that we do here, what's good for the future, stuff like that, like I said in the first place about the CCTV, the cameras and also the teachers is going to be improved and, you know they listen.

Action group student, year 8 male, Whitehorse Road

The schools used the new action group to promote student involvement in decision-making and leadership via a new, more student-led and student-centred environment (Box 3 shows a case study of the achievements at one school). This included encouraging students to chair meetings and allowing students to talk first at the start of meetings:

One school [Williamson High School], they always had a student who chaired the meeting, and [Whitehorse Road] they took the first part of the meeting so it was just students who talked and that gave them time, rather than them trying to make a contribution in a perhaps more adult kind of scenario, both those methods worked very, very well.

Bryn, external facilitator

BOX 3 Case study of student-led action at Goldstone Park

The action group at Goldstone Park was seen by staff and students as a new 'powerful', 'representative' group that identified priorities and acted on these to make comprehensive changes to the school environment. The needs assessment survey highlighted students' poor relationships with staff, with few strong connections with any staff, which were seen as underlying other problems at the school such as 'disaffection', conflict and aggressive behaviours. A specific issue was 'inconsistency' across staff in how they dealt with behavioural issues. This provided a clear focus for action:

The data showed this inconsistency issue of children, the pupil, and staff, the data didn't show staff [views] but it turned out behaviour was inconsistently managed and I think that they really ran on that well.

Vanessa, intervention manager

The success of the action group at Goldstone Park also appeared to be facilitated by 'doing stuff' straight away to ensure 'early wins', which both increased students' interest in the project and built trust between students and staff in the action group. For example, after the second meeting, when the needs assessment data were presented, the deputy head enacted students' suggestion to access a wider range of views and ideas from other year groups about the school's 'inconsistent' practices and 'unfairness' via a suggestion box and student focus groups. The students in the action group reported that they quickly realised this was 'different' from the school council and a wide-ranging action plan, informed by a range of data sources, was initiated. This involved both 'key' policy changes (e.g. the behaviour policy and rewards systems) and innovative student-led initiatives, such as a new school 'blog'.

The action group also supported the co-ordination of different intervention components. For example, the short duration of tutor periods at the school was identified as a major barrier to improving relationships between students and their tutor and also as preventing 'checking in' and 'checking out'. A key action was that the school day was reorganised to ensure that 'tutor time' would have a much greater focus on the social and emotional aspects of students' learning, and allow for greater pastoral support:

Now tutor time is very short here. Now this is one of the things that's come out of the action group. . . . So it looks like we're going to have a much bigger tutor time where we're going to be able to implement a programme through the year, and I would definitely envisage that SEAL [social and emotional aspects of learning] would be a part of that if it's not SEAL itself it would be SEAL related, because we're looking at their emotional development within tutor time and checking in and checking out.

Action group member, pastoral support manager, Goldstone Park

The success of the action group had also ensured it was seen as sustainable and could 'carry on':

The action group, has probably been the best part of it. That's what's been really highly regarded and valued. And it's kept that group going and made it a very, very effective group. And I'm sure they will want to carry on.

Head teacher, Goldstone Park

Senior managers, including head teachers, also recognised, and were motivated by, the need to increase student voice and the limited representation of students in existing groups. The action groups were seen as extremely useful in addressing this deficit in students' 'perspectives':

It was very much a 'done-to' climate the ethos in the school, and staff thinking they knew best. And some of those staff, and some of my senior team who don't have their own children as well, and I've talked a lot to my deputy about this, and those of us who have our own children, do tend to see it a bit more from the children's perspective

Head teacher, Goldstone Park

These data support the view that the emphasis within the intervention on student voice was highly acceptable and potentially a driver of improved staff–student relationships and increased student engagement. At Williamson High School, the students in the action group also attended the enhanced 3-day restorative practices training, which appeared to be an appropriate innovation for supporting students' 'understanding' and engagement with the concept of whole-school restorative change.

Several students reported that they felt that they had gained 'good experience' by participating in the action group and other related activities, such as the training. Students and staff also consistently reported that they felt that they had got 'credit' for their involvement. However, both staff and intervention team members suggested that involvement and retention may be further improved by more formal accreditation and 'additional rewards' for students. The intervention manager also suggested recognising staff involvement through better integration of the intervention with continuing professional development initiatives.

Finally, it is also important to recognise the practical barriers some schools face when organising meetings after school for students to attend, especially at schools in London and/or with large catchment areas, where students have long journeys home via public transport.

Some students suggested lunchtime meetings may have been better at Whitehorse Road, but, at the other schools, students felt that lunchtime was 'the only time to socialise', which further supports an adaptable approach to organising these meetings based on schools' local context and students' views. Providing tea, coffee, juice and biscuits also appeared to be popular with students and some staff, encouraging them to stay after school and also facilitating a 'grown-up' environment.

Action group: recommendations for a Phase III trial

The action group was an innovative and powerful mechanism for supporting student-led change to address key school-level risk and protective factors for aggressive behaviour. Students throughout schools became aware of the work of action teams and this might be an 'active ingredient' in improving students' relations with staff and commitment to their school. Using a range of methods, including directly asking and encouraging less-engaged students to take part as well as launching the project in assemblies and/or via drama workshops, is likely to facilitate active participation of a wide range of students. Schools should also ensure:

- they recruit students from different years (e.g. years 7–9) into the action group
- that the head teacher or (more likely) a deputy head teacher is a member of the group to ensure it has sufficient power to change school policies
- that action group facilitators work with the action group co-ordinator to identify the best time(s) for meetings locally and help them consider any practical barriers and how these might be overcome
- they provide a 'grown-up' environment with hot drinks and biscuits.

External facilitation: supporting school managers, advocating for students

Three external facilitators (one was also the intervention co-ordinator) worked across the four schools to support each action group and help co-ordinate the intervention locally, each working with his or her own school or schools. School SMT members' accounts of implementing the intervention included strong support for their action group's external facilitator, who they felt supported feasibility. Facilitators' experience and expertise was 'handy' to school managers who were co-ordinating the intervention locally, especially at more challenging schools and where it may be practically more difficult to organise new events quickly:

[Bryn] helped me with the agenda, and he takes part, particularly with the students, he asks a lot of questions to the students [...] there's been some other things where Bryn's expertise, if you like, his experience, has proved handy [For example] I make things fit to my timetable so it's handy to have someone like Bryn because there were a couple of times when he was a bit more assertive and I thought, OK! I will do it. So it is handy.

Deputy head teacher, Whitehorse Road School

Those schools that were explicitly motivated by the need for an 'external push', such as Goldstone Park, also strongly welcomed the external facilitator, and explained that the role 'fits' with their motivation for participating in such an intervention and their other priorities:

It just fits with what we need at this moment in time, and what we're doing and this is a huge strategic direction push for us as a school, so it blends in very nicely and helps us with what we're already trying to achieve

Head teacher, Goldstone Park

One of the facilitators explained that she felt that this external role was particularly valuable, as facilitators could 'challenge' schools and provide fresh, 'creative' input outside the local/internal 'politics' of the school:

I think really good listening skills are important and not being afraid to challenge, being objective and not getting sort of sucked into politics, being trustworthy so they feel that you can, they can really say what they need to say in front of you so building that rapport and the trust. And I think maybe a bit of creative thinking because I think you have to help them think of ideas and ways they could work which are different to the ways they work now.

Dawn, action group facilitator

As well as providing an 'external push', the facilitators' role was also seen as flexible enough to allow them to play multiple other roles and to adapt to different school contexts. For example, one facilitator explained how she could simultaneously act as an 'objective' source of support and help 'anchor' the action group to keep it focused.

The external facilitators could also support students in 'having a say' and 'having respect' as part of the process of implementing a new whole-school restorative approach at their school. The facilitators recognised the importance of their advocacy role to support student voice, especially when teachers were defensive. However, it was also suggested that change was most likely to occur and be student-led when teachers and other school staff, *as well as the students*, were all able to 'have a say' during action group meetings.

Two of the three facilitators had been school managers (one a head, one a deputy head) at secondary schools and the other was an educational consultant who worked for a number of education authorities. She felt that because of her knowledge of secondary schools from advising local education authorities, her 'skills' were important in terms of facilitation, advocacy and 'helping groups work'. However, some SMT members suggested that an external facilitator should ideally have school leadership experience.

Finally, all those involved with delivering the intervention reported that facilitators' implementation needed better co-ordination across schools, with some 'collective meetings' to share learning and example of 'best practice' in terms of facilitation and advocacy. The intervention manager described her vision for what this might look like and how greater use of technology might also make the intervention more sustainable:

I think we need a team of facilitators that are kind of coordinated [. . .] I think it would be good as well capacity-wise to be able to build in the cost of them meeting occasionally together and sharing ideas and things, because that didn't happen very much. The other thing [would be] if you were to have a virtual learning environment that was shared between all the schools that were doing it, where they could share resources – I think that would really, the schools would really benefit from that too I think. That would be a good way to support the project, but that wouldn't have to be people-rich!

Vanessa, intervention manager

External facilitation: recommendations for a Phase III trial

The external facilitators were consistently reported to have provided a highly valuable, 'external push' for schools. This support should be maintained in a Phase III trial. The existing facilitators from the pilot trial should be retained, complemented by additional educational consultants, co-ordinated by an intervention manager, meet occasionally to provide mutual support and share good practice. In future, training for external facilitators should be provided by the intervention manager. Key roles and responsibilities of external facilitators should include:

- establishing an effective ongoing working relationship with the SMT
- a 'catch-up' call with schools' intervention lead in advance of each action group meeting to support planning, allocation of tasks and administration
- advocating for both students and staff to promote an equality of 'voice' and effective decision-making involving representatives of the whole school
- supporting the co-ordination of training and curriculum implementation as required at their school(s).

Ideally, external facilitators will be educational consultants with former school leadership experience, although those without such experience but who have a strong track record in facilitating student-centred projects in schools are also likely to be effective in this role. In addition, external facilitators and intervention schools should have access to a virtual learning environment, administered by the intervention manager, in which they can share examples of 'best practice'.

Restorative training: room for improvement

It was widely agreed by both the school staff and the intervention team that high-quality training in restorative approaches was a crucial 'input' if they were to adopt this philosophy and operationalise it across the whole school. The quality of the restorative practices implemented in schools would therefore strongly be determined by the quality of training. Staff and the training provider agreed it was appropriate to undertake an initial half-day introduction for all staff on restorative principles, language and classroom methods (e.g. checking in, checking out and circle time), followed by 3-day, enhanced training for a smaller group on implementing practices such as restorative conferences. The trainer felt that the enhanced training worked best with a 'mix' of staff:

There's certain key roles to be identified but there's some of it to be opened up for these very naturally restorative people that are very passionate about it [. . .] because they are worth their weight in gold. So you do need a mix . . .

Brenda, restorative justice trainer

As well as the problems in scheduling the training reported above, there were also problems reported regarding its quality. School staff had divergent views on the content and delivery style of the training. Several SMT members reported their frustration that, although training was a critically important component and one that demanded considerable staff time and resources, it was also the weakest intervention component. A key theme that emerged from participants' accounts was that the training must be improved by making it more interactive and engaging for staff. The half-day whole-school staff session was viewed as a 'disappointment' because it was 'too didactic':

I found the training very disappointing, I think the first morning was very disappointing. It was the whole staff, it was everybody, and I just, I felt it was something that could have really fired people, and I know that there were some positives that came up, I know that some staff went away and started using checking in and checking out etc., but I think the training was too didactic, it wasn't relevant enough for a high school situation. [. . .] It was a large group, but we could have been divided up into groups and that only happened once in the 3 hours.

Assistant head teacher, Railside High School

Students in the action groups who attended the initial half-day training at their school to find out more about restorative approaches also reported that it needed to be more interactive. For example, students at Railside reported that it was 'boring' and the trainer should 'have made it more active, we was like sitting there for four hours'. Students at Goldstone Park also complained that it was just 'PowerPoint and someone talking' and 'all you got was a cup of tea'.

The enhanced 3-day training in conferencing methods also needed to be more engaging and resonate with secondary school teachers' everyday experiences. Some of the school staff suggested how the methods could be improved to make it more engaging:

It really was death by PowerPoint. And at the break time, which was at 11 o'clock, which was [after 2 hours] you know 9 o'clock to 11, we told her. And by 20 past 11, she had made it practical and interactive and worked out ways the theory could be covered on the tables rather than everybody sitting . . . and you know, fair play to her, it was the first time she'd ever worked with students and adults in a group, you know, she did really well.

Assistant head teacher, Williamson High School

As well as avoiding large-group, didactic methods, the assistant head teachers at Railside also suggested that the training would be more acceptable if it was delivered by:

Somebody who has established this in their school and can talk about nitty gritty things, like the systems, the processes, how they operate the structures on a day to day basis, then also talk about the impact it's had on children, on staff.

Assistant head teacher, Railside High School

Several SMT members felt that the training may be better received if it were delivered by someone with a teaching background. The head teacher at Goldstone also suggested that another way to address the 'credibility issue', if training were delivered by someone who did not have a teaching background, was to ensure it was more realistic.

Related to this was another recurring theme that the training lacked tangible examples focused on challenging secondary school environments:

[Although] the trainer does have examples of secondary, she talked to them in quite a primary way, I think. I think as well, when I sold it to schools, [. . .] I mean, training time for teachers is so rare and so valuable to them, and even if you pay to cover their class, which we did, they don't like it, they'd rather be in the class often. But I think they felt they didn't quite know why, why it was kind of fitting in, really.

Vanessa, intervention manager

The trainer acknowledged that starting with a more comprehensive audit of existing practices would be beneficial for understanding different, challenging secondary school environments, for example by asking initially: *'What is the school hoping to achieve?'* It was also suggested that a more detailed pre-training audit could help tailor the training to schools' needs regarding how they would 'cascade' and 'formalise' new restorative approaches and methods through the whole school. Furthermore, one staff member who attended the training sessions at Goldstone Park highlighted that the restorative training did not identify sufficiently realistic examples:

The last session for instance, the afternoon was, erm, we were given scenarios for role play, and I have to say I'm not a great lover of role play, but we'll do what we have to do. Erm, the scenarios, a lot of them, were too weak to stand up to role-play [. . .] it was more in primary level.

One observation was that training participants needed to be encouraged to engage in further learning and practice their skills between the (3-day) enhanced training sessions.

Staff training: recommendations for a Phase III trial

The provision of staff training in restorative practices was consistently identified as being a critical component in a whole-school restorative approach to behaviour change. However, a number of challenges emerged in delivering this and engaging staff. In future, trainers need to be aware of the situation of each school. Therefore, we propose that the external facilitators be themselves trained to provide training to the schools with which they work. Training would also be improved via:

- ensuring that training is undertaken at the start of the school year to pump-prime other activities and increase awareness of the intervention across the school
- a comprehensive pre-training audit to identify schools' needs, what they hoped to achieve and the most appropriate method to 'cascade' learning through the whole school
- more engaging, interactive training methods using 'realistic' examples from similar secondary schools
- ensuring that students from the action team attend the training, and they are included and engaged in it.

The curriculum component: support for the 'tailored' approach

All schools welcomed the flexible needs-led basis of this curriculum. This consisted of learning modules that schools could address using either the INCLUSIVE teaching materials or other existing materials. This allowed schools to tailor the curriculum to their needs. Schools implemented 7–12 hours of the curriculum, which is impressive bearing in mind the late initiation of the project. The intervention manager felt this represented a success also in light of the limited number of hours available for PSHE:

Citizenship has been cut right down, you know, completely cut down.

Vanessa, intervention manager

Whitehorse Road was already delivering the 'Opening Minds' social and emotional well-being curriculum in PSHE. Nonetheless, teachers found the INCLUSIVE social and emotional curriculum for year 8 students useful and addressed PSHE with a combination of 'Opening Minds' and INCLUSIVE teaching materials. The school delivered 10 hours of the INCLUSIVE curriculum and rated it very positively.

As 'restorative practice' was seen as a major gap in existing PSHE curriculum resources across all schools, and, because of the opportunity to choose modules based on schools' needs assessment data, this

curriculum was seen as a highly attractive opportunity to ‘refresh’ PSHE in line with the restorative principles underlying the intervention:

As somebody who has recently taken responsibility for PHSE in the school, I found getting a whole programme absolutely marvellous and particularly a programme that will add, with adaptation will work for us we’re up to our sixteenth hour [with the year 8 students], so we’ve managed to get through it, but that was a really difficult thing to do, because of the timing, but actually it worked out OK in the end [because] our year 8 [PSHE] programme needed a refresh, so it was a good thing to happen.

Assistant head teacher, Williamson High School

There was also evidence of the curriculum component’s popularity and sustainability at Williamson, where 8–12 hours were delivered to each year 8 class. The PSHE teacher leading this reported that they would ‘use the units again’:

They’ve had altogether about 8 hours each (year 8 class), maybe a little bit more, some of them would have had more because I have them every week so they’ve probably had about 12 lessons with me [...] We’re also going to use the units again because they were very well received as well.

PSHE teacher and action group member, Williamson High School

As discussed at length above, greater planning would facilitate implementation earlier in the academic year (largely through the opportunity for preparation in the previous year). Some staff also suggested further improving the curriculum resources in the future through more interactive methods, which would support both ‘activity and action’. For example:

We need more activity and action in it. A wider variety of learning styles. There needs to be more activities.

Assistant head teacher, Williamson High School

Student curriculum component: recommendations for a Phase III trial

The provision of the student social and emotional skills curriculum was consistently identified as being a valuable component. This would be improved via:

- greater advanced planning and preparation with each school’s PSHE lead
- more interactive activities.

This component could potentially be further developed through new modules in which students can get involved with the work of the school action group (e.g. reviewing policies, rewriting rules and other locally determined actions). This will be explored during further intervention scoping and development work.

Quantitative process measures

We collected quantitative data on process measures in terms of *coverage* (i.e. the reach of intervention outputs) from students and staff at follow-up in both intervention and comparison schools. Our aim was to pilot methods of data collection and analysis rather than to estimate effects. Process data were not collected at baseline, and thus we cannot examine change over time or compare arms while adjusting for baseline differences. Furthermore, not all outputs were assessed: for example, students were not asked about restorative conferencing and staff members were not asked about school rules and student participation in decisions. Furthermore, as with our indicative outcome measures, these findings should not be overinterpreted because of the lack of statistical power and the small number of units randomised,

which resulted in baseline differences in sociodemographic factors that could not be adequately controlled in our necessarily conservative approach to adjusted analysis (see *Data on pilot primary outcome measures* for a fuller discussion).

Analyses of student (*Table 6*) and teacher (*Table 7*) survey data showed few significant differences between groups. There were no significant differences between arms in teachers' or students' reporting of circle time and restorative conferencing. In addition to the methodological factors listed above, this is perhaps not surprising given the lateness with which these practices were implemented (i.e. in the summer term), following delays in training which were the result of our study's later-than-planned initiation date, with consequent slippage in surveys, allocation and therefore scheduling of training in intervention schools.

However, there were significant differences in teacher reports of support around behaviour management. At follow-up, teachers at intervention schools were considerably and significantly more likely to report being well supported with behaviour management by both senior members of staff ($p < 0.001$) and others implementing consistent behaviour management techniques ($p < 0.001$). These findings are subject to the same limitations as those of other quantitative analyses. However, the high significance level suggests they are unlikely to be due to chance.

TABLE 6 Student experiences of restorative practices and PSHE

Measure	Comparison [n (%)]	Intervention [n (%)]	Unadjusted effect estimate [OR (95% CI), p -value]	Adjusted effect estimate [OR (95% CI), p -value]
Do teachers at your school ever use circle time? (yes vs. no/do not know)	84 (15.6)	75 (13.9)	0.94 (0.10 to 9.14), 0.957	0.86 (0.08 to 8.72), 0.899
At this school do students have a say in writing the rules? (yes vs. no/do not know)	248 (45.1)	230 (42.4)	0.84 (0.47 to 1.52), 0.569	0.74 (0.40 to 1.36), 0.333
PSHE helps me feel more confident	345 (64.9)	332 (61.1)	0.90 (0.52 to 1.57), 0.719	0.86 (0.60 to 1.22), 0.393
PSHE helps me understand other people's feelings and problems	418 (78.3)	403 (74.8)	0.87 (0.64 to 1.17), 0.343	0.89 (0.65 to 1.21), 0.451
In PSHE I talk about how I feel	214 (39.9)	195 (36.5)	0.89 (0.54 to 1.48), 0.649	0.88 (0.55 to 1.41), 0.598
In PSHE students can be honest about how they feel	315 (59.0)	316 (58.9)	1.04 (0.52 to 2.09), 0.906	1.07 (0.62 to 1.83), 0.815
In PSHE we talk about how our words and actions affect other people	414 (77.7)	410 (75.8)	0.88 (0.59 to 1.30), 0.511	0.92 (0.63 to 1.36), 0.679
In PSHE we talk about strategies for working with others	396 (74.4)	386 (71.5)	0.89 (0.55 to 1.43), 0.627	0.85 (0.53 to 1.36), 0.488
In PSHE we talk about how to be a good friend	391 (73.5)	389 (71.9)	0.91 (0.69 to 1.21), 0.521	0.91 (0.68 to 1.22), 0.521
In PSHE we talk about where we can go for help if we feel down	398 (74.5)	380 (70.2)	0.83 (0.58 to 1.20), 0.325	0.86 (0.60 to 1.23), 0.406
PSHE is a waste of time for me. It teaches me what I already know	255 (47.8)	279 (51.9)	1.23 (0.85 to 1.78), 0.277	1.24 (0.83 to 1.83), 0.293

OR, odds ratio.

Unless otherwise specified, analyses compare responses 'I totally agree/I agree a bit' with responses 'I do not really agree/no, totally disagree'. Adjusted analyses adjust for sex, ethnicity and socioeconomic status at baseline.

TABLE 7 Teacher experiences of restorative practices, behaviour management and PSHE

Measure	Comparison [n (%)]	Intervention [n (%)]	Unadjusted effect estimate [OR (95% CI), p-value]	Adjusted effect estimate [OR (95% CI), p-value]
Do teachers at your school ever use circle time to discuss how students feel about school? (yes vs. no/do not know)	59 (35.1)	54 (32.9)	0.84 (0.20 to 3.49), 0.811	0.79 (0.16 to 3.83), 0.766
Do staff at this school ever use 'restorative conferences' to deal with disputes and repair relationships? (yes vs. no/do not know)	102 (60.7)	97 (59.2)	0.88 (0.32 to 2.44), 0.807	0.72 (0.25 to 2.11), 0.552
How well are teachers supported with behaviour management at this school by senior members of staff? (quite well/very well vs. not very well/not at all)	7 (4.4)	30 (20.6)	6.17 (2.20 to 17.30), 0.001	6.93 (2.59 to 18.55), <0.001
How well are teachers supported with behaviour management at this school by all staff implementing consistent techniques across the school? (quite well/very well vs. not very well/not at all)	9 (5.6)	60 (36.6)	9.57 (3.80 to 24.12), <0.001	12.70 (4.99 to 32.34), <0.001
Do you think that PSHE lessons at this school help to promote students' social and emotional well-being? (yes vs. no/do not know)	113 (67.3)	88 (53.3)	0.57 (0.31 to 1.04), 0.067	0.59 (0.29 to 1.17), 0.132

OR, odds ratio.

A Phase III trial should more comprehensively examine coverage of all intervention outputs and, when possible, should assess coverage at baseline. Surveys should also assess overall awareness of inputs such as the action group and staff training.

Research design and methods acceptability: key themes

In order to address RQ8, we explored how acceptable the research design and survey data collection methods were through thematic analyses of the qualitative data collected.

Trial recruitment

Head teachers consistently reported that the process of recruiting schools into the pilot trial via initial contact from the intervention manager (a former secondary school deputy head, now an educational consultant and Ofsted inspector), followed by a more detailed discussion of what the research involved with the trial manager, was appropriate. They felt that other methods of recruitment (e.g. via local authorities) were unlikely to be more effective and seemed less appropriate:

Local authorities are one route in, although it's become more disparate because of the academies and free schools.

Head teacher, Goldstone Park

It was, however, suggested that 'networks' of head teachers could also support recruitment into a larger trial via 'word-of-mouth' communication. The head teacher at the Bell Street Academy also highlighted the potential of new partnerships between schools and universities to facilitate 'research readiness', although this may limit generalisability if it were the only source of recruitment for a trial:

Lots of schools I know have very close relationships with, with academic institutions and they do lots of in-school research. I actually think it's an opportunity area.

Head teacher, Bell Street Academy

Although the recruitment process was acceptable to schools, it was 'rushed' during the pilot trial, reflecting the later-than-planned July start. These challenges were compounded by the purposive sampling criteria for the pilot, which required us to recruit very particular combinations of school. The intervention manager, who recruited schools, suggested a larger team of education consultants could recruit schools for a full trial over a longer time period:

You needed a team of consultants who've got the contacts in the areas that you want. You see, I mean, on Ofsted I work with quite a lot of people who do a lot of coaching and other consultancy in schools, and so they, if you had a team of say five people and a year to do it, you'd do it without any problem . . .

Vanessa, intervention manager

Recruitment of schools further in advance and over a longer time period would also provide more time for presentations to staff from the intervention and research teams to get them 'on board'. However, the central message was that it was acceptable and feasible for the intervention team to recruit schools.

Randomisation and experiences of being in a trial

Schools' motivation for participation was usually for them to build on existing work and to 'add value' to their core business. As a result, schools that were assigned to the comparison group were disappointed:

We were disappointed. We were disappointed because we did make an investment although as I said I didn't overdo it in case we didn't get, uh, selected. We were also disappointed because we thought it would give, erm, us a real boost and be something that was of genuine value to us. We wouldn't just be engaging as a favour or as an experiment [. . .] I'm disappointed, our Assistant Principal, was disappointed. Not aggrieved, and certainly not to the extent where we felt we'd been let down or anything like that.

Head teacher, Bell Street Academy

Despite disappointment, the process of informing schools that they would not be in the intervention group and what this meant for them in term of the rest of study was deemed to be acceptable to SMTs at comparison schools. For example:

I was quite happy with the information that I got. And I felt all along that if I had any questions it was made very clear to me that, you know, just ask and we'll tell you what the information is. Everything I asked, I got satisfactory answers. So there was nothing that you did that was wrong in any way. I was quite happy with the information that I got.

Deputy head teacher, Woodhouse School

There was also evidence that there is support for cluster RCTs from the current generation of head teachers who are more 'passionate' about 'evidence', which also made it acceptable for them to be in the comparison group:

I've no problem with [being in the comparison group] because it's the only sound way of getting information which becomes professional, rather than personal [. . .] I'm firmly of an opinion that all schools should be run on evidence-based ideas, rather than fly by your pants kind of approaches! [. . .] I'm passionate about evidence-based work in schools [. . .] we're an all-graduate profession and we should be based on evidence-based work.

Head teacher, Jamestown School

None of the eight schools dropped out of the study and follow-up survey data were collected as planned at each school. One comparison school (Milton Down) did not comply with all elements of process evaluation data collection as per protocol, as they were experiencing a period of transition following the announcement of a merger with another school during the 2011–12 academic year.

The money provided to comparison schools retrospectively at the end of the study to cover any administrative and teaching costs of participating (£500) appeared to be an incentive for Milton Down School to stay in the study and complete the follow-up student survey. The money appeared to be less relevant for motivating the other comparison schools:

Five hundred pounds is nothing. It's a drop in the ocean. You know, what you will get in a school for five hundred pounds is negligible really [. . .] people weren't doing it to get 500 pounds. They were doing it to get the investment in the programme [. . .] We didn't say 'oh, let's do this 'cos if we're not selected we get 500 quid'. We said, 'oh, let's do this because hopefully we'll be . . . in the right half'.

Deputy head teacher, Jamestown School

I think, and then you did the, um, administration of it, so I would suggest that 500 pounds is quite generous considering what it, it didn't . . . What we had to do.

Assistant head, Bell Street Academy

Overall, however, it appeared that it would be worthwhile to retain this in order to encourage all comparison schools to remain in the study at follow-up.

Survey methods

Student participation was exceptionally high in both baseline and follow-up surveys, suggesting high overall acceptability among students. However, some students reported concerns regarding the confidentiality of survey data when completing their questionnaires in school:

You know when we wrote our name and signed it, they could like count through and, count through the papers . . .

Male student, Year 8, Williamson High School

This suggests the need to ensure that information is provided to students in more accessible language about how anonymity is maintained within the trial. Some students thought that paper questionnaires 'waste trees'/'waste paper' and that an online survey would be more 'green'. However, it appears that not all schools would be able to organise students to complete web-based surveys during the same period, which is likely to be a barrier to using them as a replicable data collection method for a Phase III trial.

The most consistent theme among school staff and the intervention team was that we need to change the timing of the baseline surveys:

It's all about that build-up. I mean, bearing in mind that heads have no time, if you were to build that up and start getting heads together and explaining what you would be asking for, or other staff involved before that time, then the seeds would be sown much more about, you know, the input. It's just laying that, and also you know what it's like, it's the beginning of [autumn] term [in September].

Vanessa, intervention manager

Undertaking surveys in the summer would also have the added advantage that comparisons with follow-up surveys would be less likely to be affected by seasonality in reporting.

Teacher surveys were also piloted during this study. Although organising teacher surveys was a challenge, these occurred at baseline and follow-up as planned at seven out of the eight schools. The time constraints on these surveys were a recurring theme, and there was a widespread agreement that whole-school teacher surveys were most feasible when all staff were together at a regular staff meeting or briefing:

From the [co-ordinating] teacher's point of view I would like to have a time when they're all present and all hand it in at the same time.

Deputy head, Woodhouse School

The only other times that all teaching staff get together, outside staff meetings and briefings, are the 'inset' training days during which there is limited time for external, non-essential activities:

We've only had three 'inset' days this year, and the time is really, really kind of fought over.

Deputy head, Woodhouse School

A SMT member at one of the comparison schools (Jamestown School) reported that staff were concerned about the confidentiality of the surveys:

Sometimes some people are suspicious about surveys and data, and think 'what are they looking for? What should I say? What would look and sound right for me and the school, and the children?'

Head teacher, Jamestown School

Finally, an important finding was that all schools would have preferred non-teaching as well as teaching staff to have been surveyed, particularly at schools trying to change perceptions and 'break down barriers':

There was division in the school and there often is in many schools between support staff, non-teaching staff, teaching staff and traditionally some people having the views that non-teaching staff shouldn't be doing certain things [. . .] We're trying to break down all those barriers, we're trying to say we're one team, everybody is supporting improvement in behaviour, whether they're the office staff or not students should respond in the same way to all adults [. . .] of course in our staff meeting [when the survey was undertaken] we have non-teaching staff, and we were only giving out the surveys to the teaching staff. And that just re-enforced what we didn't want to re-enforce.

Head teacher, Goldstone School

Recruitment and survey timings: recommendations for a Phase III trial

This pilot trial was initiated in July, 3 months after originally planned, which seriously impeded our ability to recruit schools, although this was, nonetheless, completed.

To ensure more efficient recruitment in a Phase III, we recommend that:

- the project is initiated in February to enable liaison with schools to proceed 4–6 months before the summer holidays
- the project team partners with existing networks of schools, such as the Institute of Education's 'Teaching Schools', 'Challenge Partners' and UCL Partners
- baseline surveys are conducted in the summer term prior to the school year in which the intervention occurs
- a Phase III trial includes baseline and follow-up surveys with teachers but also with teaching assistants and other school staff.

Development and piloting of indicative primary outcome measures

One of the key purposes of the pilot study was to examine the performance of potential primary outcome measures of aggression and bullying in order to identify the primary outcome(s) for the full trial.

Development of primary outcome measures

We examined the response rates, discrimination and reliability of measures that examine aspects of bullying and aggression at school. Our original study protocol stated that our findings would be used to identify which of the two indicative primary outcomes, the GBS or the AAYP violence scale, performed best in assessing bullying and aggression among year 8 students. In addition, we piloted a third measure of aggression, the ESYTC school misbehaviour subscale. Below, we describe quantitative indicators of each scale's performance in relation to completion rates, prevalence, response discrimination, ICC and internal consistency.

Completion rates and prevalence

Completion rates were calculated for each outcome measure, overall and stratified by sex. Low completion rates can indicate that an item is not generally well understood, or that respondents did not feel comfortable answering it. *Table 8* shows the prevalence of missing data on items needed to calculate the GBS, AAYP and ESYTC school misbehaviour subscale scores overall and by sex.

TABLE 8 Missing data for pilot primary outcomes by sex

Outcome measure	Males [<i>n</i> (%)]	Females [<i>n</i> (%)]	Overall [<i>n</i> (%)]
Baseline			
GBS overall score	45 (8.0)	16 (3.2)	61 (5.8)
AAYP overall score	46 (8.2)	31 (6.4)	77 (7.4)
ESYTC subscale overall score	56 (10.1)	30 (6.2)	86 (8.3)
Follow-up			
GBS overall score	55 (10.3)	31 (6.7)	86 (8.7)
AAYP overall score	38 (6.9)	18 (3.8)	56 (5.5)
ESYTC subscale overall score	62 (11.8)	29 (6.3)	91 (9.2)

Table 8 shows that the number of missing data on items needed to calculate the overall scores of the three pilot primary outcome measures (GBS, AAYP and EYSTC school misbehaviour subscale) ranged from 5.8% to 8.3% at baseline. Non-completion rates for these measures were therefore low at baseline, although non-completion rates were slightly higher for male students (8.0–10.1%) than for females (3.2–6.4%). Missing data were fewest for the GBS total score at baseline: 94.2% of students who completed the baseline questionnaire provided usable GBS score data in full. The non-completion rate remained relatively low for all the pilot indicative primary outcome measures at follow-up (5.5–9.2%), decreasing somewhat for the AAYP measure (see Table 8). Girls were, once again, more likely to respond to all items on these measures than boys: at follow-up the proportion of male students who did not complete all items needed to calculate the EYSTC school misbehaviour subscale was nearly 12%. The non-completion rate for the total GBS score at follow-up was 10.3% for males compared with 6.7% for females.

Table 9 shows the responses for each item, including missing data, according to sex at baseline and follow-up. Overall, items on the GBS were consistently answered, although at follow-up 7.7% ($n = 45$) and 8.2% ($n = 48$) of male students did not complete the third and fourth items, respectively ('Have you been deliberately left out of things at this school in the last 3 months?' and 'Have you been threatened physically or actually hurt by another student recently at this school?'). The percentage of students who reported that they had been a victim of teasing, in the past 3 months, was similar across males and females, and across time (41.8–46.3%). About one-third of respondents stated that other students had spread rumours about them and one-fifth had been deliberately left out of things at school. Approximately 15% of girls at both time points reported having been physically threatened or hurt by another student in the last 3 months; this figure was slightly higher for boys (21.4% at baseline and 18.2% at follow-up).

TABLE 9 Pilot primary outcome response rates by sex at baseline and follow-up

Item	Response	Baseline		Follow-up	
		Male [n (%)]	Female [n (%)]	Male [n (%)]	Female [n (%)]
Bullying victimisation (GBS)					
Has anyone teased you or called you names at this school in the last 3 months?	No	318 (52.3)	276 (53.8)	312 (53.1)	249 (50.6)
	Yes	254 (41.8)	230 (44.8)	247 (42.0)	228 (46.3)
	Missing	36 (5.9)	7 (1.4)	29 (4.9)	15 (3.0)
How often?	Most days	92 (15.1)	67 (13.1)	84 (14.3)	73 (14.8)
	About once a week	67 (11.0)	48 (9.4)	64 (10.9)	58 (11.8)
	Less than once a week	92 (15.1)	109 (21.2)	93 (15.8)	92 (18.7)
	Not applicable	318 (52.3)	276 (53.8)	312 (53.1)	249 (50.6)
	Missing or invalid	39 (6.4)	13 (2.5)	35 (6.0)	20 (4.1)
How upsetting was it when you were teased or called names?	Not at all	95 (15.6)	46 (9.0)	106 (18.0)	65 (13.2)
	A bit	96 (15.8)	92 (17.9)	105 (17.9)	95 (19.3)
	Quite upsetting	61 (10.0)	87 (17.0)	32 (5.4)	66 (13.4)
	Not applicable	318 (52.3)	276 (53.8)	312 (53.1)	249 (50.6)
	Missing or invalid	38 (6.3)	12 (2.3)	33 (5.6)	17 (3.5)

continued

TABLE 9 Pilot primary outcome response rates by sex at baseline and follow-up (continued)

Item	Response	Baseline		Follow-up	
		Male [n (%)]	Female [n (%)]	Male [n (%)]	Female [n (%)]
Has anyone spread rumours about you at this school in the last 3 months?	No	416 (68.4)	351 (68.4)	402 (68.4)	313 (63.6)
	Yes	162 (26.6)	151 (29.4)	148 (25.2)	151 (30.7)
	Missing	30 (4.9)	11 (2.1)	38 (6.5)	28 (5.7)
How often?	Most days	35 (5.8)	31 (6.0)	24 (4.1)	19 (3.9)
	About once a week	42 (6.9)	36 (7.0)	31 (5.3)	35 (7.1)
	Less than once a week	80 (13.2)	82 (16.0)	88 (15.0)	91 (18.5)
	Not applicable	416 (68.4)	351 (68.4)	402 (68.4)	313 (63.6)
	Missing or invalid	35 (5.8)	13 (2.5)	43 (7.3)	34 (6.9)
How upsetting were the rumours?	Not at all	52 (8.6)	34 (6.6)	49 (8.3)	43 (8.7)
	A bit	79 (13.0)	60 (11.7)	59 (10.0)	55 (11.2)
	Quite upsetting	27 (4.4)	56 (10.9)	35 (6.0)	50 (10.2)
	Not applicable	416 (68.4)	351 (68.4)	402 (68.4)	313 (63.6)
	Missing or invalid	34 (5.6)	12 (2.3)	43 (7.3)	31 (6.3)
Have you been deliberately left out of things at this school in the last 3 months?	No	464 (76.3)	386 (75.2)	434 (73.8)	351 (71.3)
	Yes	108 (17.8)	117 (22.8)	109 (18.5)	111 (22.6)
	Missing	36 (5.9)	10 (1.9)	45 (7.7)	30 (6.1)
How often?	Most days	30 (4.9)	28 (5.5)	16 (2.7)	19 (3.9)
	About once a week	33 (5.4)	33 (6.4)	36 (6.1)	28 (5.7)
	Less than once a week	42 (6.9)	54 (10.5)	51 (8.7)	63 (12.8)
	Not applicable	464 (76.3)	386 (75.2)	434 (73.8)	351 (71.3)
	Missing or invalid	39 (6.4)	12 (2.3)	51 (8.7)	31 (6.3)
How upsetting was it being left out of things?	Not at all	32 (5.3)	24 (4.7)	26 (4.4)	20 (4.1)
	A bit	53 (8.7)	49 (9.6)	62 (10.5)	53 (10.8)
	Quite upsetting	21 (3.5)	41 (8.0)	17 (2.9)	37 (7.5)
	Not applicable	464 (76.3)	386 (75.2)	434 (73.8)	351 (71.3)
	Missing or invalid	38 (6.3)	13 (2.5)	49 (8.3)	31 (6.3)

TABLE 9 Pilot primary outcome response rates by sex at baseline and follow-up (continued)

Item	Response	Baseline		Follow-up	
		Male [n (%)]	Female [n (%)]	Male [n (%)]	Female [n (%)]
Have you been threatened physically or actually hurt by another student at this school recently?	No	443 (72.9)	424 (82.7)	433 (73.6)	391 (79.5)
	Yes	130 (21.4)	74 (14.4)	107 (18.2)	71 (14.4)
	Missing	35 (5.8)	15 (2.9)	48 (8.2)	30 (6.1)
How often?	Most days	26 (4.3)	13 (2.5)	24 (4.1)	9 (1.8)
	About once a week	31 (5.1)	10 (1.9)	30 (5.1)	12 (2.4)
	Less than once a week	68 (11.2)	46 (9.0)	52 (8.8)	47 (9.6)
How upsetting was it being threatened or hurt?	Not applicable	443 (72.9)	424 (82.7)	433 (73.6)	391 (79.5)
	Missing or invalid	40 (6.6)	20 (3.9)	49 (8.3)	33 (6.7)
	Not at all	44 (7.2)	14 (2.7)	41 (7.0)	21 (4.3)
	A bit	45 (7.4)	25 (4.9)	33 (5.6)	25 (5.1)
	Quite upsetting	40 (6.6)	33 (6.4)	32 (5.4)	25 (5.1)
	Not applicable	443 (72.9)	424 (82.7)	433 (73.6)	391 (79.5)
	Missing or invalid	36 (5.9)	17 (3.3)	49 (8.3)	30 (6.1)
Aggression – AAYP violence scale					
Have you ever or in the past 3 months ...					
Threatened to beat someone up, not including your brothers and sisters?	Never	441 (72.5)	421 (82.1)	416 (70.7)	403 (81.9)
	Yes, but not in the last 3 months	71 (11.7)	44 (8.6)	76 (12.9)	44 (8.9)
	Once in the last 3 months	35 (5.8)	15 (2.9)	40 (6.8)	19 (3.9)
	More than once in the last 3 months	20 (3.3)	7 (1.4)	25 (4.3)	14 (2.8)
	Missing	41 (6.7)	26 (5.1)	31 (5.3)	12 (2.4)
Threatened to beat up your brother or sister?	Never	424 (69.7)	366 (71.3)	395 (67.2)	333 (67.7)
	Yes, but not in the last 3 months	67 (11.0)	44 (8.6)	58 (9.9)	53 (10.8)
	Once in the last 3 months	35 (5.8)	33 (6.4)	41 (7.0)	34 (6.9)
	More than once in the last 3 months	43 (7.1)	39 (7.6)	60 (10.2)	57 (11.6)
	Missing	39 (6.4)	31 (6.0)	34 (5.8)	15 (3.0)

continued

TABLE 9 Pilot primary outcome response rates by sex at baseline and follow-up (continued)

Item	Response	Baseline		Follow-up	
		Male [n (%)]	Female [n (%)]	Male [n (%)]	Female [n (%)]
Threatened to cut, stab or shoot someone?	Never	546 (89.8)	478 (93.2)	536 (91.2)	466 (94.7)
	Yes, but not in the last 3 months	10 (1.6)	4 (0.8)	10 (1.7)	4 (0.8)
	Once in the last 3 months	6 (1.0)	1 (0.2)	4 (0.7)	5 (1.0)
	More than once in the last 3 months	3 (0.5)	2 (0.4)	5 (0.9)	3 (0.6)
	Missing	43 (7.1)	28 (5.5)	33 (5.6)	14 (2.8)
Cut or stabbed someone?	Never	555 (91.3)	479 (93.4)	537 (91.3)	472 (95.9)
	Yes, but not in the last 3 months	7 (1.2)	4 (0.8)	11 (1.9)	2 (0.4)
	Once in the last 3 months	1 (0.2)	1 (0.2)	1 (0.2)	2 (0.4)
	More than once in the last 3 months	3 (0.5)	1 (0.2)	4 (0.7)	1 (0.2)
	Missing	42 (6.9)	28 (5.5)	35 (6.0)	15 (3.0)
ESYTC school misbehaviour subscale					
During the last 3 months how often did you do these things at school?					
Arrive late for classes	Hardly ever or never	330 (54.3)	311 (60.6)	313 (53.2)	265 (53.9)
	Less than once a week	123 (20.2)	87 (17.0)	108 (18.4)	83 (16.9)
	At least once a week	69 (11.3)	72 (14.0)	95 (16.2)	92 (18.7)
	Most days	56 (9.2)	27 (5.3)	48 (8.2)	44 (8.9)
	Missing	30 (4.9)	16 (3.1)	24 (4.1)	8 (1.6)
Fight in or outside the classroom	Hardly ever or never	470 (77.3)	444 (86.6)	457 (77.7)	426 (86.6)
	Less than once a week	65 (10.7)	32 (6.2)	71 (12.1)	30 (6.1)
	At least once a week	19 (3.1)	12 (2.3)	24 (4.1)	19 (3.9)
	Most days	19 (3.1)	9 (1.8)	9 (1.5)	8 (1.6)
	Missing	35 (5.8)	16 (3.1)	27 (4.6)	9 (1.8)
Refuse to do homework or class work	Hardly ever or never	470 (77.3)	429 (83.6)	415 (70.6)	382 (77.6)
	Less than once a week	58 (9.5)	39 (7.6)	87 (14.8)	59 (12.0)
	At least once a week	28 (4.6)	16 (3.1)	31 (5.3)	21 (4.3)
	Most days	17 (2.8)	13 (2.5)	21 (3.6)	20 (4.1)
	Missing	35 (5.8)	16 (3.1)	34 (5.8)	10 (2.0)

TABLE 9 Pilot primary outcome response rates by sex at baseline and follow-up (continued)

Item	Response	Baseline		Follow-up	
		Male [n (%)]	Female [n (%)]	Male [n (%)]	Female [n (%)]
Be cheeky to a teacher	Hardly ever or never	330 (54.3)	345 (67.3)	295 (50.2)	290 (58.9)
	Less than once a week	142 (23.4)	82 (16.0)	143 (24.3)	109 (22.2)
	At least once a week	63 (10.4)	38 (7.4)	84 (14.3)	48 (9.8)
	Most days	32 (5.3)	30 (5.8)	41 (7.0)	36 (7.3)
	Missing	41 (6.7)	18 (3.5)	25 (4.3)	9 (1.8)
Use bad or offensive language	Hardly ever or never	406 (66.8)	390 (76.0)	350 (59.5)	341 (69.3)
	Less than once a week	75 (12.3)	44 (8.6)	110 (18.7)	66 (13.4)
	At least once a week	51 (8.4)	30 (5.8)	48 (8.2)	36 (7.3)
	Most days	40 (6.6)	30 (5.8)	54 (9.2)	40 (8.1)
	Missing	36 (5.9)	19 (3.7)	26 (4.4)	9 (1.8)
Wander around school during class time	Hardly ever or never	478 (78.6)	415 (80.9)	441 (75.0)	391 (79.5)
	Less than once a week	48 (7.9)	50 (9.7)	69 (11.7)	56 (11.4)
	At least once a week	29 (4.8)	15 (2.9)	31 (5.3)	18 (3.7)
	Most days	16 (2.6)	16 (3.1)	22 (3.7)	20 (4.1)
	Missing	37 (6.1)	17 (3.3)	25 (4.3)	7 (1.4)
Threaten a teacher	Hardly ever or never	557 (91.6)	488 (95.1)	541 (92.0)	465 (94.5)
	Less than once a week	9 (1.5)	1 (0.2)	12 (2.0)	10 (2.0)
	At least once a week	4 (0.7)	0 (0.0)	7 (1.2)	7 (1.4)
	Most days	5 (0.8)	8 (1.6)	3 (0.5)	3 (0.6)
	Missing	33 (5.4)	16 (3.1)	25 (4.3)	7 (1.4)
Hit/kick a teacher	Hardly ever or never	558 (91.8)	486 (94.7)	547 (93.0)	473 (96.1)
	Less than once a week	8 (1.3)	3 (0.6)	8 (1.4)	5 (1.0)
	At least once a week	2 (0.3)	1 (0.2)	7 (1.2)	4 (0.8)
	Most days	4 (0.7)	6 (1.2)	2 (0.3)	2 (0.4)
	Missing	36 (5.9)	17 (3.3)	24 (4.1)	8 (1.6)

continued

TABLE 9 Pilot primary outcome response rates by sex at baseline and follow-up (continued)

Item	Response	Baseline		Follow-up	
		Male [n (%)]	Female [n (%)]	Male [n (%)]	Female [n (%)]
Cheat while doing homework or tests	Hardly ever or never	493 (81.1)	424 (82.7)	476 (81.0)	399 (81.1)
	Less than once a week	60 (9.9)	50 (9.7)	64 (10.9)	66 (13.4)
	At least once a week	14 (2.3)	6 (1.2)	14 (2.4)	9 (1.8)
	Most days	8 (1.3)	16 (3.1)	10 (1.7)	9 (1.8)
	Missing	33 (5.4)	17 (3.3)	24 (4.1)	9 (1.8)
Purposely damage or destroy things belonging to the school	Hardly ever or never	524 (86.2)	464 (90.4)	504 (85.7)	449 (91.3)
	Less than once a week	35 (5.8)	17 (3.3)	47 (8.0)	18 (3.7)
	At least once a week	8 (1.3)	7 (1.4)	9 (1.5)	9 (1.8)
	Most days	7 (1.2)	8 (1.6)	6 (1.0)	5 (1.0)
	Missing	34 (5.6)	17 (3.3)	22 (3.7)	11 (2.2)
Threaten another student	Hardly ever or never	N/A	N/A	463 (78.7)	439 (89.2)
	Less than once a week	N/A	N/A	71 (12.1)	23 (4.7)
	At least once a week	N/A	N/A	15 (2.6)	12 (2.4)
	Most days	N/A	N/A	13 (2.2)	8 (1.6)
	Missing	N/A	N/A	26 (4.4)	10 (2.0)
Hit/kick another student	Hardly ever or never	N/A	N/A	399 (67.9)	389 (79.1)
	Less than once a week	N/A	N/A	117 (19.9)	58 (11.8)
	At least once a week	N/A	N/A	32 (5.4)	18 (3.7)
	Most days	N/A	N/A	17 (2.9)	15 (3.0)
	Missing	N/A	N/A	23 (3.9)	12 (2.4)
Get in a fight	Hardly ever or never	N/A	N/A	445 (75.7)	424 (86.2)
	Less than once a week	N/A	N/A	95 (16.2)	36 (7.3)
	At least once a week	N/A	N/A	13 (2.2)	12 (2.4)
	Most days	N/A	N/A	13 (2.2)	6 (1.2)
	Missing	N/A	N/A	22 (3.7)	14 (2.8)

N/A, not applicable.

The highest rate of missing data for the AAYP measure was for the third item ('Have you ever threatened to cut, stab or shoot anyone?'), which 43 males did not answer (7.1%) at baseline. Analysis of students' responses to items on the AAYP measure found that, overall, students were more likely to report aggressive behaviour at follow-up (see *Table 9*). For example, at baseline, 9.1% of males and 4.3% of females reported that they had threatened to beat someone up at least once in the last month, and this figure was slightly higher at follow-up (11.1% and 6.7%, respectively). However, the rates of 'threatening to cut, stab or shoot someone' and having 'cut or stabbed someone' in the last 3 months were almost zero for both males and females at both time points (range 0.4–1.6%). Exploratory analyses of the ESYTC school misbehaviour subscale found the non-completion rate was low for all items (< 7%) at both baseline and follow-up. The lowest rates of missing data across all three measures were for the sixth and seventh items of the ESYTC school misbehaviour subscale, with only seven females (1.4%) not answering at follow-up. ESYTC items relating to aggression provide stable but relatively low estimates of aggressive behaviour over time in this population.

Discrimination

Table 10 provides the mean scores and standard deviations for all the pilot primary outcome measures at baseline. The lowest possible value for all three scales is within one standard deviation of the mean, suggesting all scales may suffer from 'floor effects' (i.e. the scales may not be sensitive to low levels of reported bullying). This is particularly true for the AAYP score and the ESYTC school misbehaviour subscale score, both of which had standard deviations that were substantially higher than their means.

The distribution of responses was explored, and this also suggested the possibility of floor effects: responses were not normally distributed; and, for the AAYP measure, the vast majority of respondents in this study have a score of zero. For GBS and ESYTC, zero is also the most common value, but the distribution of responses is more even than that of the AAYP measure. In regards to discrimination, GBS performs best, followed by ESYTC, then AAYP.

Reliability

Reliability was assessed by examining the ICC for each score (as an indication of the stability of the measure over time) and by examining Cronbach's alpha statistics at baseline and follow-up (indicating each scale's internal consistency).

TABLE 10 Pilot primary outcome overall scores at baseline

Outcome measure	Mean (SD)
GBS overall score	0.98 (1.01)
Teasing subscale	0.73 (0.95)
Rumours subscale	0.42 (0.76)
Deliberate exclusion subscale	0.32 (0.70)
Threatened or hurt subscale	0.29 (0.67)
AAYP overall score	0.81 (1.49)
ESYTC subscale score	2.83 (4.37)

SD, standard deviation.

Intraclass correlations give an indication of the proportion of the variance explained within the participant responses (across time) relative to between-subject variance. They are used as an indication of the stability in a measure over time. None of the variables performed particularly well in this area, with surprisingly high ICCs for all three (*Table 11*). However, GBS had a substantially higher ICC than AAYP and ESYTC, suggesting that it is more stable across time.

We calculated Cronbach's alpha statistics to provide an indication of each scale's internal consistency, a measure of the extent to which items within the scale are measuring the same latent construct. Alpha values of 0.6–0.7 are considered acceptable and values of 0.7–0.9 ideal.⁹⁴ In this regard, at both reporting periods, GBS and ESYTC performed well, but the AAYP did not (see *Table 11*). This may be because of items on the AAYP scale that do not fit the typical British conceptualisation of youth violence, and these may have low correlations with other items. The results suggest that, as for the consistency of the items, GBS and ESYTC appear substantially preferable to AAYP. However, in considering these internal consistency coefficients (Cronbach's alpha statistics), we should be cautious. First, a number of items had very low prevalence rates, and scales containing a large proportion of such items will have a lower level of internal consistency. Alpha coefficients are also dependent on the number of items on the subscale, with shorter scales yielding lower coefficients.

Primary outcome measures: recommendations for a Phase III trial

A Phase III trial of INCLUSIVE should include one primary outcome focused on bullying victimisation and one measure focused on aggressive behaviours. Our GBS and ESYTC subscale measure these two outcomes, respectively, and performed satisfactorily for both. The AAYP measure is not appropriate as a primary outcome for a Phase III trial in English secondary schools.

Data on pilot primary outcome measures

Here we present data on summary scores for the three measures at baseline and follow-up.

It is important to note that the pilot study was not designed to identify differences between arms in any quantitative outcomes. First, the study was not powered to detect significant differences, and thus all estimates have very wide confidence intervals and the point estimates described below are not meaningful. Second, as might be expected in a pilot trial that allocates only eight units, there were very marked differences at baseline between the study arms in terms of students' deprivation, family structure and behaviour problems: the intervention schools being consistently more disadvantaged. Our pre-hypothesised approach to adjustment for confounding was necessarily conservative given our small sample size so residual confounding was inevitable. Third, this was a 1-year pilot (as opposed to an intended intervention period of 3 years in a Phase III trial) with a later-than-anticipated start date, so the intervention dose was insufficient to detect effects.

TABLE 11 Reliability statistics for pilot primary outcomes

Scale	ICC	Cronbach's alpha	
		Baseline	Follow-up
GBS	0.522	0.790	0.773
AAYP	0.419	0.474	0.536
ESYTC subscale	0.426	0.847	0.882

Pilot indicative primary outcomes

Table 9 shows that the prevalence of being bullied and aggressive behaviour towards others was relatively high, with relatively few differences according to sex. At baseline, > 40% of students reported being teased in the last 3 months, and 26.4% of males and 22.5% of females reported being teased once a week or more frequently. Moreover, at baseline, 26.6% of males and 29.4% of females had had rumours spread about them in the last 3 months, and 17.8% of males and 22.8% of females reported being deliberately excluded by others in the last 3 months. The rate of reporting being threatened or physically hurt in the last 3 months at baseline was 21.4% for males and 14.4% for females.

The AAYP scale found that, at baseline, 9.1% of year 8 males reported having been aggressive towards others (not including their siblings) during the past 3 months. This figure was lower for year 8 females (4.3%). Very few year 8 students reported that they had threatened to cut, stab or shoot someone in the last 3 months (< 2%) or had actually carried out these aggressive acts (< 1%). Response rates were higher for low-level provoking and school misbehaviour as measured via the ESYTC, such as using bad language, which was reported by > 15% of male students and 11.6% of female students at baseline. The majority of students reported that they had not threatened, hit or kicked a school teacher (see Table 9).

Table 12 provides the mean scores and standard deviations for all the pilot primary outcome measures at baseline for the intervention and comparison groups. There was a consistent trend across all three pilot primary outcome measures for the intervention group to report more prior experience of bullying and higher levels of aggression and school misbehaviour at baseline (see Table 12). On average, mean GBS overall scores varied from 0.91 for students in the comparison group to 1.04 for students in the intervention. With regards to aggression and school misbehaviour, the mean AAYP overall score and the mean ESYTC overall score were also both greater for intervention students. These imbalances are not unusual in a pilot cluster trial that includes a small number of randomised units ($n = 8$).

Table 13 reports the primary outcome score by arm for male and female students at baseline. This shows that rates of prior experience of bullying and levels of aggression and school misbehaviour were similar for males in both arms at baseline: the mean GBS overall score for males in the comparison group was 0.92 and for males in the intervention group was 0.91; for the AAYP the mean was 0.87 in the comparison group and 0.92 in the intervention group; and for the ESYTC school misbehaviour subscale the mean was 3.10 in the comparison group and 3.11 in the intervention group.

TABLE 12 Pilot primary outcomes [mean (SD)] at baseline by arm and overall

Outcome measure	Comparison	Intervention	Overall
GBS overall score	0.91 (0.96)	1.04 (1.05)	0.98 (1.01)
AAYP overall score	0.70 (1.34)	0.92 (1.61)	0.81 (1.49)
ESYTC overall score	2.72 (4.26)	2.94 (4.47)	2.83 (4.37)

SD, standard deviation.

TABLE 13 Pilot primary outcomes [mean (SD)] at baseline by arm stratified by sex

Measure	Male		Female	
	Comparison	Intervention	Comparison	Intervention
GBS overall score	0.92 (0.97)	0.91 (0.99)	0.91 (0.96)	1.17 (1.09)
AAYP overall score	0.87 (1.51)	0.92 (1.67)	0.48 (1.07)	0.92 (1.54)
ESYTC subscale score	3.10 (4.25)	3.11 (4.39)	2.29 (4.27)	2.77 (4.60)

SD, standard deviation.

Although male students were fairly evenly matched in terms of these pilot primary outcomes, there were clear differences between females in the comparison and intervention groups in terms of all three measures. Mean rates of victimisation in the past 3 months measured by the GBS were much higher in the intervention group (1.17) than in the comparison group (0.91). Similarly, the mean overall AAYP and ESYTC school misbehaviour scores were higher for females in the intervention group (see *Table 13*).

The study did not aim to detect intervention effects, and lacked both the statistical power and intervention duration to be able to do so. *Table 14* presents the pilot primary outcomes for the intervention and comparison conditions, and overall, at follow-up. AAYP and ESYTC scores increased between baseline and follow-up in both trial arms, as is expected normatively in year 8 students.

Pilot secondary outcomes

Quantitative analyses of secondary outcomes are shown in *Appendix 2*. As the aim of the pilot study was to pilot data collection methods, and the study was not powered to examine differences between arms, we do not report tests of between-arm differences. Each of the secondary outcomes included in the survey had adequate response rates and is likely to be included in the full trial (see *Appendix 2*).

Pilot intermediate outcome measures

In a Phase III trial we would examine intermediate outcomes concerned with students' perception of the school environment to assess the intervention logic model (see *Figure 1*) via which the intervention is hypothesised to impact on students' health and well-being. This pilot study examined these intermediate outcomes in terms of the feasibility of data collection and analysis. Data are shown in *Appendix 3*.

Economic evaluation

The aim of the economic component of this pilot study was to collect and collate evidence regarding the appropriate design of an economic evaluation in a Phase III trial. To do this, the analysis was divided into two tasks. The first task was to assess the feasibility and desirability of using the recently developed CHU-9D to measure changes in health. The results relating to the validation of the CHU-9D are presented below. The second task was to define an appropriate economic evaluation framework based on these results and consideration of the wider literature, which is discussed in *Chapter 5*. Other pilot trial data and response rates also informed our decision about an economic evaluation framework, including the pilot teacher surveys. *Table 23* reports the teacher response rates at baseline and follow-up, which also informed plans for future economic evaluation within a Phase III trial (see *Appendix 4*).

Validation of the Child Health Utility 9D questionnaire

Response rates and distribution of results

Missing data from the CHU-9D are problematic, as the utility algorithm does not currently recommend they are replaced with suitable estimates. Of 1144 questionnaires issued at baseline and returned,

TABLE 14 Pilot primary outcomes [mean (SD)] at follow-up by arm and overall

Outcome measure	Comparison	Intervention	Overall
GBS overall score	0.89 (0.94)	1.02 (0.96)	0.96 (0.95)
AAYP overall score	0.88 (1.62)	1.09 (1.72)	0.98 (1.67)
ESYTC overall score	3.52 (5.27)	4.32 (5.65)	3.92 (5.47)

SD, standard deviation.

252 (22%) contained at least one missing CHU-9D response, meaning that overall utility scores could be calculated only for 892 respondents. However, at baseline it was the very final item on the questionnaire, and the proportion of incomplete questionnaires decreased to 16% in the follow-up survey, in which it was moved forward to precede other pilot secondary outcome measures. There did not appear to be any discernible indicators as to which of the nine questions was more likely to be left incomplete. This further suggests that time and questionnaire length were the only barriers to completion, and these can be addressed.

Analysis of the baseline data indicated that > 64% of year 8 students indicated 'no problems' on the worry, sad, pain, annoyed, work, sleep and routine CHU-9D subscales (*Table 15*), indicating potential ceiling effects on these domains. This is unsurprising because the young people in the pilot were generally considered to represent a 'healthy' school-going population. However, only 31% of respondents who completed the question indicated that they were not at all tired (31%), and almost 10% indicated that they were 'extremely' tired. The potential for floor effects on each of the domains was not shown to be an issue, as few individuals reported being in any of the 'worst' health states. Only 13% (112/892) of participants indicated that they currently had 'no problems' on all nine domains at baseline, and only 66 individuals (7%) reported having no problems on all domains at both baseline and follow-up.

A histogram of the baseline survey CHU-9D utility score showed that the distribution of results was skewed (*Figure 6*). Whereas the mean baseline CHU9 utility score was 0.88 (95% CI 0.88 to 0.89), the median was 0.90 (interquartile range 0.83 to 0.95). Females reported a lower median utility score than males ($p = 0.0001$). The lowest utility score recorded by a single individual was 0.33 (recorded in the follow-up study), which is the lowest possible health state. These data do not therefore suggest potential 'floor' or 'ceiling' effects.

Construct validity

The results of tests for discriminant validity indicated a strong positive correlation (0.57; $p < 0.0001$) between the CHU-9D utility and PedsQL total score, suggesting that the two scales are measuring similar items (*Table 16*). The correlations between CHU-9D utility and PedsQL total and primary trial outcome measures were all negative and in the expected direction: as CHU-9D utility/PedsQL scores increased (indicating better health), the outcome measure scores decreased. However, although the correlation coefficients were statistically significant ($p < 0.0001$ in all instances), their sizes can be considered moderate at best.

TABLE 15 Baseline CHU-9D responses from all participants by domain

Domain	Response [<i>n</i> (%)]					Total (<i>n</i>)
	1	2	3	4	5	
Worried	699 (75)	163 (17)	45 (5)	22 (2)	6 (0)	935
Sad	726 (78)	138 (15)	33 (4)	25 (3)	13 (1)	935
Pain	630 (68)	174 (19)	70 (8)	30 (3)	17 (2)	921
Tired	284 (31)	348 (38)	127 (14)	82 (9)	79 (9)	920
Annoyed	678 (74)	144 (16)	50 (5)	29 (3)	21 (2)	922
School work	665 (73)	167 (18)	61 (6)	13 (1)	12 (1)	918
Sleep	590 (64)	204 (22)	70 (8)	30 (3)	28 (3)	922
Routine	793 (87)	94 (10)	20 (2)	4 (0)	4 (0)	915
Activities	654 (72)	122 (13)	51 (6)	41 (5)	43 (5)	911

Responses range from 1, indicating 'no problems', to 5, indicating 'lots' of problems.

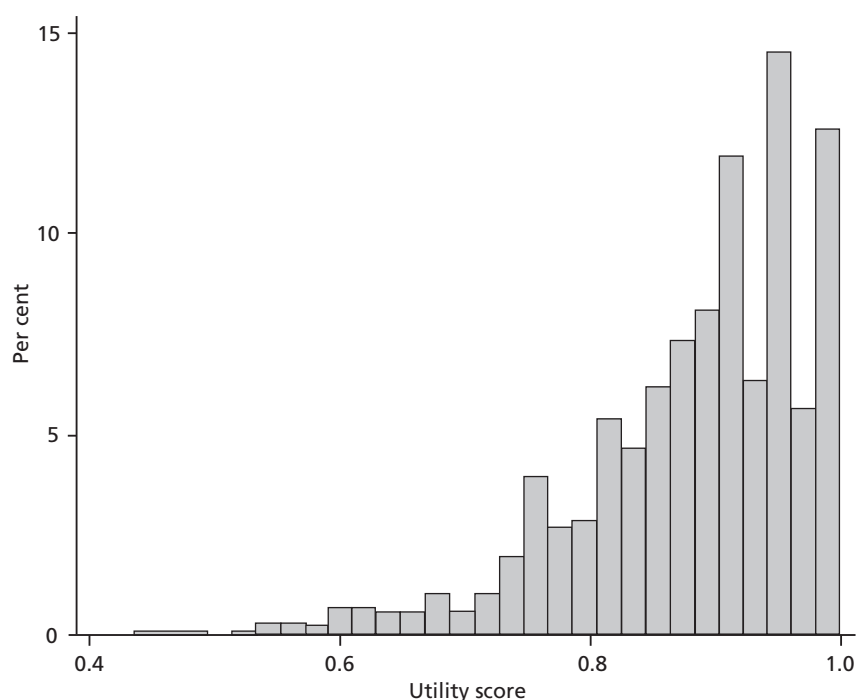


FIGURE 6 Distribution of CHU-9D utility scores.

TABLE 16 Correlations (p -value) between the CHU-9D utility score and other outcome measures

Outcome measure	CHU9 utility	PedsQL total
GBS total	-0.31 (<0.0001)	-0.35 (<0.0001)
AAYP total	-0.14 (<0.0001)	-0.16 (<0.0001)
ESYTC school misbehaviour subscale total	-0.25 (<0.0001)	-0.28 (<0.0001)
PedsQoL total	0.57 (<0.0001)	N/A
N/A, not applicable.		

In tests of convergent validity, most of the hypothesised relationships were statistically significant. However, the level of convergent validity as assessed using the PedsQL subscale scores was arguably only 'moderate': the highest was -0.38 (Table 17).

Correlating changes in the CHU-9D utility scores and changes in the three primary outcome measures showed that the relationships were statistically significant ($p \leq 0.02$ in all instances) (Table 18). However, again, the degree to which the scores were related was subjectively weak; the largest correlation was -0.11 . The changes in the PedsQL and CHU-9D utility score were much more closely related. Although ideally there would be a level of correlation between all outcome measures, by definition they have been designed to measure different outcomes. For example, the AAYP asks questions regarding the frequency of aggressive behaviour rather than questions about 'health' per se. Moreover, the period of recall is over the last 3 months, whereas the CHU-9D specifically refers to 'today'. Similarly, the GBS asks specific questions regarding experiences associated with bullying in terms of frequency and the degree of upset they caused, also over the last 3 months.

TABLE 17 Assessing convergent validity (*p*-value) (statistical tests are two tailed)

CHU-9D subscale	PedsPsych	PedsPhysical	SDQ	W-EMWS
Worried	-0.38 (<0.0001)	N/A	0.32 (<0.0001)	-0.19 (<0.0001)
Sad	-0.35 (<0.0001)	N/A	0.32 (<0.0001)	-0.23 (<0.0001)
School work	-0.33 (<0.0001)	N/A	N/A	N/A
Activities	N/A	-0.33 (<0.0001)	n/a	N/A

CHU-9D subscale, Child Health Utility 9D measure; N/A, not applicable; PedsPhysical, PedsQL physical health summary score; PedsPsych, PedsQL, psychosocial summary score; SDQ, Strengths and Difficulties Questionnaire; W-EMWS, Short Warwick–Edinburgh Mental Well-Being Scale.

TABLE 18 Sensitivity of the CHU-9D utility scores to changes in other outcome measures

Outcome measure	Change in CHU9 utility (<i>p</i> -value)
Change in GBS total	-0.10 (0.0007)
Change in AAYP total	-0.11 (0.005)
Change in ESYTC total	-0.09 (0.02)
Change in PedsQL total	0.34 (<0.0001)

Chapter 5 Discussion

Key findings

The INCLUSIVE intervention was feasible to implement and acceptable to school staff and students at four English secondary schools with varying levels of deprivation and a range of Ofsted ratings of school effectiveness.

The first objective of the pilot trial was to assess whether the prespecified feasibility and acceptability criteria were met. All these criteria were met in full. At all four intervention schools, the needs assessment survey had an > 80% response rate (range 91–97%); an action group was formed, met at least six times during the year and revised school policies relating to aggression and its determinants, including through whole-school actions involving staff and students from across the school; > 20 staff completed restorative practice training (range = 30–75), peer mediation and/or 'buddying' schemes were reviewed and enhanced in line with restorative principles, and restorative practices were introduced (e.g. circle time, checking in/out, etc.); and the student curriculum was delivered to all year 8 students (range 7–12 hours). The intervention was acceptable to school SMT members and student and staff action group members. The intervention was also highly rated by both student and staff action group members who were surveyed (e.g. > 93% of action group members surveyed felt it was a good way to ensure students contributed to decision-making at their school). Furthermore, the trial methods were acceptable to all SMT members, randomisation occurred as planned and no schools dropped out of the study, with follow-up student survey responses rates of 91–94% in the intervention schools and 87–96% in comparison schools.

The second objective was to explore students', school staff members' and facilitators' experiences of implementing and trialling the INCLUSIVE intervention in order to optimise the intervention and refine evaluation methods prior to a subsequent Phase III trial. Qualitative data indicated that all intervention components were acceptable. Students consistently reported that the intervention inputs (i.e. the needs assessment survey, action group, external facilitator, staff training and curriculum components) and restorative practices were appropriate and acceptable to them. The intervention's emphasis on 'having a say' and 'having respect' was a strong source of acceptability among students. Changes enacted included new school rules that students felt were fair (and so were less likely to break) and new timetables giving students more time with their form tutors (so problems with other students could be resolved more quickly). Qualitative data also suggested student participation may be a core component in improving relationships and engagement across the school.

Among school staff, the intervention tapped into a strong desire for locally adaptable processes for improving staff–student relationships and addressing aggressive behaviour. SMT members, particularly head teachers, reported that the intervention was a good 'fit' with schools' 'core business' of improving attainment and the national policy agenda focused on improving student behaviour. Although staff reported that some 'restorative-type' approaches were already being used in their school, this intervention was highly attractive because it could provide a new framework, process and resources for embedding restorative practices more consistently and more widely across the whole school. The extent to which the intervention balanced fidelity of standard processes and key components with scope for local participation and adaptation of some elements was seen as a strength. This intervention could therefore work across different school contexts, and our evaluation suggested that feasibility and acceptability were particularly strong in more disadvantaged and challenging schools.

School staff and intervention providers consistently reported that a longer lead-in was needed, as well as improved co-ordination of activities and a longer duration of intervention (our intention always being that the intervention would be 3 years in a subsequent Phase III trial). School leaders noted that outputs did not

begin until the second term and restorative conferencing did not begin to be implemented until the summer term because of delays in staff training. However, our qualitative data consistently reported that, rather than reflecting any intrinsic limitations of the intervention, these delays were caused by the delayed start date and the timing of the baseline surveys in this study. There was a consensus among staff and facilitators that major impacts on behaviour were likely, but only in the subsequent school year, once policies and practices had been changed, and that a 2-year intervention was the minimum needed to expect change in the school environment. Nonetheless, teachers at intervention schools were significantly more likely to report at follow-up being well supported with behaviour management by both senior members and by others implementing consistent behaviour management techniques. Staff reported that training could be further improved through a more comprehensive pre-training audit to identify schools' needs, greater use of interactive training methods and more 'realistic' examples from similar secondary schools.

The third objective was to pilot and field test potential primary, secondary and intermediate outcome measures and economic methods prior to a Phase III trial. This involved examining measures' completion, discrimination and reliability. The GBS and the ESYTC school misbehaviour subscale were acceptable, discriminating and reliable measures of bullying and aggression. We found significant limitations with using the AAYP questions to assess youth violence in the UK context owing to limited discrimination, and the potential for 'floor effects', and limited reliability. Our pilot economic analyses support the use of the CHU-9D scale with this population and the feasibility of CUA, although this should be supplemented with a cost-consequence analysis (CCA), which is discussed further below. The study did not aim to detect intervention effects, and lacked both the statistical power and intervention duration to be able to do so. There was no evidence of harm.

We found that the collection of data from students and teachers was feasible and allowed us to plan a health economic evaluation for a Phase III trial. The CHU-9D was an acceptable measure of health utility in this sample, although piloting work should be undertaken to ensure all students have time to complete the measure in full. There was little evidence of a potential floor effect and there was a high degree of correlation between the CHU-9D utility and PedsQL total scores, the latter being a well-established health-related QoL instrument for children and young people, which is in line with findings reported elsewhere.⁹⁵ The analyses clearly show that the correlations between the CHU-9D utility and indicative primary outcome measures were all relatively weak. However, the extent to which this finding is problematic or unexpected is debateable. For example, the AAYP items ask questions regarding the frequency of serious aggressive behaviour rather than questions about 'health' per se. Moreover, the period of recall for primary indicative outcomes was 3 months, whereas the CHU-9D specifically refers to how students feel 'today'.

Strengths

We undertook a mixed-methods approach to assess the feasibility and acceptability of the INCLUSIVE intervention and trial methods according to prespecified criteria in a highly challenging group of schools. This study is one of the most comprehensive school-based pilot trials undertaken to date in the UK in terms of the breadth and depth of data collection. It is particularly critical to understand the process of delivering such locally determined, highly complex, multicomponent school-based interventions via a mix of data collection methods. This study provides a template for future Phase II exploratory trials and process evaluations. Another major strength was the use of prespecified feasibility and acceptability criteria, which were all met in full. Young people were also heavily involved in the development of the trial methods and with implementing the intervention.

We chose to pilot this intervention in range of purposively selected secondary schools of varying levels of deprivation and school quality in London and south-east England. Two of our schools (one intervention, one comparison) were extremely challenging contexts (i.e. rates of FSM eligibility above the London

average and rated only as 'satisfactory' by Ofsted in terms of the school management). The process of piloting the intervention and trial methods in such a diverse range of school contexts supports the idea that the intervention is feasible and acceptable across English secondary schools, and provides a very strong basis to proceed to a Phase III trial.

Limitations

This pilot was focused on assessing processes in order to assess and refine the feasibility of the INCLUSIVE intervention and trial methods. Our pilot involved randomisation and outcome evaluation solely to assess their feasibility and acceptability. The inability of this pilot to estimate intervention effects should not, therefore, be considered a limitation. It is customary to say that a pilot trial such as this is 'underpowered'. It would be more accurate to say that power is not an issue in this trial, as the trial was neither designed nor able to examine the effects of the intervention.

Although the GBS and the ESYTC school misbehaviour subscale were found to be acceptable, discriminating and reliable measures of bullying and aggression in this context, this is not to say that they are not without limitations. For example, the GBS scoring scale used here does not distinguish among more extreme, intense forms of bullying (as the most intense form of one type of bullying would score the same as the most intense form of all four kinds of bullying). The scoring scale should therefore be considered further prior to protocol development in a Phase III trial. We also acknowledge that there are some limitations with the measures of psychological well-being piloted and field tested in this study. For example, the SDQ may be more useful as a *behavioural* measure of psychological functioning than as a measure of psychological distress per se; and the SWEMWBS is a relatively new measure of emotional well-being, although this study provided a useful opportunity to pilot the SWEMWBS with young people aged 12 or 13 years in English secondary schools. It should also be noted that, although an appropriate measure of alcohol consumption with students aged 12 or 13 years,⁶⁹ this binary measure of whether students have 'drank more than a sip' of alcohol in the past month (30 days) may not discriminate between levels of alcohol consumption among an older population. If students are followed up until age 15 years or beyond in a Phase III trial, then the outcome measure should assess frequency of heavy alcohol consumption.

Overall, the data collected were comprehensive. However, one gap was identified at the end of the study: although a survey of all action group members was undertaken to examine their views on implementation and acceptability, we did not collect systematic data regarding student action group members' social background and educational attainment. This should be examined in the action group survey in a subsequent Phase III study to support scalability.

Finally, a further limitation is that, although young people were consulted throughout and the voices of students and staff captured through our qualitative data, other key stakeholders were not involved. In future, parental engagement is needed to access parents' views and inform the intervention and research, and representatives from the Department of Education, Ofsted and synergistic public health strategies (e.g. social and emotional aspects of learning) should also be consulted and involved throughout.

Comparison with the literature

Our pilot trial supports the conclusions of an earlier UK pilot study concerning the feasibility and acceptability of an intervention employing an educational facilitator, needs assessment and action groups to bring about health-promoting modifications to the school social environment.^{69,96} Our study builds on this study by suggesting that such an intervention can feasibly be orientated towards aggressive behaviours and bullying; can also incorporate a curriculum component; and can feasibly be subject to a RCT in English schools. Together, these two studies strongly suggest that this type of school environment

intervention, which has been reported to bring about significant benefits in the USA⁵⁴ and Australia,^{56,97} is potentially transferable to English schools.

Supporting the evidence from AAYP,⁵⁴ our evaluation has suggested that this type of school environment intervention might prove especially feasible and acceptable in more challenging schools serving more disadvantaged communities. Therefore, the INCLUSIVE intervention might have the potential not only to bring about overall population health benefits but also to reduce health inequalities. This should be examined within a Phase III trial encompassing a diversity of schools. Our pilot also confirms the findings of previous studies that suggest that restorative practices, such as circle time and restorative conferencing, are feasible and acceptable to deliver in English schools.^{60,62}

Ttofi and Farrington have discussed the most important components for school-based bullying prevention projects, informed by their systematic review and meta-analysis⁹⁸ undertaken for the Campbell Collaboration.⁴⁹ The existing evidence supports the use of a whole-school approach, such as the one piloted in this study, and the INCLUSIVE intervention includes several features of effective programme components: teacher training, improved classroom management and changes to disciplinary methods.⁹⁸ This further supports the intervention, and they recommend more RCTs and conclude that ‘the time is ripe to mount a new program of research on the effectiveness of anti-bullying programs’ (p. 48).

Since the pilot study commenced, there has also been one study published comparing the acceptability and validity of the CHU-9D and EQ5D-Y scales of child health utility in a UK population aged 6 and 7 years.⁹⁵ Although their performances were reasonably similar, the authors concluded that the CHU-9D was preferable, largely because the CHU-9D health classification system was explicitly developed using children’s input and the reinforcement of the time frame (the CHU-9D refers to ‘today’ at the end of every response option, which was considered an advantage for children). This study also reported ‘fair to moderate’ test–retest reliability coefficients for the CHU-9D and extremely high completion rates. Our completion rates were lower, but the CHU-9D was placed at the very back of the baseline survey questionnaire and some students did not have time to finish it. This assumption is supported by data on response rates from the follow-up: the CHU-9D was moved from the end of the booklet to before the other secondary outcomes and the proportion of incomplete responses reduced to 16%. The possibility that some participants did not fully understand all of the questions cannot be completely ruled out, although no systematic patterns were identified, which suggests time was the main barrier to completion.

Developing a framework for economic evaluation for a full trial

We have shown that a Phase III cluster RCT with an economic evaluation of cost-effectiveness would be feasible. However, it would not be without challenges. Not least because one finding of the recent Campbell Collaboration (systematic review and meta-analysis) of school-based interventions to reduce bullying was that ‘There never has been a cost–benefit anti-bullying program analysis’ (p. 46).⁹⁸ Our pilot economic analyses support the use of the CHU-9D scale for this population (see *Chapter 4, Economic evaluation*). Below we discuss, first, the value of CUA within a subsequent Phase III trial, although this should be supplemented with a CCA, measuring the health effects on school staff as well as students. We do not, however, view supplementary decision modelling of long-term outcomes as being a valuable addition to the trial (see *Which health effects should be included?*). We also outline below why we believe an economic evaluation should adopt a public sector cost perspective and focus on promoting equality (*A public sector cost perspective focused on promoting equality*).

Choice of outcome measure and analytic methods

Public health methods guidance from NICE states that economic evaluations should usually be undertaken with health effects measured using a non-monetary outcome indicator, preferably QALYs.^{89,90} We found that the basic psychometric properties of the CHU-9D, and its acceptability, were satisfactory, and that it

would be a suitable instrument to use in a Phase III trial. However, it is clear from the weak correlations between the CHU-9D and the proposed primary outcome measures that QALYs are unlikely to capture all of intervention effects. Indeed, Coast has questioned the validity of ‘funneling’ all outcomes into one measure.⁹⁹ Instead, she has proposed the concept of a CCA. In this approach, the consequences of different options are contrasted clearly and explicitly in a simple tabular form. There is no formal attempt to synthesise outcomes; rather, they are presented as a series of outcomes in their own right per trial arm, so that differences can be highlighted. In theory, different decision-makers can focus on the (costs and) outcomes that are relevant to them, rather than being presented with some form of aggregated result, such as a QALY. However, the main limitation of a CCA is that trade-offs between different outcomes are not made explicit and results are difficult to interpret when a particular intervention is associated with ‘better’ outcomes on some scales, but ‘poorer’ scores on others. For a Phase III trial, we would plan to perform CCA as a ‘secondary analysis’, as recommended by NICE.⁸⁹

Which health effects should be included?

It is reasonable to suggest that reducing students’ aggressive behaviours could impact on teachers’ health and QoL. Indeed, the ESYTC school misbehaviour subscale contains questions that refer to aggressive behaviour directed towards staff. However, we chose not to include a preference-based outcome measure in the teacher survey, as it was unlikely to have provided helpful information in terms of designing the Phase III trial. That is, the pilot teacher survey was designed to show that it was possible to ‘reach’ staff and that survey completion rates were acceptable, which it did (see *Appendix 4*). Moreover, as the choice of instrument for teachers is likely to be the EQ-5D, further investigating its psychometric properties did not seem to be a valuable exercise, as this has already been validated. We would plan to include the EQ-5D in the teacher surveys in a Phase III trial.

The evaluation time horizon should be chosen to incorporate all important costs and consequences. In the context of INCLUSIVE, this statement can be interpreted as being ‘lifelong’, as the impact of bullying is associated with increases in risk behaviours such as substance use,^{11–13} long-term emotional, behavioural and mental health problems,^{14,16,100} and self-harm and suicide.²⁰ The standard approach to incorporating longer-term outcomes in an economic evaluation based alongside a RCT is to undertake supplementary decision modelling: that is to link observed trial effects to longer-term health outcomes and their associated costs. For example, RCTs of treatments for human immunodeficiency virus typically report changes in immunological and virological markers of disease progression, and economic evaluations of these treatments link these changes to subsequent changes in mortality and morbidity, often using observational data.

In 2009, Hummel *et al.*⁹² built an economic model for the NICE public health programme to assess the cost-effectiveness of universal interventions that aimed to promote emotional and social well-being in secondary schools. Although they identified a number of published longitudinal studies, they concluded that existing data sets did not contain appropriate information to estimate these longer-term effects and emphasised the ‘many caveats’ with respect to the number of assumptions made by the modelling required to estimate long-term effects. The NICE public health methods guidance clearly states that complex modelling should be avoided if it is likely to create cost-effectiveness estimates that are highly uncertain.⁸⁹ For this reason, we do not currently view supplementary decision modelling as being a valuable addition to the trial.

A public sector cost perspective focused on promoting equality

Public health methods guidance from NICE recommends that the base-case cost-effectiveness estimate is presented from a public sector perspective, as this allows the costs and benefits of more than one central/local government body to be taken into account.⁸⁹ In this pilot study, the acceptability of basic questions regarding NHS use within the previous 12 months, contact with the police, truancy and exclusion from school were assessed. Although the completion rates for all four questions were encouragingly high (> 93%), as only information to estimate the price of exclusions exists, it is difficult to know exactly how a full economic evaluation using data recorded in this manner would accurately value truancy, NHS use and

police contact, as measured here. Moreover, during the pilot it transpired that the police and/or community support officers are routinely based in schools, which could count as a 'contact'. For these reasons, we plan to revise the NHS, police and truancy questions for a future Phase III trial.

Data linkage may be able to support longer-term cost-effectiveness estimates via accessing anonymised routinely collected records on individuals' use of the health service, education and the criminal justice system. The potential for collecting unique pupil numbers from schools at baseline in order to facilitate long-term data linkage should be explored prior to a Phase III trial if possible. A further issue that was not included in the pilot study relates to the estimation of staff time taken up in dealing with aggressive behaviour by students. We believe this to be an important issue and plan to include questions to measure this in the teacher survey in a Phase III trial. Note that we do not anticipate any particular problems in measuring the resources required to deliver INCLUSIVE.

Finally, although promoting equality is not typically central to the design of an economic evaluation, its importance is emphasised in the NICE public health methods guidance.⁸⁹ In the context of INCLUSIVE, this statement could be interpreted as meaning that cost-effectiveness estimates should be presented for different socioeconomic groups if effect sizes are found to differ. We propose to adopt this approach in a Phase III trial.

Chapter 6 Conclusion and recommendations for further research

This study has allowed us to examine systematically the feasibility and acceptability of the INCLUSIVE intervention for initiating change locally in bullying and aggression through the school environment. It has also enabled us to assess the feasibility of conducting a Phase III RCT and to inform decisions about outcomes to be explored in such a trial.

Progression to a Phase III trial

Our results suggest that INCLUSIVE is a feasible and acceptable intervention. All progression criteria were met despite challenges arising from the later-than-planned start date and the particularly challenging schools included in our purposive sampling frame. Therefore, in accordance with the MRC framework for evaluating complex interventions, this suggests progression to a Phase III trial is appropriate.⁶⁶ A cluster RCT would establish the effectiveness and cost-effectiveness of such a whole-school restorative approach in addressing aggressive behaviours. The evidence provided by such a trial will be important in helping schools, local authorities and the NHS improve health, improve behavioural and educational outcomes for young people, and reduce health inequalities. The RQs it should address include:

- Is the INCLUSIVE intervention implemented over 3 school years more effective and cost-effective than standard practice in reducing bullying and aggression in English secondary schools?
- Is the INCLUSIVE intervention more effective than standard practice in improving students' QoL, well-being, and psychological function and attainments and in reducing school exclusion and truancy, substance use, sexual risk, NHS use and police contacts among students?
- Is the INCLUSIVE intervention more effective than standard practice in improving staff QoL and attendance and reducing staff 'burn-out'?
- What factors moderate and mediate the effectiveness of the INCLUSIVE intervention?

However, we do not want to underplay some of the challenges identified in planning and delivering a whole-school restorative approach, especially in very large secondary schools. Although these challenges did not impede implementation during this study, to ensure the project is delivered over a full 3-year period of implementation, refinements and additional resources may be needed so that staff training reaches the whole school in an engaging format(s), needs assessment data are accessible and curriculum materials are responsive to schools' needs. Our recommendations focus on how we would optimise the intervention and build on, and refine, our existing trial methods to undertake a Phase III trial in 2014.

Intervention design and delivery

Whole-school approach

Launching the intervention throughout the school was identified as a way of ensuring greater engagement with the intervention across all school groups. This would be strengthened via:

- devoting more resources to timely launch events, targeting both students and staff
- a greater web presence, including interactive online content to engage schools during a Phase III trial
- the use of locally adaptable newsletters to inform staff and students about the activities and outputs arising from the intervention
- arranging annual events to celebrate progress and achievements.

Needs assessment survey

Needs assessment data collection and tailored reports for each school were an acceptable and powerful external input that helped all the school action groups to identify priorities. This should remain integral to the intervention approach and logic model. It could be improved by:

- ensuring that needs assessment data are delivered to intervention schools in a timely and accessible manner at the start of the intervention (i.e. September–October)
- adopting a needs assessment approach that aims to identify the ‘positive’ features of the school environment as well as challenges and needs (i.e. an ‘assets-based’ approach)
- continuing to compare each school against the average, but ensuring facilitators aid in the interpretation of these data, including through benchmarking against other schools with a similar socioeconomic intake as well as the average overall
- ensuring that all reporting is accessible and student centred
- using annual surveys in intervention schools to monitor progress and identify new/ongoing priorities.

Action groups

Action groups are an innovative and powerful mechanism for supporting student-led change to address key school-level risk and protective factors for aggressive behaviour. The action groups could be improved via:

- recruiting students from a mixture of years (e.g. years 7–9) into the action group and, when necessary, inviting particular students to participate in order to ensure a diverse group
- ensuring that the head or deputy head teacher is a member of the group to make sure it has sufficient power to change school policies
- external facilitators working with the action group co-ordinator to identify the best time(s) for meetings locally and helping them consider any practical barriers and how these might be overcome
- providing what students and staff identified as a ‘grown-up’ environment for group meetings.

External facilitators

External facilitators were consistently reported to have provided a highly valuable, ‘external push’ for schools. This support should be maintained in a Phase III trial. Ideally, the existing facilitators should be retained if possible, and additional educational consultants also recruited. External facilitators will continue to be educational consultants with former school leadership experience. An intervention manager will be required, to provide training and support to ensure programme fidelity by the external facilitators, including through use of a virtual learning environment to share resources and examples of best practice online. The intervention team should be managed separately from the research team, and should be housed within an educational institution, which we envisage would be the Institute of Education at the University of London. Key roles and responsibilities for external facilitators will include:

- establishing an effective ongoing working relationship with the SMT in their schools
- arranging a ‘catch-up’ call with school’s intervention lead in advance of each action group meeting to support planning, allocation of tasks and administration
- advocating for both students and staff to promote ‘equality of voice’ and effective decision-making involving representatives of the whole school
- supporting the co-ordination of training and curriculum implementation as required at their school(s).

Staff training in restorative practices

Staff training in restorative practices was consistently identified as being a critical component in implementing a whole-school restorative approach to behaviour change. However, a number of challenges emerged in terms of delivery and staff engagement. In a future trial, it would be essential to ensure trainers are aware of each school’s particular context, and improve the timeliness and reach of training.

Therefore, the external facilitators would themselves be trained to provide training in the schools with which they work. Training would also be improved via:

- ensuring that training is undertaken at the start of the school year to pump-prime other activities and increase awareness of the intervention across the school
- a comprehensive pre-training audit to identify schools' needs, what they hoped to achieve and the most appropriate method to 'cascade' learning through the whole school
- more engaging, interactive training methods using 'realistic' examples from similar secondary schools
- ensuring that students from the action team attend the training, and they are included and engaged in it.

Social and emotional skills curriculum

The student social and emotional skills curriculum was consistently identified as being a valuable and flexible component. In our pilot trial, as a result of the time needed for curriculum planning, this was delivered only in the third (summer) term. The curriculum could be refined further via:

- greater advanced planning and preparation with each school's PSHE lead
- the addition of more interactive activities.

New intervention partnerships

New intervention partnerships should also be developed with the Department of Education and Ofsted to maximise synergy with the broader policy environment and assessment frameworks within which English secondary schools operate. Similarly, consultations with further public health stakeholders in England should take place to explore how this intervention can be integrated and mainstreamed with ongoing policy programmes that aim to increase access to psychological therapies and support the social and emotional aspects of learning more strategically.

Trial design and methods

The methods were feasible and acceptable, with all schools remaining in the study and extremely high student response rates (> 93% at baseline and follow-up), but further refinements are, nonetheless, suggested based on the learning from the pilot.

Primary outcomes

The primary outcomes investigated within a Phase III trial should include one measure of bullying victimisation and one measure of the perpetration of aggressive behaviours. Our GBS and ESYTC measures performed satisfactorily and should therefore constitute these two outcomes, respectively. The trial should be powered on the basis of these two outcomes.

Secondary outcomes

The secondary outcomes investigated in a Phase III trial would include all those hypothesised for the pilot as well as validated measures of drug use and sexual risk behaviour (age at sexual debut; and contraception use at last sex). We will also seek to measure educational attainment, as the intervention is hypothesised to have demonstrable effects in this area, and this would probably be powerful evidence in enabling the scaling up of the intervention.

Sample of schools

The sample of schools in a Phase III trial should continue to be diverse but should reflect the overall population profile of schools in the study area (south-east England), rather than aiming to oversample particularly challenging schools (as the pilot did to ensure a diverse range of contexts were included). This pilot trial was initiated in July 2011, 3 months after originally planned, which seriously impeded our ability to recruit schools, although this was nonetheless completed.

Recruitment of schools for a Phase III trial will require a longer lead-in period. We recommend that:

- the project be initiated in February to enable liaison with schools to proceed for 4–6 months before the summer holidays
- the trial team partners with existing networks of schools such as the Institute of Education's 'Teaching Schools' network and other school practice networks, such as UCL Partners schools network and 'Challenge Partners' (www.challengepartners.org)
- comparison schools continue to be offered £500 in total to cover the expenses for data collection as well as a report of information from baseline and follow-up surveys (once the trial has been completed).

Surveys

Surveys (of students and teachers) should be conducted in the summer term each year, with baseline surveys undertaken in the summer term prior to the school year in which the intervention occurs. Additionally we recommend that:

- information is provided to students and staff in more accessible language, including about how anonymity is maintained within the trial
- a Phase III trial should include baseline and follow-up surveys with teachers but also with teaching assistants and other school staff
- staff surveys should also be conducted in the summer term prior to the school year in which the intervention occurs and prior to randomisation
- baseline and follow-up surveys should be undertaken with the action team, including items on students' social and educational characteristics.

Although this study found pen-and-paper questionnaires were feasible, were acceptable and produced a high response rate among students and staff, and some schools suggested that using school information technology facilities to undertake an online survey would not be feasible, we recommend that the cost of new technologies (e.g. small, touch-screen 'tablet' devices with 4G web capability) is reviewed prior to a Phase III trial. Although there is an up-front cost in purchasing such equipment, such online survey methods may improve data quality significantly and deliver savings overall, as a result of minimising any costs associated with printing questionnaires and inputting, checking and archiving data. In addition to reviewing the economic costs and benefits, we would also consult the NCB YRG and teachers to get their views on this method, and on the risks and benefits of online surveys in schools, prior to making a decision.

Quantitative data on intervention fidelity

Quantitative data on intervention fidelity should be collected in a Phase III trial. In addition to provider checklists and the documentary evidence that was collected in this pilot trial, the following data should also be collected:

- structured independent assessments of intervention delivery drawing on audio recordings and observations of a sample of action team meetings, training sessions and curriculum sessions
- information on the professional background and other characteristics of each action group facilitator in order to analyse how implementation and/or effects may vary
- measures of coverage regarding all relevant intervention inputs and outputs and, when relevant, including baseline assessment.

Qualitative data

Qualitative data should continue to be collected in a main trial as part of its integral process evaluation in order to assess unexpected processes, explore causal pathways and assess variation in implementation by context.

Economic evaluation

Economic evaluation will be a core element of a Phase III cluster RCT. Our pilot economic analyses support the use of the CHU-9D scale with this population and the feasibility of CUA, although this should be supplemented with a CCA. However, at present, we do not believe that undertaking (complex) modelling to link observable trial outcomes to longer-term (health) events is warranted, given the inherent limitations with the existing evidence base. Such an exercise is likely to produce cost-effectiveness estimates that are so uncertain as to be of little practical use. Anonymised data linkage may also support longer-term cost-effectiveness analyses via routinely collected health-service, education and criminal justice system data. The potential for collecting unique pupil numbers from schools at baseline, which could be linked to national health, education and crime databases, should be explored prior to a Phase III trial in order to facilitate long-term data linkage if possible. These findings suggest that student participation, particularly 'having a say' in revising the school rules, is a core component of the intervention, and therefore our economic analyses within a Phase III trial would also estimate these ongoing costs in order to facilitate potential longer-term analysis via data linkage.

Scalability and generalisability

A key criterion for assessing the importance of public health interventions is their potential scalability.¹⁰¹ This must be considered in all phases of evaluation^{66,102} and needs to be a key focus when planning a Phase III trial of INCLUSIVE. The INCLUSIVE intervention is potentially scalable because of:

- its clear feasibility and acceptability in English schools
- its focus on schools' 'core business' of teaching, discipline and pastoral care, and its perceived potential by school managers to contribute towards schools' government-required mission of increasing educational attainment and improving behaviour
- its strong balance of fidelity to intervention processes and core components with adaption of non-core outputs to local needs and existing policies and practices.^{71,72}

Furthermore, the intervention has the potential to achieve a wider range of health benefits and reductions in health inequalities beyond reductions in student bullying and aggression. This is because the intervention aims to address a variety of school-level and individual-level risk factors for a range of intercorrelated health behaviours¹⁰³⁻¹⁰⁵ that affect health across the life-course.⁵⁸ The intended secondary health outcomes at this pilot stage included improvements in students' QoL and psychological well-being, as well as reductions in psychological distress and substance use. In addition to these, we intend a Phase III trial to examine intervention effects on drug use and sexual risk behaviour. This plurality of intended health benefits also contributes to the potential scalability of the intervention, in contrast to numerous single-focus curriculum interventions addressing each of these outcomes separately, which are unlikely to find a place in busy school timetables.

The potential for scale-up will be a key focus of the process evaluation within a Phase III trial. As well as working with existing school networks, we will explore with them issues to consider in scale up and the most appropriate ways to proceed to this should the intervention be determined to be effective. A key factor to consider is the appropriate provider. In the Phase III trial, the intervention facilitators will be based at the Institute of Education working in collaboration with the various networks of secondary schools discussed above, such as the Institute of Education's 'Teaching Schools' network and 'Challenge Partners'. This is appropriate in that it allows us both to recruit schools more efficiently and to manage the intervention carefully in order to maximise fidelity. It is also appropriate because the Institute of Education has a long-standing role in capacity building and professional development from working with a variety of networks of secondary schools. These arrangements are therefore promising ones for the delivery of a scaled-up intervention, should it be demonstrated as effective in a Phase III trial.

Acknowledgements

The pilot trial was funded by the HTA programme at the NIHR. We would also like to acknowledge the following funding bodies for their financial support towards the costs of the intervention: the Paul Hamlyn Foundation; the Big Lottery Fund; and the Coutts Charitable Trust. We would also like to thank all of the project field workers for their hard work and commitment to this study, and also our TSC for their constructive feedback and support throughout. Finally, we would like to thank all the students and staff at the eight participating schools, and the NCB young researchers, for their support and enthusiasm throughout the project.

Contributions of authors

Professor Russell Viner (Professor of Adolescent Health) and **Professor Chris Bonell** (Professor of Sociology and Social Intervention) conceived and designed the intervention and study, directed the research project and contributed to the drafting and editing of the report; Professor Chris Bonell also undertook some of the qualitative data collection.

Dr Adam Fletcher (Senior Lecturer in Social Science and Health) managed the trial, led on the qualitative data collection and analysis, contributed to the design of the intervention and study, and led on writing this report.

Ms Natasha Fitzgerald-Yau (Research Assistant, Adolescent Health) and **Dr Daniel Hale** (Research Associate, Adolescent Health) supported the data collection process, facilitated the involvement of young people and contributed to the analysis of quantitative data and the writing of the report.

Dr Elizabeth Allen (Senior Lecturer in Medical Statistics), **Professor Diana Elbourne** (Professor of Healthcare Evaluation) and **Ms Rebecca Jones** (Research Associate, Medical Statistics) were the trial statisticians and led on random allocation and the analysis and interpretation of quantitative data.

Professor Lyndal Bond (Professor of Adolescent Health) advised on the design of the intervention and the trial and commented on drafts of the final report.

Ms Meg Wiggins (Senior Research Officer in Childhood, Families and Health) advised on intervention management, undertook some of the qualitative data collection and commented on drafts of the final report.

Dr Alec Miners (Senior Lecturer in Health Economics) and **Dr Rosa Legood** (Lecturer in Decision Modelling) led on the economic evaluation.

Professor Stephen Scott (Professor of Child Health and Behaviour) contributed to the development of the study, advised on process evaluation methods and commented on drafts of the final report.

Dr Deborah Christie (Clinical Psychologist and Honorary Reader in Paediatric Adolescent Psychology) contributed to the development of the study, including questionnaire development, and commented on drafts of the final report.

References

1. Scott S, Knapp M, Henderson J, Maughan B. Financial cost of social exclusion: follow up study of antisocial children into adulthood. *BMJ* 2001;**323**:191.
2. Krug EG, Mercy JA, Dahlberg LL, Zwi AB. The world report on violence and health. *Lancet* 2002;**360**:1083–8. [http://dx.doi.org/10.1016/S0140-6736\(02\)11133-0](http://dx.doi.org/10.1016/S0140-6736(02)11133-0)
3. McKee M, Raine R. Riots on the streets. *BMJ* 2011;**343**:d5248. <http://dx.doi.org/10.1136/bmj.d5248>
4. Fletcher A, Gardner F, McKee M, Bonell C. The British government's Troubled Families Programme. *BMJ* 2012;**344**:e3403.
5. World Health Organization. *World Report on Violence and Health*. Geneva: World Health Organization; 2002. URL: www.who.int/violence_injury_prevention/violence/world_report/en/full_en.pdf (accessed November 2012).
6. Craig WM, Henderson K, Murphy JG. Prospective teachers' attitudes toward bullying and victimization. *School Psychol Int* 2000;**21**:5–21. <http://dx.doi.org/10.1177/0143034300211001>
7. Newman RS, Murray B, Lussier C. Confrontation with aggressive peers at school: students' reluctance to seek help from the teacher. *J Edu Psychol* 2001;**93**:398. <http://dx.doi.org/10.1037/0022-0663.93.2.398>
8. Nansel TR, Craig W, Overpeck MD, Saluja G, Ruan W. Cross-national consistency in the relationship between bullying behaviors and psychosocial adjustment. *Arch Pediat Adol Med* 2004;**158**:730. <http://dx.doi.org/10.1001/archpedi.158.8.730>
9. Radford L, Corral S, Bradley C, Fisher H, Bassett C, Howat N, et al. *Child Abuse and Neglect in the UK Today*. London: National Society for the Prevention of Cruelty to Children; 2011.
10. Arseneault L, Walsh E, Trzesniewski K, Newcombe R, Caspi A, Moffitt TE. Bullying victimization uniquely contributes to adjustment problems in young children: a nationally representative cohort study. *Pediatrics* 2006;**118**:130–8. <http://dx.doi.org/10.1542/peds.2005-2388>
11. Forero R, McLellan L, Rissel C, Bauman A. Bullying behaviour and psychosocial health among school students in New South Wales, Australia: cross sectional survey. *BMJ* 1999;**319**:344–8. <http://dx.doi.org/10.1136/bmj.319.7206.344>
12. Kaltiala-Heino R, Rimpelä M, Rantanen P, Rimpelä A. Bullying at school – an indicator of adolescents at risk for mental disorders. *J Adolescence* 2000;**23**:661–74. <http://dx.doi.org/10.1006/jado.2000.0351>
13. Juvonen J, Graham S, Schuster MA. Bullying among young adolescents: the strong, the weak, and the troubled. *Pediatrics* 2003;**112**:1231–7. <http://dx.doi.org/10.1542/peds.112.6.1231>
14. Hawker DS, Boulton MJ. Twenty years' research on peer victimization and psychosocial maladjustment: a meta-analytic review of cross-sectional studies. *J Child Psychol Psychiatry* 2000;**41**:441–55. <http://dx.doi.org/10.1111/1469-7610.00629>
15. Bond L, Carlin JB, Thomas L, Rubin K, Patton G. Does bullying cause emotional problems? A retrospective study of young teenagers. *BMJ* 2001;**323**:480–4.
16. Arseneault L, Bowes L, Shakoor S. Bullying victimization in youths and mental health problems: 'much ado about nothing'. *Psychol Med* 2010;**40**:717–29. <http://dx.doi.org/10.1017/S0033291709991383>

17. Van der Wal MF, De Wit CA, Hirasing RA. Psychosocial health among young victims and offenders of direct and indirect bullying. *Pediatrics* 2003;**111**:1312–17. <http://dx.doi.org/10.1542/peds.111.6.1312>
18. Woods S, Wolke D. Direct and relational bullying among primary school children and academic achievement. *J School Psychol* 2004;**42**:135–55. <http://dx.doi.org/10.1016/j.jsp.2003.12.002>
19. Glew GM, Fan M-Y, Katon W, Rivara FP, Kernic MA. Bullying, psychosocial adjustment, and academic performance in elementary school. *Arch Pediat Adol Med* 2005;**159**:1026. <http://dx.doi.org/10.1001/archpedi.159.11.1026>
20. Crick NR, Ostrov JM, Werner NE. A longitudinal study of relational aggression, physical aggression, and children's social–psychological adjustment. *J Abnorm Child Psych* 2006;**34**:127–38. <http://dx.doi.org/10.1007/s10802-005-9009-4>
21. Price LH, Kao H-T, Burgers DE, Carpenter LL, Tyrka AR. Telomeres and early-life stress: an overview. *Biol Psychiat* 2013;**73**:15–23. <http://dx.doi.org/10.1016/j.biopsych.2012.06.025>
22. Wolke D, Woods S, Stanford K, Schulz H. Bullying and victimization of primary school children in England and Germany: prevalence and school factors. *Brit J Psychol* 2001;**92**:673–96. <http://dx.doi.org/10.1348/000712601162419>
23. Beinart S, Anderson B, Lee S, Utting D. *Youth at risk?: A National Survey of Risk Factors, Protective Factors and Problem Behaviour Among Young People in England, Scotland and Wales (JRF Findings 432)*. York: Joseph Rowntree Foundation; 2002.
24. Olweus D. *Bullying at School: What We Know and What We Can Do*. Cambridge, MA: Blackwell; 1993.
25. Bender D, Lösel F. Bullying at school as a predictor of delinquency, violence and other anti-social behaviour in adulthood. *Crim Behav Ment Health* 2011;**21**:99–106. <http://dx.doi.org/10.1002/cbm.799>
26. Stansfeld S, Haines M, Booy R, Taylor S, Viner RM, Head JJ, et al. *Health of Young People in East London: The RELACHS Study 2001*. London: Institute of Community Health Sciences, Barts and the London Queen Mary's School of Medicine and Dentistry; 2003.
27. Battistich V, Hom A. The relationship between students' sense of their school as a community and their involvement in problem behaviors. *Am J Public Health* 1997;**87**:1997–2001. <http://dx.doi.org/10.2105/AJPH.87.12.1997>
28. National Center for Educational Statistics. *Indicators of School Crime and Safety: 2007*. Washington, DC: National Center for Educational Statistics; 2007.
29. Jansen DE, Veenstra R, Ormel J, Verhulst FC, Reijneveld SA. 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**:440. <http://dx.doi.org/10.1186/1471-2458-11-440>
30. Sainsbury Centre for Mental Health. *Diversion: A Better Way for Criminal Justice and Mental Health*. London: Sainsbury Centre for Mental Health; 2009.
31. Department for Education and Skills (DfES). *White Paper. The Future of Higher Education*. London: DfES; 2003.
32. Department for Children, Families and Schools. *Your Child, Your Schools, Our Future: Building a 21st Century Schools System*. London: The Stationery Office (TSO); 2009.
33. Department of Health. *Healthy Lives, Brighter Futures*. London: Her Majesty's Stationery Office (HMSO); 2009.

34. Chamberlain T, Britain G. *Tellus4 National Report*. London: Department for Children, School and Families; 2010.
35. Home Office. *The Offending, Crime and Justice Survey 2006*. London: Her Majesty's Stationery Office (HMSO); 2006.
36. Brame B, Nagin DS, Tremblay RE. Developmental trajectories of physical aggression from school entry to late adolescence. *J Child Psychol Psyc* 2001;**42**:503–12. <http://dx.doi.org/10.1111/1469-7610.00744>
37. Boxer P, Goldstein SE, Musher-Eizenman D, Dubow EF. Exposure to 'low-level' aggression in school: associations with aggressive behavior, future expectations, and perceived safety. *Violence Vict* 2003;**18**:691–705. <http://dx.doi.org/10.1891/vivi.2003.18.6.691>
38. Steer A. *Learning Behaviour: Lessons Learned – A Review of Behaviour Standards and Practices in our Schools*. London: Department for Children, Schools and Families; 2007.
39. Schagen S, Blenkinsop S, Schagen I, Scott E, Eggers M, Warwick I, et al. Evaluating the impact of the National Healthy School Standard: using national datasets. *Health Educ Res* 2005;**20**:688–96. <http://dx.doi.org/10.1093/her/cyh023>
40. Department for Education (DfE). *Schools, Pupils and Their Characteristics, January 2011*. London: DfE; 2011.
41. Department for Education (DfE). *Positive for Youth*. London: DfE; 2011.
42. Morrell G, Scott S, McNeish D, Webster S. *The August Riots in England: Understanding the Involvement of Young People. A Report Prepared for the Cabinet Office*. London: NatCen; 2012.
43. Smith JD, Schneider BH, Smith PK, Ananiadou K. The effectiveness of whole-school antibullying programs: a synthesis of evaluation research. *School Psychol Rev* 2004;**33**:547–60.
44. Hahn R, Fuqua-Whitley D, Wethington H, Lowy J, Crosby A, Fullilove M, et al. Effectiveness of universal school-based programs to prevent violent and aggressive behavior: a systematic review. *Am J Prev Med* 2007;**33**:S114–29.
45. Limbos MA, Chan LS, Warf C, Schneir A, Iverson E, Shekelle P, et al. Effectiveness of interventions to prevent youth violence. *Am J Prev Med* 2007;**33**:65–74. <http://dx.doi.org/10.1016/j.amepre.2007.02.045>
46. Vreeman RC, Carroll AE. A systematic review of school-based interventions to prevent bullying. *Arch Pediat Adol Med* 2007;**161**:78. <http://dx.doi.org/10.1001/archpedi.161.1.78>
47. Wilson SJ, Lipsey MW. School-based interventions for aggressive and disruptive behavior: update of a meta-analysis. *Am J Prev Med* 2007;**33**:S130–43.
48. Park-Higgerson HK, Perumean-Chaney SE, Bartolucci AA, Grimley DM, Singh KP. The evaluation of school-based violence prevention programs: a meta-analysis. *J School Health* 2008;**78**:465–79. <http://dx.doi.org/10.1111/j.1746-1561.2008.00332.x>
49. Farrington DP, Ttofi MM. *School-Based Programs to Reduce Bullying and Victimization. Campbell Systematic Reviews* 2009:6.
50. Viner RM, Ozer EM, Denny S, Marmot M, Resnick M, Fatusi A, et al. Adolescence and the social determinants of health. *Lancet* 2012;**379**:1641–52. [http://dx.doi.org/10.1016/S0140-6736\(12\)60149-4](http://dx.doi.org/10.1016/S0140-6736(12)60149-4)
51. 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).

52. Dahlgren G, Whitehead M. *Policies and Strategies to Promote Social Equity in Health*. Stockholm: Institute for Future Studies; 1991.
53. Marmot M. Creating healthier societies. *B World Health Organ* 2004;**82**:320.
54. Flay BR, Graumlich S, Segawa E, Burns JL, Holliday MY, Aban Aya Investigators. Effects of 2 prevention programs on high-risk behaviors among African American youth: a randomized trial. *Arch Pediatr Adolesc Med* 2004;**158**:377–84. <http://dx.doi.org/10.1001/archpedi.158.4.377>
55. Bond L, Patton G, Glover S, Carlin JB, Butler H, Thomas L, et al. The Gatehouse Project: can a multilevel school intervention affect emotional wellbeing and health risk behaviours? *J Epidemiol Community Health* 2004;**58**:997–1003. <http://dx.doi.org/10.1136/jech.2003.009449>
56. Patton GC, Bond L, Carlin JB, Thomas L, Butler H, Glover S, et al. Promoting social inclusion in schools: a group-randomized trial of effects on student health risk behavior and well-being. *Am J Public Health* 2006;**96**:1582–7. <http://dx.doi.org/10.2105/AJPH.2004.047399>
57. Viner R. Co-occurrence of adolescent health risk behaviors and outcomes in adult life: findings from a national birth cohort. *J Adolescent Health* 2005;**36**:98–9. <http://dx.doi.org/10.1016/j.jadohealth.2004.11.012>
58. Buck D, Frosini F. *Clustering of Unhealthy Behaviours Over Time*. London: The King's Fund; 2012.
59. Hopkins B. *Just Schools: A Whole School Approach to Restorative Justice*. London: Jessica Kingsley; 2004.
60. Youth Justice Board for England and Wales (YJB). *National Evaluation of the Restorative Justice in Schools Programme*. London: YJB; 2004.
61. Kane JG LG, McCluskey S, Riddell J, Stead J, Weedon E. *Restorative Practices in Scottish Schools*. Edinburgh: Scottish Executive; 2007.
62. Skinnis L, Hough M. *An Evaluation of Bristol RAI's*. London: Institute for Criminal Policy Research, King's College London; 2009.
63. Minnesota Department of Children, Families and Learning. *A Three-Year Evaluation of Alternative Approaches to Suspensions and Expulsions*. Report to the legislature; 2002.
64. Shaw G. Restorative practices in Australian schools: changing relationships, changing culture. *Confl Resolut Quart* 2007;**25**:127–35. <http://dx.doi.org/10.1002/crq.198>
65. Wong DS, Cheng CH, Ngan RM, Ma SK. Program effectiveness of a restorative whole-school approach for tackling school bullying in Hong Kong. *Int J Offender Ther* 2011;**55**:846–62. <http://dx.doi.org/10.1177/0306624X10374638>
66. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;**337**:a1655. <http://dx.doi.org/10.1136/bmj.a1655>
67. Fletcher A, Bonell C, Sorhaindo A. 'We don't have no drugs education': the myth of universal drugs education in English secondary schools? *Int J Drug Policy* 2010;**21**:452–8.
68. Bonell C, Fletcher A, Sorhaindo A, Wells H, McKee M. How market-oriented education policies might influence young people's health: development of a logic model from qualitative case studies in English secondary schools. *J Epidemiol Community Health* 2012;**66**:e24. <http://dx.doi.org/10.1136/jech.2011.137539>
69. Bonell C, Sorhaindo A, Strange V, Wiggins M, Allen E, Fletcher A, et al. A pilot whole-school intervention to improve school ethos and reduce substance use. *Health Educ* 2010;**110**:252–72. <http://dx.doi.org/10.1108/09654281011052628>

70. 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**:1209–20. [http://dx.doi.org/10.1016/S0277-9536\(02\)00120-X](http://dx.doi.org/10.1016/S0277-9536(02)00120-X)
71. Hawe P, Shiell A, Riley T. Complex interventions: how 'out of control' can a randomised controlled trial be? *BMJ* 2004;**328**:1561. <http://dx.doi.org/10.1136/bmj.328.7455.1561>
72. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychol* 2008;**41**:327–50. <http://dx.doi.org/10.1007/s10464-008-9165-0>
73. Bond L, Butler H. The Gatehouse Project: A Multi-level Integrated Approach to Promoting Wellbeing in Schools. In Killoran A, Kelly M, editors. *Evidence-Based Public Health: Effectiveness and Efficiency*. Oxford: Oxford University Press; 2009. pp. 250–69.
74. Stephenson J, Strange V, Allen E, Copas A, Johnson A, Bonell C, *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**:e224.
75. Moon AM, Mullee MA, Thompson RL, Speller V, Roderick P. Health-related research and evaluation in schools. *Health Educ* 1999;**99**:27–34. <http://dx.doi.org/10.1108/09654289910248481>
76. Green J, Thorogood N. *Qualitative Methods for Health Research*. London: Sage; 2009.
77. McAra L, McVie S. Youth crime and justice: key messages from the Edinburgh Study of Youth Transitions and Crime. *Criminol Crim Just* 2010;**10**:179–209. <http://dx.doi.org/10.1177/1748895809360971>
78. Bond L, Wolfe S, Tollit M, Butler H, Patton G. A comparison of the Gatehouse Bullying Scale and the Peer Relations Questionnaire for students in secondary school. *J School Health* 2007;**77**:75–9. <http://dx.doi.org/10.1111/j.1746-1561.2007.00170.x>
79. Smith DJ. *School Experience and Delinquency at Ages 13 to 16*. Edinburgh: Centre for Law and Society, University of Edinburgh; 2006.
80. Varni JW, Burwinkle TM, Seid M. The PedsQL TM 4.0 as a school population health measure: feasibility, reliability, and validity. *Qual Life Res* 2006;**15**:203–15.
81. Stevens K. Assessing the performance of a new generic measure of health-related quality of life for children and refining it for use in health state valuation. *Appl Health Econ Health Policy* 2011;**9**:157–69. <http://dx.doi.org/10.2165/11587350-000000000-00000>
82. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psyc* 2006;**38**:581–6. <http://dx.doi.org/10.1111/j.1469-7610.1997.tb01545.x>
83. Clarke A, Friede T, Putz R, Ashdown J, Martin S, Blake 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**:487. <http://dx.doi.org/10.1186/1471-2458-11-487>
84. 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**:b2534. <http://dx.doi.org/10.1136/bmj.b2534>
85. 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 Psyc* 2010;**51**:199–209.
86. Epstein JL, McPartland JM. The concept and measurement of the quality of school life. *Am Edu Res J* 1976;**13**:15–30. <http://dx.doi.org/10.3102/00028312013001015>

87. Roeser RW, Midgley C, Urdan TC. Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: the mediating role of goals and belonging. *J Educ Psychol* 1996;**88**:408. <http://dx.doi.org/10.1037//0022-0663.88.3.408>
88. Goodenow C. Classroom belonging among early adolescent students relationships to motivation and achievement. *J Early Adolesc* 1993;**13**:21–43. <http://dx.doi.org/10.1177/0272431693013001002>
89. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal*. London: NICE; 2008.
90. National Institute for Health and Care Excellence (NICE). *Methods for the Development of NICE Public Health Guidance (Third Edition)*. London: NICE; 2012.
91. National Institute for Health and Care Excellence (NICE). *Methods for the Development of NICE Public Health Guidance (Second Edition)*. London: NICE; 2009.
92. Hummel S, Naylor P, Chilcott J, Guillaume L, Wilkinson A, Blank L, et al. *Cost-Effectiveness of Universal Interventions Which Aim to Promote Emotional and Social Wellbeing in Secondary Schools. Report for NICE Sheffield*. Sheffield: University of Sheffield; 2009.
93. Ozer EJ. Contextual effects in school-based violence prevention programs: a conceptual framework and empirical review. *J Prim Prev* 2006;**27**:315–40. <http://dx.doi.org/10.1007/s10935-006-0036-x>
94. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *BMJ* 1997;**314**:572. <http://dx.doi.org/10.1136/bmj.314.7080.572>
95. Canaway AG, Frew EJ. Measuring preference-based quality of life in children aged 6–7 years: a comparison of the performance of the CHU-9D and EQ-5D-Y – the WAVES pilot study. *Qual Life Res* 2013;**22**:173–83. <http://dx.doi.org/10.1007/s11136-012-0119-5>
96. Bonell C, Sorhaindo A, Allen E, Strange V, Wiggins M, Fletcher A, et al. Pilot multi-method trial of a school-ethos intervention to reduce substance use: building hypotheses about upstream pathways to prevention. *J Adolescent Health* 2010;**47**:555–63.
97. Bond L, Thomas L, Coffey C, Glover S, Butler H, Carlin JB, et al. Long-term impact of the gatehouse project on cannabis use of 16-year-olds in Australia. *J School Health* 2004;**74**:23–9. <http://dx.doi.org/10.1111/j.1746-1561.2004.tb06597.x>
98. Ttofi MM, Farrington DP. Effectiveness of school-based programs to reduce bullying: a systematic and meta-analytic review. *J Exp Criminol* 2011;**7**:21–56. <http://dx.doi.org/10.1007/s11292-010-9109-1>
99. Coast J. Is economic evaluation in touch with society's health values? *BMJ* 2004;**329**:1233. <http://dx.doi.org/10.1136/bmj.329.7476.1233>
100. 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**:368–83. <http://dx.doi.org/10.1177/109019810102800310>
101. Ogilvie D, Cummins S, Petticrew M, White M, Jones A, Wheeler K. Assessing the evaluability of complex public health interventions: five questions for researchers, funders, and policymakers. *Milbank Q* 2011;**89**:206–25. <http://dx.doi.org/10.1111/j.1468-0009.2011.00626.x>
102. 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**:2299–306. <http://dx.doi.org/10.1016/j.socscimed.2012.08.032>
103. Jackson C, Sweeting H, Haw S. Clustering of substance use and sexual risk behaviour in adolescence: analysis of two cohort studies. *BMJ Open* 2012;**2**:e000661. <http://dx.doi.org/10.1136/bmjopen-2011-000661>

104. Hawkins JD, Catalano RF, Kosterman R, Abbott R, Hill KG. Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Arch Pediat Adol Med* 1999;**153**:226. <http://dx.doi.org/10.1001/archpedi.153.3.226>
105. Bonell CP, Parry W, Wells H, Jamal F, Fletcher A, Harden A, *et al.* The effects of the school environment on student health: a systematic review of multi-level studies. *Health Place* 2013;**21**:180–91. <http://dx.doi.org/10.1016/j.healthplace.2012.12.001>

Appendix 1 Trial Steering Committee membership

Professor Rona Campbell, School of Social and Community Medicine, University of Bristol (chairperson); Dr John Coleman, Department of Education, University of Oxford; and Professor Carole Torgerson, School of Education, Durham University.

Appendix 2 Pilot secondary outcome analyses

Table 19 reports the pilot secondary outcomes at baseline. There were consistent differences in QoL as measured by the PedsQL between the intervention and comparison group, favouring the latter, but there were no differences in terms of the CHU-9D, which is analysed in detail in the *Economic evaluation* in Chapter 4. At baseline, those in the comparison group were observed to have lower levels of psychological distress (as measured by the SDQ) and improved emotional well-being (as measured by the SWEMWBS). In terms of substance use, the comparison group reported lower rates of smoking and drinking than the intervention group. Overall, only 1.4% of students reported that they had smoked in the last month, and 5.9% stated that they had drunk alcohol (categorised as having more than a 'sip') in the last month. There were few differences in rates of NHS use, truancy, school exclusions and police contact (Table 19). Nearly half of the sample had used NHS services once or more in the last year. A small minority reported that they had previously been excluded temporarily or permanently from school (4%), played truant (6%) or had contact with the police (e.g. stopped or picked up) in the last 12 months (11.7%).

Table 20 reports the pilot secondary outcomes at follow-up at both the intervention and comparison schools (mean scores and standard deviations/proportions). As at baseline, at follow-up students in the comparison group also reported a slightly better QoL (PedsQL), lower levels of psychological distress (SDQ), improved psychological well-being (SWEMWBS) and lower rates of smoking and drinking in the last month.

TABLE 19 Pilot secondary outcomes at baseline

Measure	Comparison [mean (SD)]	Intervention [mean (SD)]	Overall [mean (SD)]
QoL (PedsQL)			
PedsQL overall score	84.08 (12.72)	82.04 (12.46)	82.99 (12.62)
Physical	88.48 (12.53)	86.59 (13.09)	87.48 (12.86)
Emotional	78.62 (19.92)	78.41 (18.24)	78.51 (19.03)
Social	89.28 (15.84)	86.36 (16.96)	87.71 (16.51)
School	77.08 (18.12)	74.13 (17.36)	75.49 (17.76)
Psychosocial	81.65 (14.89)	79.64 (14.12)	80.58 (14.51)
Health utility – CHU-9D			
CHU-9D overall score	0.88 (0.10)	0.88 (0.09)	0.88 (0.10)
Psychological distress (SDQ)			
SDQ overall score	10.04 (5.62)	11.17 (5.25)	10.63 (5.46)
Emotional problems	2.65 (2.24)	2.80 (2.18)	2.73 (2.21)
Conduct problems	2.02 (1.73)	2.20 (1.78)	2.11 (1.75)
ADHD symptoms	3.91 (2.40)	4.39 (2.35)	4.16 (2.38)
Peer problems	1.64 (1.73)	1.85 (1.66)	1.75 (1.70)
Psychological well-being (SWEMWBS)			
WEMWBS overall score	24.35 (5.43)	23.01 (5.83)	23.62 (5.68)
Smoking			
No smoking this month	498 (98.8)	534 (98.3)	1032 (98.6)
Smoking this month	6 (1.2)	9 (1.7)	15 (1.4)
Alcohol consumption			
No alcohol this month	469 (95.1)	504 (93.2)	973 (94.1)
Alcohol this month	24 (4.9)	37 (6.8)	61 (5.9)
NHS use			
None in the last year	262 (50.3)	296 (54.2)	558 (52.3)
Once or more in the last year	259 (49.7)	250 (45.8)	509 (47.7)
Contact with police			
No	470 (89.5)	486 (87.1)	956 (88.3)
Yes	55 (10.5)	72 (12.9)	127 (11.7)
Truancy			
No	492 (93.5)	524 (94.4)	1016 (94.0)
Yes	34 (6.5)	31 (5.6)	65 (6.0)
Exclusion from school			
No	504 (95.8)	536 (96.2)	1040 (96.0)
Yes	22 (4.2)	21 (3.8)	43 (4.0)

ADHD, attention deficit hyperactivity disorder; SD, standard deviation.

TABLE 20 Pilot secondary outcomes at follow-up

Measure	Comparison [mean (SD)]	Intervention [mean (SD)]
QoL (PedsQL)		
PedsQL overall score	84.71 (12.45)	82.53 (12.81)
Physical	89.94 (11.63)	87.88 (13.76)
Emotional	78.80 (20.16)	78.63 (19.36)
Social	88.07 (15.86)	85.16 (16.12)
School	78.71 (16.99)	75.22 (17.24)
Psychosocial	81.94 (14.54)	79.75 (14.49)
Health utility – CHU-9D		
CHU-9D overall score	0.87 (0.10)	0.87 (0.10)
Psychological distress (SDQ)		
SDQ overall score	9.47 (5.42)	10.68 (5.68)
Emotional problems	2.39 (2.28)	2.49 (2.15)
Conduct problems	1.96 (1.76)	2.30 (1.76)
ADHD symptoms	3.67 (2.31)	4.14 (2.30)
Peer problems	1.81 (1.91)	1.89 (1.78)
Psychological well-being (SWEMWBS)		
WEMWBS overall score	24.21 (5.18)	24.13 (5.01)
Other outcomes		
Smoking in the last month	14 (3.2)	16 (3.5)
Alcohol consumption in the last month	34 (8.0)	55 (12.1)
NHS use in the last year	269 (58.6)	273 (59.7)
Truancy	48 (10.2)	53 (11.1)
Exclusion from school	33 (7.1)	32 (6.7)
Contact with police	53 (11.3)	71 (14.9)

ADHD, attention deficit hyperactivity disorder; SD, standard deviation.

Appendix 3 Pilot intermediate outcome analyses

There were small differences between intervention and comparison groups at baseline in terms of the pilot intermediate outcomes: intervention schools appeared to have a slightly more positive school climate as measured using the BBSCQ. The BBSCQ overall score and mean across all four domains (teacher relationships, belonging, participation and commitment) were higher in the intervention group than in the comparison group. Students of intervention schools also reported a lower level of delinquent misbehaviour at school than those in the comparison group (*Table 21*).

Table 22 reports the pilot intermediate outcomes at follow-up at both the intervention and comparison schools (mean scores and standard deviations/proportions). As at baseline, at follow-up the BBSCQ overall score and total mean score for each subscale were higher in the intervention schools, whereas delinquent behaviour increased among students in both arms.

TABLE 21 Pilot intermediate measures at baseline

Measure	Comparison [mean (SD)/n (%)]	Intervention [mean (SD)/n (%)]	Overall [mean (SD)/n (%)]
School environment (BBSCQ)			
BBSCQ overall score	1.79 (0.40)	1.90 (0.39)	1.85 (0.39)
Teacher relations	1.90 (0.53)	2.00 (0.51)	1.95 (0.52)
Belonging	1.95 (0.49)	2.10 (0.50)	2.02 (0.50)
Participation	1.72 (0.50)	1.82 (0.53)	1.77 (0.52)
Student commitment	1.38 (0.41)	1.44 (0.43)	1.41 (0.42)
Antischool actions (ESYTC SRD)			
SRD overall score	2.78 (3.05)	2.40 (2.90)	2.59 (2.98)

SD, standard deviation.

TABLE 22 Pilot intermediate measures at follow-up

Measure	Comparison [mean (SD)]	Intervention [mean (SD)]
School environment (BBSCQ)		
BBSCQ overall score	1.96 (0.42)	2.11 (0.42)
Teacher relations	2.13 (0.57)	2.28 (0.56)
Belonging	2.10 (0.51)	2.24 (0.53)
Participation	1.89 (0.57)	2.07 (0.57)
Student commitment	1.44 (0.46)	1.53 (0.49)
Antischool actions (ESYTC SRD)		
SRD overall score	2.91 (3.16)	3.03 (3.08)

SD, standard deviation.

Appendix 4 Teacher survey response rates

Pilot teacher surveys at baseline and follow-up found high response rates to questions across a range of domains, and low levels of missing data (typically < 10%) (Table 23).

TABLE 23 Teacher response rates at baseline and follow-up

Item	Response	Baseline [n (%)]	Follow-up [n (%)]
Avon Longitudinal Study of Parents and Children			
Most pupils at this school want to do well in tests and exams	Totally agree	236 (61.3)	225 (67.4)
	I agree a bit	138 (35.8)	96 (28.7)
	I do not really agree	10 (2.6)	5 (1.5)
	Totally disagree	0 (0.0)	0 (0.0)
	Missing	1 (0.3)	8 (2.4)
Pupils who get good marks or work hard are teased by the other pupils	Totally agree	10 (2.6)	3 (0.9)
	I agree a bit	77 (20.0)	69 (20.7)
	I do not really agree	213 (55.3)	184 (55.1)
	Totally disagree	81 (21.0)	71 (21.3)
	Missing	4 (1.0)	7 (2.1)
Most pupils in this school are interested in learning	Totally agree	173 (44.9)	148 (44.3)
	I agree a bit	178 (46.2)	166 (49.7)
	I do not really agree	26 (6.8)	12 (3.6)
	Totally disagree	2 (0.5)	0 (0.0)
	Missing	6 (1.6)	8 (2.4)
Many pupils do not do as well as they could because they are afraid that other pupils will not like them as much	Totally agree	12 (3.1)	10 (3.0)
	I agree a bit	82 (21.3)	63 (18.9)
	I do not really agree	189 (49.1)	169 (50.6)
	Totally disagree	97 (25.2)	84 (25.2)
	Missing	5 (1.3)	8 (2.4)
There is good extracurricular provision in this school	Totally agree	189 (49.1)	170 (50.9)
	I agree a bit	156 (40.5)	114 (34.1)
	I do not really agree	35 (9.1)	37 (11.1)
	Totally disagree	2 (0.5)	5 (1.5)
	Missing	3 (0.8)	8 (2.4)
There are very few pupils at this school whose behaviour in class prevents other pupils from learning	Totally agree	116 (30.1)	99 (29.6)
	I agree a bit	143 (37.1)	131 (39.2)
	I do not really agree	99 (25.7)	79 (23.7)
	Totally disagree	24 (6.2)	16 (4.8)
	Missing	3 (0.8)	9 (2.7)

continued

TABLE 23 Teacher response rates at baseline and follow-up (continued)

Item	Response	Baseline [n (%)]	Follow-up [n (%)]
Most pupils behave well in class	Totally agree	145 (37.7)	131 (39.2)
	I agree a bit	206 (53.5)	173 (51.8)
	I do not really agree	27 (7.0)	19 (5.7)
	Totally disagree	5 (1.3)	1 (0.3)
	Missing	2 (0.5)	10 (3.0)
There is not much bullying or name calling of each other by pupils	Totally agree	80 (20.8)	67 (20.1)
	I agree a bit	176 (45.7)	155 (46.4)
	I do not really agree	111 (28.8)	95 (28.4)
	Totally disagree	13 (3.4)	8 (2.4)
	Missing	5 (1.3)	9 (2.7)
Most of the students in my class enjoy being together	Totally agree	164 (42.6)	133 (39.8)
	I agree a bit	199 (51.7)	183 (54.8)
	I do not really agree	16 (4.2)	8 (2.4)
	Totally disagree	0 (0.0)	0 (0.0)
	Missing	6 (1.6)	10 (3.0)
I notice when students are doing a good job and let them know about it	Totally agree	278 (72.2)	236 (70.7)
	I agree a bit	103 (26.8)	87 (26.0)
	I do not really agree	1 (0.3)	2 (0.6)
	Totally disagree	0 (0.0)	0 (0.0)
	Missing	3 (0.8)	9 (2.7)
At this school, students have a lot of chances to help decide and plan school activities, events and policies	Totally agree	120 (31.2)	102 (30.5)
	I agree a bit	177 (46.0)	158 (47.3)
	I do not really agree	70 (18.2)	57 (17.1)
	Totally disagree	13 (3.4)	7 (2.1)
	Missing	5 (1.3)	10 (3.0)
Student activities at this school offer something for everyone	Totally agree	116 (30.1)	98 (29.3)
	I agree a bit	188 (48.8)	156 (46.7)
	I do not really agree	68 (17.7)	62 (18.6)
	Totally disagree	5 (1.3)	7 (2.1)
	Missing	8 (2.1)	11 (3.3)
Students have a say in decisions affecting them at this school	Totally agree	118 (30.6)	95 (28.4)
	I agree a bit	190 (49.4)	174 (52.1)
	I do not really agree	65 (16.9)	49 (14.7)
	Totally disagree	5 (1.3)	5 (1.5)
	Missing	7 (1.8)	11 (3.3)

TABLE 23 Teacher response rates at baseline and follow-up (continued)

Item	Response	Baseline [n (%)]	Follow-up [n (%)]
International Youth Development Study			
At this school, if students are violent or aggressive on school grounds or at school events, how often are they:			
Issued a written warning	Never	32 (8.3)	22 (6.6)
	Rarely	50 (13.0)	41 (12.3)
	Sometimes	111 (28.8)	97 (29.0)
	Often	83 (21.6)	85 (25.4)
	Always	83 (21.6)	54 (16.2)
	Missing	26 (6.8)	35 (10.5)
Parents/guardians called or contacted	Never	0 (0.0)	0 (0.0)
	Rarely	6 (1.6)	3 (0.9)
	Sometimes	62 (16.1)	50 (15.0)
	Often	161 (41.8)	138 (41.3)
	Always	145 (37.7)	123 (36.8)
	Missing	11 (2.9)	20 (6.0)
Referred to a school counsellor or school nurse	Never	28 (7.3)	15 (4.5)
	Rarely	74 (19.2)	75 (22.5)
	Sometimes	172 (44.7)	140 (41.9)
	Often	54 (14.0)	55 (16.5)
	Always	15 (3.9)	8 (2.4)
	Missing	42 (10.9)	41 (12.3)
Referred to the leadership group (e.g. head of year, assistant head)	Never	1 (0.3)	0 (0.0)
	Rarely	8 (2.1)	5 (1.5)
	Sometimes	60 (15.6)	50 (15.0)
	Often	151 (39.2)	136 (40.7)
	Always	151 (39.2)	120 (35.9)
	Missing	14 (3.6)	23 (6.9)
Referred to participate in a group or programme	Never	25 (6.5)	18 (5.4)
	Rarely	90 (23.4)	95 (28.4)
	Sometimes	173 (44.9)	138 (41.3)
	Often	36 (9.4)	38 (11.4)
	Always	10 (2.6)	3 (0.9)
	Missing	51 (13.2)	42 (12.6)
Encouraged to participate in peer mediation	Never	45 (11.7)	42 (12.6)
	Rarely	107 (27.8)	93 (27.8)
	Sometimes	132 (34.3)	105 (31.4)
	Often	39 (10.1)	41 (12.3)

continued

TABLE 23 Teacher response rates at baseline and follow-up (continued)

Item	Response	Baseline [n (%)]	Follow-up [n (%)]	
Placed in school detention	Always	10 (2.6)	9 (2.7)	
	Missing	52 (13.5)	44 (13.2)	
	Never	12 (3.1)	11 (3.3)	
	Rarely	11 (2.9)	5 (1.5)	
	Sometimes	63 (16.4)	48 (14.4)	
	Often	159 (41.3)	157 (47.0)	
Isolated on their own at school	Always	123 (31.9)	92 (27.5)	
	Missing	17 (4.4)	21 (6.3)	
	Never	13 (3.4)	5 (1.5)	
	Rarely	18 (4.7)	23 (6.9)	
	Sometimes	102 (26.5)	78 (23.4)	
	Often	172 (44.7)	149 (44.6)	
Excluded from school temporarily	Always	62 (16.1)	56 (16.8)	
	Missing	18 (4.7)	23 (6.9)	
	Never	6 (1.6)	2 (0.6)	
	Rarely	39 (10.1)	42 (12.6)	
	Sometimes	186 (48.3)	161 (48.2)	
	Often	108 (28.1)	83 (24.9)	
Excluded from school permanently	Always	32 (8.3)	24 (7.2)	
	Missing	14 (3.6)	22 (6.6)	
	Never	31 (8.1)	24 (7.2)	
	Rarely	207 (53.8)	200 (59.9)	
	Sometimes	96 (24.9)	62 (18.6)	
	Often	20 (5.2)	14 (4.2)	
HSE	Always	10 (2.6)	7 (2.1)	
	Missing	21 (5.5)	27 (8.1)	
	How fair are the rules at this school?	Very fair	188 (48.8)	148 (44.3)
		Quite fair	172 (44.7)	165 (49.4)
		Quite unfair	18 (4.7)	13 (3.9)
		Very unfair	4 (1.0)	0 (0.0)
Missing		3 (0.8)	8 (2.4)	
Do teachers at this school try to make sure that students obey the rules?	All teachers	99 (25.7)	66 (19.8)	
	Most teachers	224 (58.2)	203 (60.8)	
	Some teachers	58 (15.1)	57 (17.1)	
	Very few teachers	1 (0.3)	0 (0.0)	
	Missing	3 (0.8)	8 (2.4)	

TABLE 23 Teacher response rates at baseline and follow-up (continued)

Item	Response	Baseline [n (%)]	Follow-up [n (%)]	
Safety				
Do you feel safe at this school?	All the time	248 (64.4)	75 (22.5)	
	Most of the time	106 (27.5)	182 (54.5)	
	Some of the time	18 (4.7)	68 (20.4)	
	Never	1 (0.3)	1 (0.3)	
	Missing	12 (3.1)	8 (2.4)	
Bullying victimisation (GBS adapted)				
Have you been threatened physically or actually hurt by a student at this school in the last 3 months of school time?	No	353 (91.7)	293 (87.7)	
	Yes	21 (5.5)	31 (9.3)	
	Missing	11 (2.9)	10 (3.0)	
How often?	Once	8 (2.1)	22 (6.6)	
	Occasionally	6 (1.6)	3 (0.9)	
	About once a week	0 (0.0)	1 (0.3)	
	Most days	2 (0.5)	0 (0.0)	
	Missing	369 (95.8)	308 (92.2)	
	Have you had support to deal with this?	No	6 (1.6)	8 (2.4)
		Yes	7 (1.8)	14 (4.2)
Not sure		3 (0.8)	4 (1.2)	
Missing		369 (95.8)	308 (92.2)	
Peter Smith DAPHNE (adapted)				
Have you ever been victimised through mobile phone use or on the internet by a student (e.g. sent abusive text messages or e-mails)?	No	357 (92.7)	299 (89.5)	
	Yes, once or twice	15 (3.9)	13 (3.9)	
	Yes, several times	1 (0.3)	0 (0.0)	
	Missing	12 (3.1)	22 (6.6)	
Do you know any teacher at this school who has ever been victimised through mobile phone use or on the internet by a student?	No	284 (73.8)	243 (72.8)	
	Yes	87 (22.6)	65 (19.5)	
	Missing	14 (3.6)	26 (7.8)	

A decorative graphic consisting of numerous thin, parallel green lines that curve from the left side of the page towards the right, creating a sense of movement and depth.

**EME
HS&DR
HTA
PGfAR
PHR**

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

Published by the NIHR Journals Library