

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



LSHTM Research Online

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

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

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

Usage Guidelines:

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

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

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

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

Emily Ashbrook Warren

Volume 2

Table of Contents

Appendices.....	4
Appendix 1: Locating and testing the healthy context paradox: examples from the INCLUSIVE trial.....	5
Appendix 2: Methodological reflections on using qualitative research to explore the causal mechanisms of complex health interventions	31
Appendix 3: Are randomised controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services	56
Appendix 4: INCLUSIVE process evaluation study protocol.....	68
Appendix 5: INCLUSIVE original study protocol.....	137
Appendix 6: INCLUSIVE updated study protocol	151
Appendix 7: Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial.....	154
Appendix 8: Modifying the secondary school environment to reduce bullying and aggression: the INCLUSIVE cluster RCT	167
Appendix 9: Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention.....	365
Appendix 10: Moderated mediation analyses to assess intervention mechanisms for impacts on victimisation, psycho-social problems and mental wellbeing: Evidence from the INCLUSIVE realist randomized trial	375
Appendix 11: Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al.	383
Appendix 12: Online appendices related to the QCA	388
Appendix 13: Copyrights for included publications	393
Appendix 13.1: Proof of copyright for “Are randomised controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services”	393
Appendix 13.2: Proof of copyright for “Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial”	394
Appendix 13.3: Proof of copyright for “Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): update to a cluster randomised controlled trial protocol”	395
Appendix 13.4: Proof of copyright for “Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial”	396
Appendix 13.5: Proof of copyright for “Modifying the secondary school environment to reduce bullying and aggression: the INCLUSIVE cluster RCT”	397
Appendix 13.6: Proof of copyright for “Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention”	401
Appendix 13.7: Proof of copyright for “Moderated mediation analyses to assess intervention mechanisms for impacts on victimisation, psycho-social problems and mental wellbeing: Evidence from the INCLUSIVE realist randomized trial”	402

Appendix 13.8: Proof of copyright for “Action groups as a participative strategy for leading whole-school health promotion: Results on implementation from the INCLUSIVE trial in English secondary schools” .403
Appendix 13.9: Proof of copyright for “Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention” 409
Appendix 13.10: Proof of copyright for “Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al.” 410

Appendices

Appendix 1: Locating and testing the healthy context paradox: examples from the INCLUSIVE trial

Locating and testing the healthy context paradox: examples from the INCLUSIVE trial

Accepted at BMC Medical Research Methodology

G.J. Melendez-Torres ^{1*}; Emily Warren ²; Obioha C Ukoumunne ³; Russell Viner ⁴;

Chris Bonell ²

¹ Peninsula Technology Assessment Group, University of Exeter Medical School, Exeter, UK

² Department of Public Health, Environments and Society, London School of Hygiene and Tropical Medicine, London, UK

³ NIHR Applied Research Collaboration South West Peninsula (PenARC), University of Exeter Medical School, Exeter, UK.

⁴ Department of Social Science, UCL Institute of Child Health, London, UK

*corresponding author; g.j.melendez-torres@exeter.ac.uk, +44 7847 360259

Abstract

Background. The healthy context paradox, originally described with respect to school-level bullying interventions, refers to the generation of differences in mental wellbeing amongst those who continue to experience bullying even after interventions successfully reduce victimisation. Using data from the INCLUSIVE trial of restorative practice in schools, we relate this paradox to the need to theorise potential harms when developing interventions; formulate the healthy context paradox in a more general form defined by mediational relationships and cluster-level interventions; and propose two statistical models for testing the healthy context paradox informed by multilevel mediation methods, with relevance to structural and individual explanations for this paradox.

Methods. We estimated two multilevel mediation models with bullying victimisation as the mediator and mental wellbeing as the outcome: one with a school-level interaction between intervention assignment and the mediator; and one with a random slope component for the student-level mediator-outcome relationship predicted by school-level assignment. We relate each of these models to contextual or individual-level explanations for the healthy context paradox.

Results. Neither model suggested that the INCLUSIVE trial represented an example of the healthy context paradox. However, each model has different interpretations which relate to a multilevel understanding of the healthy context paradox.

Conclusions. Greater exploration of intervention harms, especially when those accrue to population subgroups, is an essential step in better understanding how interventions work and for whom. Our proposed tests for the presence of a healthy context paradox provide the analytic tools to better understand how to support development and implementation of interventions that work for all groups in a population.

Trial registration. Current Controlled Trials, ISRCTN10751359

Keywords. Cluster-randomised trials, contextual effects, mental wellbeing, bullying, intervention harms

Locating and testing the healthy context paradox: examples from the INCLUSIVE trial

BACKGROUND

Garandeau and Salmivalli[1] recently theorised the existence of a *healthy context paradox*. Using the example of school bullying interventions, they described that interventions that reduce the prevalence of victimisation (and thus improve overall rates of mental wellbeing) may actually worsen the mental wellbeing of those students who continue to experience victimisation during and after the intervention. Anti-bullying interventions may therefore strengthen rather than attenuate differences in mental wellbeing between victimised and non-victimised individuals. In this brief paper, we extend Garandeau and Salmivalli's valuable contribution by: relating their work to our previous discussion of the need to consider dark logic models[2] theorising potential harms when developing interventions; formulating the healthy context paradox in a more general form defined by mediational relationships; reiterating that the healthy context paradox is a phenomenon that can only be detected in cluster-level interventions; proposing two statistical models for testing the healthy context paradox; and relating these statistical models to the meta-mechanisms (contextual and individual) that might be implicated in the healthy context paradox. We demonstrate these points using a mediational model from the INCLUSIVE trial[3], a school-randomised trial of a restorative practice intervention to prevent bullying and improve mental wellbeing amongst secondary school students in southeast England (see Box 1). Throughout this paper, our definition of mediation is classical; that is, a variable that explains part or all of the causal relationship between an independent variable and an outcome.[4] From an interventional perspective, a mediator is a variable on the causal

path between an intervention and an outcome through which a significant indirect effect can be detected.

The healthy context paradox and dark logic models

The healthy context paradox is a welcome contribution to the literature in that it provides intervention developers and implementers with additional insights into how school-based interventions might inadvertently cause harms. In our prior work on dark logic models, we described that harms could take the form of either *paradoxical effects*, wherein an intervention that seeks to improve an outcome in fact worsens it, or *harmful externalities*, wherein an intervention aiming to generate benefits in one domain generates harms in another.[2] As a heuristic, the healthy context paradox provides a way for intervention developers and evaluators to advance and refine intervention theory through understanding how interventions may not equally benefit all students.[2, 5]

In relation to whole-school anti-bullying interventions, the healthy context paradox is an example of a paradoxical effect where the harm does not affect the entire study population (all students) but a subpopulation defined in terms of the intermediate effects (those who are bullied) of the intervention. That is, the harm affects a subpopulation defined in part by the impact of the intervention on the mediator. The healthy context paradox is also an *equity harm*[6], meaning that an intervention that improves health overall may worsen it for some, leading to exacerbations of existing inequalities between groups. For example, the INCLUSIVE theory of change suggested that schools where students know that their teachers are taking action to address bullying, and where being a bully goes against social norms, will have students with better mental wellbeing.[7] However, the healthy context paradox would suggest that students who are victimized may have worse mental health

than before the intervention.[1] The healthy context paradox also suggests that this may be because they have fewer co-bullied peers to relate with and now suffer worse social isolation.

To generalise, implicit in the healthy context paradox is a specific mediational relationship defined by a psychosocial or behavioural mediator and a wellbeing outcome. In the study by Garandeau and Salmivalli[1], bullying victimisation is the mediator, but the mediator could be any similar variable capturing intermediate outcomes, for example bullying perpetration, which is also known to be linked to poor mental wellbeing; other forms of relational aggression, such as sexual harassment or dating and relationship violence; or even variables such as school commitment. The outcome of interest generally relates to mental wellbeing, but could hypothetically relate to any outcome where intervention effects on the outcome are mediated by intervention effects which make a behaviour or other experience less normative within a setting. The healthy context paradox thus relates to harms in wellbeing that affect a subpopulation that does not experience the benefits experienced by the broader study population.

In our original work on dark logic models[2], we proposed that a critical path through which intervention harms might arise is the interaction between the social structure within which an intervention is delivered and the agency of those interacting with the intervention, which may trigger unintended consequences. Such an interaction might also occur within the healthy context paradox. These processes might harm all those who continue to experience victimisation or those with particular vulnerabilities. Dark logic models for future school-based health interventions that seek to address bullying or other critical mediating behaviours should theorise potential adverse

mechanisms that could operate at structural and individual levels and develop ways to measure the interaction between these.

Understanding the healthy context paradox in relation to contextual effects

Implicitly but importantly, the healthy context paradox can be detected only in interventions that are allocated at the cluster level. That is, it is impossible to detect a healthy context paradox in a situation where an intervention is allocated at the individual level. It is important to stress that our focus is on *detection* of the healthy context paradox as opposed to its *generation*. The existence of an intervention-generated equity harm[6] of the type described above, specifically a general improvement in wellbeing arising from reductions in a behavioural or experiential mediator with a worsening in wellbeing for those who still report high levels of the mediator, might manifest but cannot be detected in an individually randomised trial. Consider, for example, if the INCLUSIVE trial had tested an intervention consisting only of individually-administered social-emotional learning without the school components, and was thus amenable of a trial that randomises individuals within schools. Even where this intervention is effective and the prevalence of bullying victimisation sharply declines, and victimisation becomes less normative, remaining victims may, for example, receive less support and experience worse mental health. While this equity harm exists in the same form as above, it cannot be identified as a *contextual* paradox because there is no basis to contrast cluster-level and individual-level impacts. This is despite the fact that in our hypothetical example, the equity harms generated by the intervention clearly worked through contextual, school-based mechanisms related to provision of support.

This conceptual basis for detecting the healthy context paradox in cluster randomised trials can be represented statistically, and these representations form the

basis for proposing tests of the healthy context paradox. The rest of this section focuses on developing these representations using concepts from multilevel models, also known as generalised linear mixed-effects models or hierarchical linear models.[8] Multilevel models are frequently used in the analysis of cluster randomised trials as they can jointly consider the impact of cluster-level variables (such as treatment allocation) and individual-level variables (such as demographic characteristics) on outcomes.[9] In our example, clusters refer to schools in a trial, and individuals refer to students. First, we focus on how multilevel models estimate intervention impacts in cluster randomised trials; second, we consider how multilevel models can be used to identify contextual effects; and third, we reinforce why detection of the healthy context paradox can only occur in multilevel data structures, such as students nested within schools, specifically where interventions are allocated at cluster or school level.

Estimating intervention impacts with multilevel models. When interventions are allocated at the cluster (or school) level, multilevel models use individual students' reports of the study's outcome to estimate differences between intervention and control schools through school-level means of those student reports.[10] Understood statistically, the healthy context paradox exists in the contradiction between school-level differences (between intervention and control groups) and student-level impact of an intervention (which may be more heterogeneous than a school mean can represent). Put otherwise, even if school-level means suggest that a school *on average* has experienced an improvement on wellbeing, it is possible that a minority of individual students within intervention schools experienced comparative worsening in their mental health, and that this worsening can be related to a specific individual characteristic or vulnerability. This heterogeneity in intervention effect forms the basis of the healthy context paradox.

From cluster-level predictors to contextual predictors. This difference between school means and student impacts is an important first step in developing a statistical representation of the healthy context paradox. The next step is to understand how continuous predictors measured at the student level, such as mediators, can create both student effects and contextual school effects. In a multilevel modelling context, Raudenbush and Bryk[8] describe the contextual effect as the difference between the within-cluster coefficient (relationship of an individual value of a predictor and an individual value of an outcome within a cluster) and the between-cluster coefficient (relationship between cluster-level mean of a predictor and cluster-level mean of an outcome) for a predictor with an outcome. While contextual effects exist anywhere individuals are grouped in clusters (i.e. students grouped in schools), we can also describe contextual effects as follows: is there an impact of a school-mean predictor on an outcome above and beyond the student-level relationship between predictor and outcome? Their classic example[8, 11] relates socioeconomic position to performance on standardised tests. Socioeconomic position can be measured at the student level, with individual students' reports, and also at the school level, with the average of students' reports. To restate this example as a question: is there an impact on students' test scores, above and beyond the student-level associations between socioeconomic position and test scores, of studying in a school with low average socioeconomic position?

Contextual effects can only be detected where predictors are measured at the individual level and can be aggregated to cluster-level means, leading to simultaneous testing of both the association between the variable measured at the individual level and the outcome and the association between the variable aggregated to the cluster level and the outcome. As a result, contextual effects are not relevant for intervention

allocation status as that is only a cluster-level variable, but rather contextual effects are relevant for understanding the association between mediators and outcomes.

Detection of the healthy context paradox in multilevel data structures. To be clear, this is not to say that the healthy context paradox reduces to testing a contextual effect, as we discuss below. Instead, a necessary precondition to understanding the healthy context paradox is to parse individual-level and cluster-level variation in the relationship between a predictor and an outcome and therefore identify a contextual effect. This is because, consistent with the range of mechanisms identified by Garandeanu and Salmivalli[1], the intervention could potentially influence a) the school context within which the link between bullying and mental wellbeing occurs or b) students' experiences of how bullying links to mental wellbeing even where school-level contexts have improved. For example, INCLUSIVE may have changed school culture or it may have worsened a student's bullying even when bullying has become less prevalent. Thus, not only conceptually but statistically, because contextual effects can only be directly measured in multilevel data structures, the healthy context paradox can only be detected in cluster-allocated, or school-allocated, interventions.

METHODS

Testing for the healthy context paradox

Putting this all together, any test of the healthy context paradox requires the decomposition of the intervention's mediational pathways into mediational pathways at both school and student levels. The insight of Raudenbush and Bryk[8] about the estimation of contextual effects has been influential in understanding mediation in cluster randomised trials. The healthy context paradox corresponds to a specific type of mediation known as cross-level mediation, and specifically 2-1-1 mediation.[12] The 2-1-1 mediation model exists when an intervention is allocated at level 2, or at the school

level; influences a level 1, or student, mediator (e.g. bullying); and also influences a level 1 outcome (e.g. mental wellbeing). Of note is that both the mediator and the outcome can be measured at the student level and aggregated at the school level to generate school-level means. Pituch and Stapleton[11] have observed that specific approaches to testing cross-level mediation generate greater power and greater insights in distinguishing between the impact of the intervention on the individual and cluster levels; that is, a school-randomised intervention may effect a specific mediational pathway both through the student-level relationship between mediator and outcome and through a school-level contextual effect which modifies the nature of this student-level relationship. To parse these relationships, they suggest testing a mediational model with the school-level paths between intervention and school-level mediator mean, intervention and school-level outcome mean, school-level mediator mean and school-level outcome mean; and with the student-level path between the mediator and the outcome (see Figure 1). In this formulation, the school-level mediator-outcome relationship is therefore the contextual effect (or the effect of the school level of the mediator), the student-level mediator-outcome relationship is the individual effect (or how a student's report of the mediator links to a student's report of the outcome), and the sum of the school-level and student-level mediator-outcome coefficients multiplied by the coefficient relating mediator and intervention is the total indirect effect.

Importantly, in the first instance, this requires including an uncentred (i.e. at its original value) mediator at the student level alongside a school-level mean for the mediator.

Drawing on this 2-1-1 mediation model, we propose that what Garandeanu and Salmivalli[1] describe as a moderated mediation model is better described as a special case of mediation where the intervention through its school-level effects moderates the relationship between mediator and outcome. This is because moderated mediation is

most generally understood as a situation where a fourth variable explains heterogeneity across the trial sample in the magnitude of the indirect effect.[13] However, if the healthy context paradox is an equity harm generated by a cluster-level intervention, then the intervention itself cannot be that fourth variable and cannot moderate the link between intervention and mediator. The intervention can, however, moderate the link between mediator and outcome. This will be familiar to those approaching mediation from the potential outcomes framework as a treatment-by-mediator interaction.[14]

If the general form of the healthy context paradox is treatment-by-mediator interaction and the mediator-outcome relationship can be measured at both individual and cluster levels in cluster-randomised trials, it follows that there are two potential treatment-by-mediator interactions to be estimated: one at the cluster level, in which the contextual effects are moderated by intervention; and one at the individual level, in which individual effects are moderated by intervention. We propose tests of each of these below as Test 1 and Test 2 respectively, and provide indicative code for implementation in Mplus (see Appendix 1). Both of these treatment-by-mediator interactions provide a test for the existence of the healthy context paradox, and are suggestive of different possible meta-mechanisms for the paradox's existence in a given trial.

The mediation models we develop draw on two key study outcomes: bullying victimisation assessed using the Gatehouse Bullying Scale[15], at 36-month follow-up, and a measure of functional and psychological mental wellbeing, the Short Warwick-Edinburgh Mental Wellbeing Scale or SWEMWBS[16], at 36-month follow-up. Higher scores on the Gatehouse Bullying Scale represent higher levels of victimisation, while higher scores on the SWEMWBS represent higher levels of mental wellbeing. We restrict our consideration here to mediators and outcomes as measured on linear

scales; estimation of direct and indirect effects is more complicated where either mediator or outcome requires a different, non-normal link function.[17]

A baseline mediational model from INCLUSIVE

To estimate this model, we use the 2-1-1 model described above, including regressing the school-level mediator mean on intervention status, the school-level outcome mean on the school-level mediator mean and intervention status, and the student-level outcome on the student-level mediator. This accomplishes the separation of contextual and individual effects in the mediator-outcome relationship.

We note at this point that a non-significant relationship between mediator and outcome at either school or student level should not preclude undertaking either Test 1 or Test 2, as, for example, the average of two effects could produce a misleading null effect overall. However, a non-significant path between intervention and mediator suggests that the candidate mediator should not be considered further.

Test 1: contextual effects

To estimate this model, the relationship between mediator and outcome at the school level is moderated by intervention status. This is an extension to the standard structural equation model-based mediation method, where the interaction between intervention allocation and the mediator score is entered as an additional predictor of the outcome.[14] Thus, the findings from this model will generate different estimates of the contextual effect between mediator and outcome. A standard significance test can be used on the interaction term to test for differential contextual effects on the intervention arising from a moderated relationship between mediator and outcome.

Where a healthy context paradox is present at a contextual level, the results of the test will indicate that a contextual effect for the mediator-outcome relationship has a magnitude indicating less benefit (or greater harm) in the intervention as compared to

the control group, even where the intervention-mediator relationship suggests a meaningful and positive impact of the intervention. The interpretation of this is that the intervention may have triggered structural mechanisms that are linked to a worsening of school context for those who experience bullying; but also for those who do not. This is because, in this circumstance, the intervention reduces levels of bullying at the school level; may still improve mental wellbeing overall at school level, including through direct effect on the outcome; but potentiates a worsening school-level link between bullying and mental health, so that intervention schools with higher levels of bullying experience an even greater negative contextual impact on average levels of mental wellbeing. This may be enough to outweigh positive benefits from the intervention at individual and contextual levels, because intervention schools with higher levels of bullying have an even larger association with worsening school mental health than control schools.

Test 2: individual effects

To estimate this model, the relationship between mediator and outcome at student level is moderated by intervention status. This is a standard random slope model where the student-level relationship is moderated by a school-level variable, here intervention status.[11] A direct test of significance is usually available for this relationship. However, a complication of this model is that to estimate this relationship without bias, the individual-level mediator must be centred within schools.[18] This means that the school-level relationship between mediator and outcome is no longer the contextual effect alone but rather the sum of the contextual and individual effects.[11] While this is not a barrier to testing, it should be borne in mind in interpretation, as in this model the value of the school-level relationship between mediator and outcome may be closer to the sum of the student-level and school-level

paths in the baseline mediation model. As with most random slope models, it can be useful to co-vary the random slope component with the intercept for the dependent variable.

Analogous to above, where a healthy context paradox is present at the student level, the results of the test will indicate that the student-level relationship between mediator and outcome has a magnitude indicating less benefit (or greater harm) in the intervention as compared to the control group, even where the intervention-mediator relationship is significant. The interpretation of this is that even as the intervention improves scores on the mediator and on the outcome overall for students, those students experiencing worse values for the mediator (e.g. bullying) also experience proportionally worse and more inequitable values for the outcome (e.g. mental wellbeing).

RESULTS

Baseline mediation model

Our baseline model (see Figure 1) showed that the impact of the intervention on mental wellbeing was mediated by improvements in bullying victimisation. However, these improvements were mediated at the student level ($\beta = -0.687$, $SE = 0.060$) without a significant school-level contextual effect. That is, the school-level path from victimisation to mental wellbeing ($\beta = 0.340$, $SE = 0.454$) was not significant. Because a mediation pathway was also included at the student level, the school-level path represents the contextual effect of the mediator. The interpretation of this model is that part of INCLUSIVE's beneficial effect on mental wellbeing was through reducing victimisation; but that the link between victimisation and mental wellbeing can be understood in this baseline model at the student level (students with lower victimisation on average had better mental wellbeing) without a contextual effect at the

school level (beyond the student-level relationship, schools with lower victimisation did not have students with better mental wellbeing on average).

Test 1 in INCLUSIVE

We examined the interaction between the school-level mean of the mediator, bullying victimisation, and intervention allocation status and entered this as a third predictor of the outcome. The function of this predictor, as discussed above, is to induce a different value of the contextual effect of the mediator on the outcome depending on intervention status (i.e., a different value for each of the intervention and control arms).

As in the baseline model, there is a significant and meaningful mediational pathway from intervention to outcome (see Figure 2). However, the test for contextual effects in the healthy context paradox suggests that this paradox is not supported in INCLUSIVE, and a Wald test did not find that this model was significantly different from the baseline model ($df=1, p=0.20$). Neither the interaction of intervention with mediator ($\beta=1.089, SE=0.847$) nor the direct path at the between level from mediator to outcome ($\beta=-0.060, SE=0.527$) were significant. Indeed, the intervention by mediator interaction, while non-significant, would be interpreted as having an opposite effect; namely an important enhanced effect on mental wellbeing of the intervention in schools experiencing residually higher rates of bullying victimisation. In short, applying test 1 to data from INCLUSIVE does not suggest that the intervention worsened the link at school level from bullying victimisation to decreased mental wellbeing. Were this to have been the case, we would have expected the intervention by mediator interaction to have a significant effect with a negative sign.

Test 2 in INCLUSIVE

In addition to the 'baseline' mediation model, we: a) cluster mean-centred the mediator (that is, redefined student-level scores on bullying victimisation as deviations

from the school-level mean), b) specified a random slope component for the relationship between student-level mediator and the outcome, and c) regressed this random slope component on intervention status at the school level. The function of point c) is to determine a different value of the individual-level relationship between the mediator and the outcome depending on intervention status.

Again, a significant and meaningful mediational pathway from intervention status to mental wellbeing persists (Figure 3). However, in this analysis, the student-level relationship between bullying victimisation and mental wellbeing is regressed on intervention status. Thus, the baseline estimate of the relationship at the student level between the mediator and the outcome ($\beta=-0.634$, $SE=0.082$) properly refers to the mediator-outcome relationship in students in control schools. The regression of student-level slope on intervention thus yields the difference between intervention and control groups on the relationship between student-level mediator and outcome ($\beta=-0.148$, $SE=0.117$). The interpretation of this coefficient is that in intervention schools, the relationship between bullying victimisation and mental wellbeing is stronger; that is, students experiencing victimisation experience an even greater decrement in mental wellbeing, consistent with the healthy context paradox. However, this path is not significant and thus the model does not support the existence of the healthy context paradox at individual level. A Wald test did suggest that this model was significantly different from a baseline model ($df=3$, $p=0.03$); however, this was due to a significant random slope component for the student-level mediator-outcome relationship. A Wald test comparing this model to a model with no relationship between student-level slope and intervention status and with no covariance between slope and random intercept did not support the existence of the healthy context paradox ($df=2$ $p=0.29$).

DISCUSSION

Of the two proposed tests, Test 2 is probably closest to how Garandean and Salmivalli understood the healthy context paradox. However, we believe that Test 1 is important as well. This is because it sheds light on potential school-level structural explanations for the healthy context paradox—structural explanations that form an important part of the theoretical basis for this paradox—and can account for exacerbation in differences *between intervention and control schools* in the mental health of those experiencing victimisation. In contrast, Test 2 sheds light on exacerbation in differences *within intervention schools* in the mental health of those experiencing victimisation. To the extent that understanding mechanisms in interventions is an inductive task, the results of each test provide analytic purchase in inferring the mechanisms for evidenced health inequalities both between schools and within school, suggesting that these are either primarily contextual or primarily individual, or both. These two tests thus relate to two different meta-mechanisms, structural and individual, that can drive the existence of the healthy context paradox. Given the increasing focus on complex systems approaches to evaluation[19], understanding how interventions work over multiple systems of influence can help in developing intervention theory. The healthy context paradox may also be useful in other areas of public health that seek to reduce the frequency or prevalence of specific population characteristics or behaviours, thus stigmatising those who are ‘left behind’ by the intervention. For example, the healthy context paradox could be tested for interventions targeting diet and physical activity, where interventions that stigmatise overweight can worsen contextual or individual relationships between overweight and mental wellbeing; or interventions that seek to reduce sexual risk, thus stigmatising those who continue to engage in risk behaviours.

CONCLUSIONS

We are grateful to Garandean and Salmivalli for this important contribution to the understanding of how school-level interventions may not equally benefit all students. Greater exploration of intervention harms, especially when those accrue to population subgroups, is an essential step in better understanding how interventions work and for whom, and thus in supporting the decision to implement interventions in contexts different from the ones where interventions may have been originally evaluated.[20] Our proposed tests for the presence of a healthy context paradox provide the analytic tools to better understand how to make school contexts effective places for all children and young people to reach their full potential.

LIST OF ABBREVIATIONS

SE Standard error

SWEMWBS Short Warwick-Edinburgh Mental Wellbeing Scale

DECLARATIONS

Ethics approval. Ethics approval for the original study was provided by the University College London ethics committee (ref 5248/001). This study was performed in accordance with relevant guidelines and regulations (i.e. Declaration of Helsinki). Written, informed consent was obtained from head teachers for random allocation and intervention, and from individual students, staff, and intervention facilitators for data collection.

Consent for publication. Not applicable.

Availability of data and materials. The dataset analysed in this study relates to the INCLUSIVE trial. It is not publicly available as it contains sensitive information about trial participants who were below the age of majority at the time of the intervention. However, it is available on request with evidence of ethics approval and analysis protocol.

Competing interests. The authors have no competing interests to declare.

Funding. This study was funded by the National Institute for Health Research in England under its Public Health Research Board (12/153/60) and the Education Endowment Foundation. Obioha Ukoumunne is funded and G.J. Melendez-Torres is part-supported by the National Institute for Health Research (NIHR) Applied Research Collaboration South West Peninsula (PenARC) at the Royal Devon and Exeter NHS Foundation Trust. The views expressed in this publication are those of the authors and do not necessarily reflect those of the UK NHS, the National Institute for Health Research, or the Department of Health for England.

Authors' contributions. GJMT developed the analysis and led the drafting of the manuscript. EW, RV and CB were investigators on the original INCLUSIVE trial and supported development of the analysis and drafting of the manuscript. OCU supported development of the analysis and drafting of the manuscript. All authors read and approved the final manuscript.

Acknowledgements. Not applicable.

Box 1. The INCLUSIVE trial

We use data from INCLUSIVE[3], a school randomised trial of restorative practice in schools involving 40 schools (n=6667 at baseline, n=5960 at 36-month follow-up) serving students age 11-16 in south-east England from 2014 to 2017. Overall, the intervention, which comprised restorative practice, student participation in school decisions and a student social-emotional learning curriculum, was found to reduce student-reported bullying victimisation and improve mental wellbeing as well as benefit other secondary outcomes at 36-month follow-up for children aged 14-15 years. Full details of methods and overall results are presented elsewhere.[3] When we refer to the INCLUSIVE trial in terms of mediation, we use bullying victimisation as the mediator and mental wellbeing as the outcome.

Figure 1. Baseline mediation model

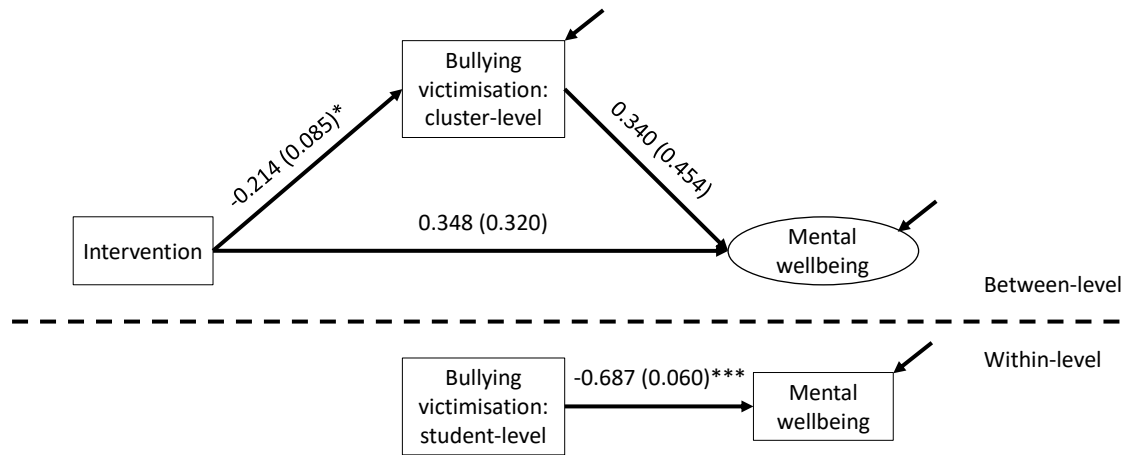


Figure 1. Baseline mediation model

Figure 2. Testing for contextual effects

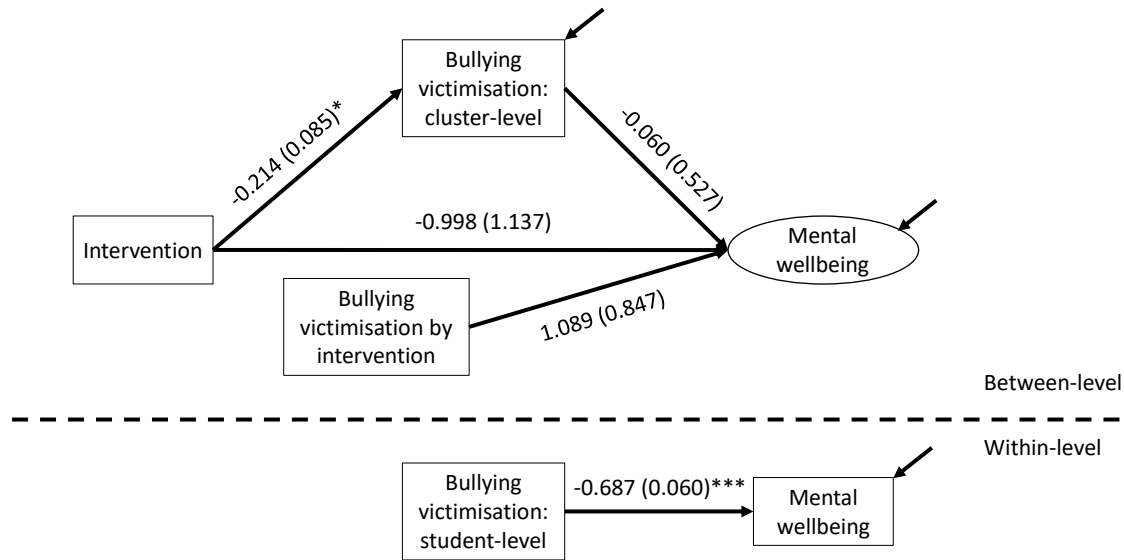


Figure 2. Testing for contextual effects

Figure 3. Testing for individual effects

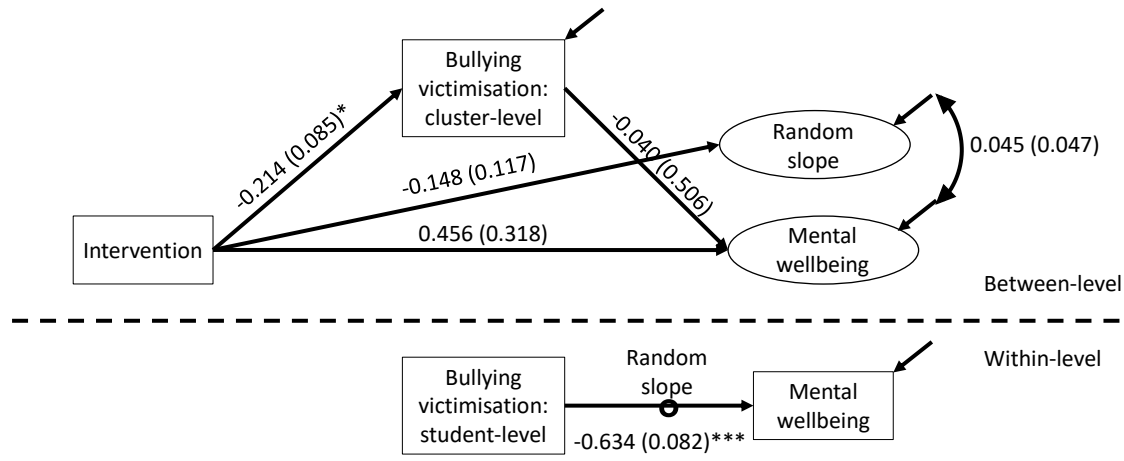


Figure 3. Testing for individual effects

References

1. Garandeau, C.F. and C. Salmivalli, *Can Healthier Contexts Be Harmful? A New Perspective on the Plight of Victims of Bullying*. *Child Development Perspectives*, 2019. **13**(3): p. 147-152.
2. Bonell, C., et al., *'Dark logic': theorising the harmful consequences of public health interventions*. *Journal of Epidemiology and Community Health*, 2015. **69**(1): p. 95-98.
3. Bonell, C., et al., *Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial*. *The Lancet*, 2018. **392**(10163): p. 2452-2464.
4. Baron, R.M. and D.A. Kenny, *The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations*. *Journal of Personality and Social Psychology*, 1986. **51**(6): p. 1173-1182.
5. Bonell, C., G.J. Melendez-Torres, and S. Quilley, *The potential role for sociologists in designing RCTs and of RCTs in refining sociological theory: A commentary on Deaton and Cartwright*. *Social Science & Medicine*, 2018. **210**: p. 29-31.
6. Lorenc, T. and K. Oliver, *Adverse effects of public health interventions: a conceptual framework*. *Journal of Epidemiology and Community Health*, 2014. **68**(3): p. 288.
7. Warren, E., et al., *Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention*. *Trials*, 2020. **21**(774).
8. Raudenbush, S.W. and A.S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Advanced Quantitative Techniques in the Social Sciences. 2002: SAGE Publications.
9. Eldridge, S.M., et al., *Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care*. *Clinical Trials*, 2004. **1**(1): p. 80-90.
10. Hayes, R.J. and L.H. Moulton, *Cluster Randomised Trials*. 2nd ed. Chapman & Hall/CRC Biostatistics Series. 2017: CRC Press.
11. Pituch, K.A. and L.M. Stapleton, *Distinguishing Between Cross- and Cluster-Level Mediation Processes in the Cluster Randomized Trial*. *Sociological Methods & Research*, 2012. **41**(4): p. 630-670.
12. Preacher, K.J., M.J. Zyphur, and Z. Zhang, *A general multilevel SEM framework for assessing multilevel mediation*. *Psychological Methods*, 2010. **15**(3): p. 209-233.
13. Muller, D., C.M. Judd, and V.Y. Yzerbyt, *When moderation is mediated and mediation is moderated*. *Journal of Personality and Social Psychology*, 2005. **89**(6): p. 852-863.
14. MacKinnon, D.P., M.J. Valente, and O. Gonzalez, *The Correspondence Between Causal and Traditional Mediation Analysis: the Link Is the Mediator by Treatment Interaction*. *Prevention Science*, 2020. **21**(2): p. 147-157.
15. Bond, L., et al., *A comparison of the Gatehouse Bullying Scale and the Peer Relations Questionnaire for students in secondary school*. *Journal of School Health*, 2007. **77**(2): p. 75-79.
16. Melendez-Torres, G.J., et al., *Measurement invariance properties and external construct validity of the short Warwick-Edinburgh mental wellbeing scale in a large national sample of secondary school students in Wales*. *Health and Quality of Life Outcomes*, 2019. **17**(139).
17. Albert, J.M., et al., *Generalized causal mediation and path analysis: Extensions and practical considerations*. *Statistical Methods in Medical Research*, 2018. **28**(6): p. 1793-1807.
18. Enders, C.K. and D. Tofighi, *Centering predictor variables in cross-sectional multilevel models: A new look at an old issue*. *Psychological Methods*, 2007. **12**(2): p. 121-138.

19. Moore, G.F., et al., *From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation*. *Evaluation*, 2019. **25**(1): p. 23-45.
20. Bonell, C., et al., *Will it work here? A realist approach to local decisions about implementing interventions evaluated as effective elsewhere*. *Journal of Epidemiology and Community Health*, 2020. **Ahead of print**.

Appendix 2: Methodological reflections on using qualitative research to explore the causal mechanisms of complex health interventions

Accepted at Evaluation

Chris Bonell ¹; Emily Warren ¹; G.J. Melendez-Torres ^{2*};

¹ Department of Public Health, Environments and Society, London School of Hygiene and Tropical Medicine, London, UK

² Peninsula Technology Assessment Group, University of Exeter Medical School, Exeter, UK

*corresponding author: Chris.bonell@lshtm.ac.uk

Abstract

We reflect on how qualitative research can be used to develop or refine theories about how the mechanisms triggered by intervention enactment might generate outcomes, referring to examples from a 'realist trial' of a whole-school health intervention. Qualitative research can explore mechanisms directly, by asking participants how they think interventions work, or indirectly, by exploring participant experiences of intervention-related actions to understand the conditions and consequences of these actions. Both of these approaches can inform theorisation of how mechanisms are triggered and generate outcomes, and how this is contingent on context. We discuss methods for sampling, data collection and data analysis, and recommend dimensional analysis as a means to analyse qualitative data on mechanisms. We then consider how to draw on qualitative research to inform hypotheses to be tested statistically.

Key words

Evaluation; qualitative; realist; mechanisms; complex interventions

Introduction

In this paper, we consider how qualitative research can be used to inform theorising of the mechanisms which enactment of complex interventions trigger and which generate outcomes. Our reflections are illustrated with examples from our 'realist trial' of a whole-school health intervention. Our ideas about what mechanisms are, and how they should be examined, are informed by realist concepts. Nonetheless, our aim is to provide useful suggestions for those evaluating complex interventions, regardless of the model of evaluation to which they subscribe.

Complex interventions are commonly described as social interventions with components that interact with each other and with the context in which they are enacted (Moore et al., 2014). There is increased interest in evaluating complex interventions not only by quantifying their effects or describing their implementation, but also by understanding the underlying mechanisms by which they generate outcomes (Bonell et al., 2012; Moore et al., 2014; Burchett et al., 2020). Understanding such mechanisms can help us understand how interventions work. When linked with an understanding of how these mechanisms are contingent on context, as realist approaches to evaluation aim to do, this can also help us understand for whom and where interventions work (Pawson and Tilley, 1997). This can help inform assessments of potential transferability (Burchett et al., 2020), intervention refinement (Bonell et al., 2021), and broader scientific understanding and future interventions (Davey et al., 2019). Such approaches can be broadly described as theory-driven evaluation because they aim to develop and refine theory as to how interventions work (Marchal et al., 2012).

A substantial literature considers what mechanisms are (Lemire et al., 2020) in terms of their ontological features and the epistemological status of knowledge we have about them. Informed by this literature, we see mechanisms as new human responses, actions and interactions triggered by the provision of new economic, informational or other

resources, and by the resultant enactment of intervention activities (May, 2013; Pawson and Tilley, 1997; Lemire et al., 2020). Mechanisms consist of “underlying entities, processes, or structures which operate in particular contexts to generate outcomes of interest” (Astbury and Leeuw, 2010) (p.368). Mechanisms might operate at the individual, group, institutional or societal level (Marchal et al., 2012). Mechanisms can involve intra-individual changes in cognitions or emotions generating changes in behaviour (Carey et al., 2019) or changes to inter-individual interactions (Giddens, 1984). Mechanisms have been categorised as involving: macrosocial influences on microsocial outcomes (e.g. media influencing attitudes); microsocial influences on microsocial outcomes (e.g. peers educating peers); or microsocial influences on macrosocial outcomes (e.g. communities lobbying politicians) (Hedstrom and Swedberg, 1998).

Realist evaluation literature suggests that mechanisms are not directly observable, are contingent on local context and generate outcomes. They are triggered by, but are distinct from, intervention activities. They are not reducible to the variables used to assess them or the lines used to suggest them in logic model diagrams (Astbury and Leeuw, 2010). Interventions may trigger one or more different mechanisms. Whereas intervention activities are the ‘form’ of intervention (events), mechanisms are the ‘function’ of these interventions (the generative processes triggered by them) and outcomes are the events thus generated (Lacouture et al., 2015; Hawe et al., 2004). Realists define context in terms of the individual actor capacities and relationships, institutional setting and wider social structures which precede intervention, and which interact with the mechanisms triggered to generate outcomes (Lemire et al., 2020).

Some authors have noted confusion in the literature surrounding the distinctions between, on the one hand, intervention activities and mechanisms, and, on the other, between mechanisms and context (Lemire et al., 2020; Barnes et al., 2003). Our own view is that, while intervention activities and mechanisms are ontologically distinct (as indicated above), mechanisms and context are not ontologically distinct phenomena (contexts will include pre-existing mechanisms which may then interact with

mechanisms triggered by intervention activities) but that nonetheless using these distinct terms is useful in framing propositions about how interventions work and collecting data to examine these propositions.

In traditional trials and other quantitative evaluations of complex health interventions, mechanisms are indirectly assessed via moderation (e.g. assessing if intervention effects differ for different populations) and mediation (e.g. assessing if intervention effects on outcomes are explained by their effects on intermediate measures) analyses (Gardner et al., 2006). In such studies, interventions are increasingly informed by theories of change but these are of variable sophistication and plausibility, often consisting of little more than strings of variables with no consideration of actual mechanisms or how these might interact with context to generate outcomes (Breuer et al., 2016; Moore and Evans, 2017).

In contrast, realist evaluators use context-mechanism-outcome (CMO) configurations which postulate what mechanisms might be triggered by use of intervention resources and how these mechanisms might interact with context to generate outcomes. These CMO configurations are tested and refined through iterative analyses drawing on qualitative and/or quantitative data. This approach has a number of advantages: it provides a theorisation of mechanisms which goes beyond strings of variables to explore deeper processes; it distinguishes between the reality of how mechanisms operate and the data used to examine these; and it considers how mechanisms might operate variously to generate different outcomes in different contexts. Some realist or critical realist evaluators have modified the configurations so that these incorporate agency, or focus on how interventions and contexts interact to trigger mechanisms which generate outcomes (Porter, 2015; Lemire et al., 2020). As suggested above, we prefer to use CMO configurations to frame propositions about how mechanisms operate while recognising the importance of agency and that there is no sharp ontological distinction between context and mechanism.

A key argument of this paper is that it is important to use qualitative research to refine CMO configurations or other forms of theories of change before subjecting them to quantitative analysis. Initial theories informed by existing academic theory may be sharply at odds with how those delivering or receiving the intervention perceive interventions as working. Qualitative evidence may propose important refinements or augmentations to our theories about how mechanisms might generate outcomes. Therefore, initial CMO configurations or other theories may need to be revised in the light of analysis of data on the lived experiences of those involved in implementation and receipt (van Urk et al., 2016). There is therefore an important role for qualitative research in refining our theories about mechanisms before these are quantitatively tested (Unrau, 2001). There is also an important role for qualitative research in examining mechanisms that are too complex to be amenable to typical measurement approaches in quantitative evaluation, such as those including complex chains of causation or feedback loops (Cohn et al., 2013).

While the above arguments for qualitative research on mechanisms are widely recognised in the literature, there is little detailed guidance on conducting qualitative research on mechanisms. The Medical Research Council (MRC) frameworks for complex interventions and for process evaluation have suggested some general principles about the importance of process evaluation examining mechanisms (Moore et al., 2014; Skivington et al., 2021). The process evaluation framework suggests, for example, that process evaluations use mixed methods to examine mechanisms, using qualitative research to explore mechanisms that are unanticipated and/or too complex to be captured quantitatively. However, the framework does not offer specific guidance on how qualitative data might be sampled, collected or analysed in order to do this.

A recent paper by Thirsk and Clark has argued for the importance of hermeneutics-oriented qualitative research on mechanisms (Thirsk and Clark, 2017). Thirsk and Clarke argue that a hermeneutic approach can focus on the reality of mechanisms, and not merely the subjective meanings attributed to these by participants. They also suggest that analysis of qualitative data on mechanisms can be informed by researchers'

pre-understandings of a phenomenon while still leaving the research open to new findings:

Previous experience and understanding of a topic does not prevent a researcher from being open to new understanding of the topic but is an asset that enables the researcher to be better prepared for understanding.” (Thirsk and Clark, 2017)(p.5)

There have also been some useful suggestions from realist evaluators on using qualitative research to examine the mechanisms which interventions trigger (Manzano, 2016; Pawson, 1996). Pawson, for example, has offered a number of suggestions about how to structure interviews so that these can be drawn on to assess the validity of CMO configurations (Pawson, 1996). But while useful, this existing literature does not aim to offer comprehensive guidance on the different ways in which qualitative research might examine mechanisms, on how to conduct qualitative research to explore mechanisms, or on how qualitative research might inform testable hypotheses.

Drawing on examples of qualitative research conducted within a ‘realist trial’ to examine the mechanisms of a whole-school intervention to prevent bullying and improve student mental and physical wellbeing, we aim to reflect methodologically on: how qualitative research can contribute to understanding mechanisms; how we should decide which participants can provide authoritative data on mechanisms and construct samples of these; how we should analyse transcripts to develop ideas about mechanisms; and how we should draw on these analyses to develop testable hypotheses. As suggested above, our aim is to make useful suggestions for conducting qualitative research on mechanisms which are informed by our own realist evaluation but which are also useful to those who subscribe to different models of evaluation.

Case study

Learning Together is a whole-school intervention that aims to enable prevention of bullying and improve student mental and physical health. It provides various intervention resources (such as manuals, curriculum materials, training) to facilitate students and staff in secondary schools enacting 'action groups' which review local survey data on student views and needs, and then decide locally appropriate actions, supported by an external facilitator. Actions can involve reviews to policies as well as changes to school management and environment. Additionally, staff are trained in using restorative practice to build strong relationships between and among staff and students, and to resolve conflicts by exploring different perspectives, identifying harms and repairing relationships. Students are also taught a social and emotional skills curriculum.

Learning Together was informed by an a priori theory of change built on the theory of human functioning and school organisation (Markham and Aveyard, 2003), as well as some initial CMO configurations which proposed how mechanisms which the interventions triggered might interact with context to generate outcomes. The theory of change proposed that student involvement in risk behaviours may be reduced by promoting student commitment to their school's 'instructional' (teaching and learning) and 'regulatory' (discipline and community) orders. This in turn requires that schools 'reframe' provision to focus on student needs and erode 'boundaries' between: staff and students; academic and broader personal development; and the culture of the school and its local community (Markham and Aveyard, 2003). Initial CMO configurations proposed, for example, that these mechanisms would be more likely to generate beneficial outcomes for socio-economically disadvantaged students, for whom commitment to school is less likely to be the socialised norm and whom schools are more likely to engage by the above reframing and boundary-eroding mechanisms.

Learning Together was evaluated using an explicitly realist cluster-randomised controlled trial with 3-year follow across 40 English secondary schools. The overall trial analyses reported that the intervention was associated with reduced bullying victimisation, smoking, alcohol use, smoking tobacco and police involvement, and improved mental wellbeing, psychological functioning and quality of life (Bonell et al.,

2018). A process evaluation quantified fidelity, reach and acceptability using observations of training, action groups and other activities, checklists completed by staff delivering the curriculum action groups and restorative practice, and staff and student questionnaires. The process evaluation also used semi-structured observations, interviews and focus groups to collect qualitative data on observed and reported processes of implementation and receipt of intervention activities, and the consequences of these for staff and students (Warren E, 2019). The realist trial used this qualitative research to inform refinement and augmentation of the initial CMO configurations (Warren et al., 2020). These were then examined using qualitative comparative analysis (QCA) as well as moderation, mediation and moderated-mediation regression analyses (Melendez-Torres et al., 2021). Since this paper is a methodological reflection rather than an empirical report, interested readers should seek further methodological details in the above cited papers.

How qualitative research can explore mechanisms

Qualitative research can inform development, refinement or augmentation of theory relating to mechanisms. This can occur via two approaches, which are not mutually exclusive. Firstly, qualitative research can directly explore the accounts of providers, recipients and other participants about how they think interventions work. This approach in effect develops 'second-order' constructs describing causal mechanisms by interpreting and critically weighing participants' 'first-order' constructs of these (Schütz, 1962). Realist evaluators have pioneered this useful approach, usually exploring participants' views on a priori CMO configurations, to validate, refine or falsify these (Manzano, 2016; Pawson, 1996). As Pawson has argued:

[T]he researcher's theory is the subject matter of the interview, and the subject is there to confirm or falsify and, above all, to refine that theory. (Pawson, 1996) (p.299)

Pawson suggested that the interview should first 'teach' the participant about the possible theories before 'learning' from the participant which of these theories align with their experiences and how they might be refined:

The subject's task is to agree, disagree and to categorize themselves in relation to the attitudinal patterns as constructed in [the researcher's] questions but also to refine their conceptual basis. It is at this point that mutual knowledge is really achieved. The subject is saying in effect 'this is how you have depicted the potential structure of my thinking, but in my experience it happened like this... (p. 306)

In evaluating Learning Together, we used such techniques to ask participants to consider our a priori CMO configurations and articulate their own theories. For example, one student responded to such enquiries by describing how he theorised restorative practice could resolve conflicts:

I just thought [restorative practice] was a brilliant idea because it's showing the younger kids how to be mature about difficult situations and teaching them how to deal with it. And rather than just getting angry, sitting down and talking through things is a better solution. And it's just showing them that. (Focus group with year 9 students, Meadowood School)

We did not conclude that such participants' theories were straightforwardly true representations of mechanisms. Instead, we iteratively compared and contrasted such accounts with other qualitative data to help refine our CMO configurations.

In engaging in such comparisons, we often drew on qualitative data in a second, less direct way. We examined participants' accounts of their experiences of enacting intervention-related activities, the conditions within which this occurred, and what actions or other consequences flowed from these. We used this to refine our theories as to how mechanisms were triggered and how these interacted with context to generate outcomes. Any single participant might only be able to discuss the actions involve in one sub-section of a causal mechanism (Pawson, 1996; Giddens, 1984). However, as

researchers, we could draw on multiple accounts, as well as our own observations, to develop a more authoritative and comprehensive understanding of mechanisms. This drawing from multiple sources to iteratively develop, augment or refine theorised mechanisms has been described in previous empirical studies (Unrau, 2001).

There may be a quasi-quantitative aspect to such work, looking for regularities in what conditions seem to be associated with accounts of certain actions or what reported consequences seem to be associated with particular actions. But the exploration should go beyond this 'black box' assessment of such regularities. Sayer, for example, has described how qualitative research should explore exactly what it is about certain conditions that appear to enable particular actions to occur or what is it about certain actions that allow certain consequences to follow (Sayer, 2000).

In the case of Learning Together, multiple participant accounts were drawn on to theorise how the conditions present within action groups in some school contexts enabled staff and students to develop mutual understanding and hence better relationships. For example, one student participant described coming to understand teachers' perspectives on the group:

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

Interviews with other students highlighted how such insights encouraged students to develop stronger and more affective relationships with teachers. As one student commented:

If you have a bond with your teacher... you want to do well for the teacher because you feel like she's paid attention to you and gave her respect [in action group meetings]. And the way you can respect her back is by working hard. (Focus group with year 8 students, St. Anselm's School)

But it was also apparent that such processes only occurred, and indeed could only occur, in schools in which action groups were led by senior staff. This was because only senior staff possessed the authority to ensure that action groups were well attended and well facilitated enough that empathy might develop among staff and students through their interactions on such groups. This allowed us to begin to theorise that action groups could trigger mechanisms by which staff and students developed better relationships and mutual empathy, which in turn might engender increased student commitment to school. Reference to other accounts helped us to theorise that such mechanisms were contingent on schools' ability to ensure senior staff committed to attend action groups and that these were well facilitated so that participative and productive conversations ensued. It is important to note that no single participant proposed such an entire theory but this theory was developed indirectly from iterative in-depth exploration of multiple accounts of intervention-related actions.

In another example, interviews explored students' experiences of restorative practice sessions. One interview was with a boy who had been involved in taking a photo of another boy on the toilet. This boy described his emerging sense of responsibility in the restorative session and how the conditions present in the session enabled him to take responsibility for his actions:

I normally would have been moaning [about being punished], saying "No" ... But this time I actually felt what I had done was really wrong. It just made me realise... I mean it's... just when I saw him sitting there in that state [crying during the meeting]. (Interview with year 8 student, Harper's School)

Along with other qualitative data, such data informed refinements to our a priori CMO configurations. We theorised that restorative practice actions could trigger mechanisms involving the development of a sense of accountability among participants. Other qualitative data suggested that this might only generate reductions in bullying when a critical mass of staff were committed to delivering restorative practice and a critical mass of incidents of bullying ensured that such practices were widely deployed. Thus, qualitative data could help us refine how we theorised mechanisms in terms of

individual meaning and how this interacted with intervention processes (May, 2013; Sayer, 2000).

Both approaches might be used in the same study, or even in the same interview, to explore mechanisms from different perspectives. Manzano has suggested that, in realist evaluations, initial exploratory interviews help the evaluator to articulate tentative theories of how intervention activities might trigger generative mechanisms and the contextual contingencies that might affect this. Later interviews might then aim to consider and refine these tentative theories by exploring with participants the particularities of their experiences and what light these might shed on the researcher's emerging understandings of how local conditions are implicated in mechanisms as well as the way in which implementation of interventions modified practitioner's actions and interactions (Manzano, 2016). Our own experience is that, while useful, qualitative research need not always follow such phases in order. It can also be useful for qualitative research to begin from participants' account of their actions and how these might have been modified by intervention activities before moving on to explore participants' own theories of how interventions work and how context affects this.

Sampling

As described above, qualitative research aims to examine the mechanisms triggered by enactments of interventions, the contextual contingencies involved and their consequences. This is best fulfilled by in-depth research in a manageable number of varying case studies. There should be diversity in terms of sampling different contexts and different participants involved in different intervention activities. If possible to ascertain at the sampling stage, it is also good to sample settings or individuals who report different consequences of their involvement in intervention activities. In the case of Learning Together, for example, schools were selected as case studies based on rates of eligibility for free school meals (as a measure of different school cultures and student bodies) as well as facilitator reports of the success of implementation (as a rough measure of the apparent consequences of intervention activities). It is also useful for the

research to have some flexibility in its design so that, if initial analyses of qualitative data suggest different mechanisms or different interactions with context than initially theorised, there is scope to explore these by including new cases.

Within case studies, there should similarly be purposive sampling of individual participants involved in different aspects of intervention delivery and receipt. In the case of Learning Together, for example, interviews were undertaken with school leaders as well as classroom teachers, with intervention developers as well as facilitators, and with teachers and students involved in action groups, restorative practice and the curriculum.

A key issue for sampling as well as analysis is which accounts are most useful to examine which aspects of mechanisms. Manzano has suggested that: provider managers will have the broadest overview of patterns of local successes and failures and so provide particularly authoritative information on how mechanisms are contingent on context; local practitioners will be able to add information about the specific conditions which might affect mechanisms; and clients will offer detailed accounts of their personal experiences of impacts but may have less to say about mechanisms or how these interact with context (Manzano, 2016).

However, in our own study, school leaders and external facilitators often merely offered accounts focused on implementation of intervention activities rather than the consequences of these, so their accounts were less useful in exploring mechanisms. These groups also sometimes presented the 'official' theory of change and offered fewer insights into lived experience of the enactment of the curriculum or restorative practice or the actual consequences of these for preventing bullying. We found that student accounts were often more useful in exploring the mechanisms triggered through intervention-related actions and how these interacted with context to generate reductions in bullying. For example, student accounts were much more useful in exploring the consequences of participation in restorative practice.

Data collection

Observations can allow researchers to witness the actions and interactions that constitute or spin off from the enactment of interventions, which may offer insights into how outcomes might be generated. However, it is likely that observations will only shed light on the sections of mechanisms most proximal to the intervention actions triggering them rather than to the generation of outcomes. In the case of Learning Together, for example, observations provided insights into the nature of interactions within restorative practice sessions but offered fewer insights into the consequences of these. Interviews and focus groups were therefore also essential. As suggested earlier, interviews and focus groups can take a direct approach, exploring participants' views on theories of change. The direct approach will involve the researcher tightly controlling the interview process and asking questions about mechanisms. Manzano gives the following examples (Manzano, 2016):

For example: 'How was your work different before the programme was implemented?', 'Is this new programme going to work for everyone?', 'Could you explain to me the types of people and places where you think it may be more effective?' Stronger questions about context should encourage people to compare subgroups, location, times, before and after. The objective is to draw the interviewee into comparison to explore contextual effectiveness. (p.354-5)

As discussed, interviews and focus groups can also take a less direct approach, exploring participants' accounts of actions, the meanings and goals ascribed to these actions, their conditions and consequences. This approach will involve a more participant-centred interview, exploring participants' accounts of their experiences. Prompts might explore how participants' actions were influenced by: intervention resources and other intervention-related activities; local policies or norms; or the distribution of economic, informational or other resources within a setting (May, 2013). In the case of Learning Together, for example, interviews with those participating on action groups explored how the group's activities were enabled or constrained by the intervention manual and presence of the external facilitator, the training which staff

underwent, as well as the broader culture, management structure, priorities and resourcing of the school. Interviews also explored the consequences that enactment of action groups had for how school processes operated and how staff and students acted and interacted, and how this was affected by the broader context of the school.

Analysing qualitative data on mechanisms

Existing literature offers some guidance on analysis. Realists have suggested, for example, that qualitative data be coded in terms of “‘description of the actual intervention’, ‘observed outcomes’, ‘context conditions’ and ‘underlying mechanisms’” (p. 195) to inform refinement of theory (Marchal et al., 2012). But existing literature has not aimed to offer comprehensive guidance on analysis.

Participant accounts are themselves an interpretation of their experiences of reality so that analysis of such accounts is a ‘double hermeneutic’ exercise (Thirsk and Clark, 2017; Giddens, 1984). As Thirsk and Clark argue, this does not mean that qualitative research cannot examine phenomena external to the participants (Thirsk and Clark, 2017), but it does mean that this is unavoidably mediated by participants’ own interpretations of these phenomena. This is not a weakness of qualitative research but a strength because how participants understand an intervention will often be central to its mechanism (Thirsk and Clark, 2017). With Learning Together, for example, interviews explored how staff and students talked about the intervention. The quote below illustrates one recurring theme, that staff and students tended to emphasise its participative nature, which appeared to be associated in multiple accounts with its ability to transform relationships within a school:

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

Qualitative analysis can be used to develop completely new intervention theory, where little or none previously existed for an intervention, or inform refinement or augmentation of existing theorisation. Although our intervention was informed by a priori theory of change and initial CMO configurations, we put these aside when analysing our qualitative data to inductively generate analyses of how mechanisms worked. We chose not to focus analysis of qualitative data on validating and refining our existing theory so that we could make full use of all our qualitative data and so that our qualitative analysis was not overly influenced by our starting theories.

If qualitative research is to inform theorisation of mechanisms, analysis needs to do more than identify recurring themes. Depending on the intervention in question and the frameworks informing evaluation, analysis may need to consider interactions between microsocial and macrosocial levels (Hedstrom and Swedberg, 1998) or the multiple levels of individual, group, institution and/or society (Marchal et al., 2012). Within realist evaluation, analysis will also need to engage with how mechanisms interact with context or, alternatively, how mechanisms arise from the interaction of intervention activities and context (Lemire et al., 2020) or how agency, context and mechanisms interact to generate outcomes (Porter, 2015).

In our own evaluation, we found that analytical approaches based on grounded theory were useful since these methods explicitly aim to develop theory focused on the identification of social processes (Charmaz, 2014). We used a variant of grounded theory called dimensional analysis because this offers a framework for thinking about how social mechanisms operate with regard to their broad context (the boundaries of a phenomenon), conditions (the specific factors facilitating, blocking or otherwise shaping social action associated with a phenomenon), process (the actions or interactions involved in a phenomenon), consequences (what occurs as a result of the actions involved in a phenomenon) and outcomes (changes in people or groups of people as a result of the phenomenon) (Schatzman, 1991). Although this terminology differs from that used in realist evaluation, we felt that use of this approach aligned well

with realist evaluation. While grounded theory methods and dimensional analysis were developed within the symbolic interactionist approach to sociology, they may be, and indeed are widely, used within other approaches (Oliver, 2011; Charmaz, 2014; Hoddy, 2018). The analytical techniques they involve, such as constant comparison, theoretical sampling and abductive reasoning, can be undertaken within other approaches, such as realist evaluation. We used dimensional analysis to analyse staff and student accounts describing or implying process (for example, increasing commitment to school), linked within and across interviews to accounts of conditions (for example, having positive experiences of participating in the action group), consequences (for example, decreasing involvement in anti-school peer groups) and outcomes (for example, reduced involvement in behaviours such as aggression).

In terms of practical procedures, analysis will need to draw on different accounts in order to develop a hermeneutic, pluralistic theorisation of mechanisms from the point at which they are triggered by intervention enactment to how they generate outcomes. For example, through qualitative analysis we developed theory as to how enactment of action groups might trigger mechanisms generating student commitment and, through this, reducing student involvement in aggression. This was pieced together from insights gathered from many different interviews, focus groups and observations. Analyses will need to compare and contrast different accounts, deciding which accounts provide more or less authoritative insights into particular sections or aspects of the mechanism. This requires axial coding which draws on an initial wave of in-vivo coding to generate cross-cutting and higher-order concepts.

As argued above, as well as theorising or refining theorisation about mechanisms by analysing participants' own theories, qualitative analyses can also explore the conditions necessary for mechanisms to 'trigger'. This can take into account quasi-quantitative analysis of patterns of contingencies but it also requires an analysis of exactly what it is about certain conditions that enable certain actions, or what is it about the characteristics of certain actions that enable certain consequences. Such analysis will usefully employ techniques associated with grounded theory, such as deviant case

analysis. For example, some of the insights into what conditions were necessary to ensure action group actions could trigger mechanisms generating increased student commitment came from observations and interviews in a school in which the action group did not attract broad staff participation and hence failed to encourage staff and students to better understand each other's perspectives.

Informing quantitative hypotheses

Through these different approaches, qualitative research can develop, augment or refine theory of how mechanisms appear to interact with context to generate outcomes. As discussed above, we oriented our analysis of qualitative data to theorising mechanisms without limiting this to the validation or refinement of our initial theories. However, once our analyses were complete, we used the completed qualitative analysis to refine and augment our CMO configurations.

Depending on the precise form of theorisation (or in the case of realist evaluation, the components that configurations include (Porter, 2015; Lemire et al., 2020)), qualitative research might inform refinement of theories about how mechanisms interact with pre-existing context to generate outcomes (as we and most other realist evaluators conceive it) or about how mechanisms triggered by intervention activities in interaction with context generate outcomes.

However, we have to accept that any such theories, even if refined through qualitative research, might be wrong for several reasons. Firstly, analysis of participants' accounts will be limited by the extent to which these accounts are themselves fallible (Sayer, 2000). Drawing on the accounts of multiple others to develop a hermeneutic, pluralistic account will to some extent compensate for this. However, even a research account based on multiple accounts will sometimes be wrong.

Secondly, qualitative analyses drawing on patterns of regularities to consider what contingent factors appear to be important for actions or consequences to occur will inevitably be based on a relatively small number of observations and accounts of events and may therefore be subject to chance coincidences. This is particularly likely to be a problem where intervention impacts are not large (which is commonly the case in most public health interventions for example) so that it is hard for case studies to determine what factors are influential.

Thirdly, analyses of what it is about certain conditions or actions that enable certain other actions and consequences will be limited by the available data, the theories used to inform analysis and the broader conceptual hinterland of the researcher. For example, in the Learning Together evaluation, broad staff participation was identified as a key enabler of action groups being able to trigger sharing of perspectives between staff and students. This was informed by powerful evidence from a small number of cases. Other factors might have also been apparent had processes in other schools been explored.

We are arguing therefore that there is a role for correlational quantitative research in checking whether broader patterns of regularities appear to align with the theories developed or refined through qualitative research. This is a controversial area. Some realist evaluators are open to the use of quantitative alongside qualitative data to examine mechanisms (Marchal et al., 2012). Others, including some from a critical realist perspective, are critical of quantitative research examining regularities, arguing that explanation of social mechanisms cannot be reduced to a search for regularities. Andrew Sayer, for example, argues that this approach fails to appreciate that, in 'open systems', simple regularities rarely occur (Sayer, 2000):

... events arise from the workings of mechanisms which derived from the structures of objects, and they take place within geo-historical contexts. This contrasts with approaches which treat the world as if it were no more than patterns of events, to be registered by recording punctiform data regarding 'variables' and looking for regularities among them... Given the variety and

changeability of the contexts of social life, this absence of regular associations between 'causes' and 'effects' should be expected. (Sayer 2000: 15-16)

While we recognise this very real risk, we do not think it is always wrong to explore regularities quantitatively. Examples of regularities identified through statistics that are informative include that: people in nations with high levels of income inequality generally experience worse health outcomes across all social classes once other confounding differences, such as gross domestic product, are accounted for (Wilkinson and Pickett, 2009); within countries, those of lower socioeconomic status experience worse health (Marmot, 2004); and schools which engage all students in learning generally have lower rates of student violence and substance use (Bonell et al., 2013).

We also note that more advanced forms of statistical analysis can be used to examine more complex forms of regularities. Examining effect modification enables an assessment of how the association of two factors is contingent on the presence of one or more other factors, allowing for an assessment of what works for whom and where. Examining mediation enables an assessment of whether the causal association between two factors can be explained by a pathway through a third intervening factor, allowing for an insight of how an intervention might work. To the extent that variables are ontologically different from mechanisms (Falleti and Lynch, 2009), mediation analyses themselves do not 'test' mechanisms. However, they provide analytic traction in developing and refining models of intervention functioning.

Statistical research can even examine moderated mediation, which assesses whether the mediation of the causal effect of one factor on another is contingent on the presence of another factor. For example, we used moderated-mediation analyses to examine whether our refined CMO configurations aligned with broader regularities. These statistical analyses suggested that intervention beneficial effects on bullying and mental health outcomes were mediated by a quantitative measure of student sense of belonging in school - but that such mediation only occurred in a subset of schools with good baseline measures of management capacity, student belonging and low levels of

bullying. In other schools, the intervention was similarly effective in reducing measures of bullying and mental distress but these effects were not mediated by increased student belonging (Melendez-Torres et al., 2021).

Hence, we believe that there is a role for quantitative research in assessing whether CMO configurations or other hypotheses that emerge from, or are refined via, qualitative research appear to explain broader regularities. The use of qualitative research to inform these hypotheses is important in ensuring that quantitative research is limited to assessing plausible hypotheses and does not merely dredge statistical data looking for spurious associations.

The hypotheses that qualitative research offers up for testing should be orientated towards a view of causation which recognises the contingency of correlations in open social systems. As already reported, we used CMO configurations as developed by realist evaluators (Pawson and Tilley, 1997). Qualitative analyses usually generate ‘thick’ descriptions and theories but these can inform the specification of more abstracted CMO configurations. These can then be tested using the various statistical analysis methods described above. Where these hypotheses are not supported by quantitative evidence, this should encourage reflection. It may be that the quantitative measures fail to capture the phenomena identified in qualitative research or that quantitative samples are insufficiently powered to identify real patterns of regularities. But it may also indicate that qualitative research has provided misleading evidence as to the mechanisms triggered in a setting or how these interact with context to generate outcomes.

As suggested earlier, some mechanisms may be too complex to subject to quantitative analysis using data collected using measurement strategies standard in evaluations. In such cases, quantitative research may be able to focus on testing some but not all aspects of mechanisms and, for other aspects, qualitative research may be as far as the analysis can be taken. Some mechanisms might in principle be open to quantitative examination but be developed at a point in an evaluation when it is too late to identify

suitable quantitative measures. In such cases, these hypotheses might form the focus of future studies.

Conclusion

Qualitative research can be useful in developing, augmenting or refining theories about the generative mechanisms which interventions trigger. This might occur directly, whereby interviews and focus groups explore participants' theories of how interventions work (Manzano, 2016; Pawson, 1996). Or it might be indirect, exploring intervention activities, the conditions which enable these and the consequences of these (Sayer, 2000) to build up a picture of a mechanisms. The direct approach requires interviews or focus groups to be tightly controlled by the researcher, who asks direct questions about participants' ideas about mechanisms (Manzano, 2016; Pawson, 1996). The indirect approach requires a more participant-centred agenda exploring experiences of intervention activities, and the conditions and consequences of these.

Analysis of qualitative data needs to focus on building or refining theory rather than merely identifying themes. We found that dimensional analysis (Schatzman, 1991) provided a useful framework for theorisation. Qualitative research might then inform more abstracted CMO configurations or other hypotheses about how mechanisms generate outcomes. Qualitative and quantitative research can together build stronger though still indirect and fallible evidence of how mechanisms might generate different outcomes in different contexts. This mixed-method approach to analysing mechanisms should help ensure that evaluation contributes to assessments of potential transferability (Burchett et al., 2020), intervention refinement (Bonell et al., 2021) and broader scientific understanding and future interventions (Davey et al., 2019).

References

Astbury B and Leeuw FL (2010) Unpacking black boxes: mechanisms and theory building in evaluation *American Journal of Evaluation* 31(3): 363-381.

- Barnes M, Matka E and Sullivan H (2003) Evidence, understanding and complexity: evaluation in non-linear systems. *Evaluation* 9: 265–284.
- Bonell C, Allen E, Warren E, et al. (2018) Initiating change in the school environment to reduce bullying and aggression: a cluster randomised controlled trial of the Learning Together (LT) intervention in English secondary schools. *The Lancet* 392(10163): 2452-2464.
- Bonell C, Fletcher A, Morton M, et al. (2012) 'Realist Randomised Controlled Trials': a new approach to evaluating complex public health interventions. *Social Science and Medicine* 75(12): 2299-2306.
- Bonell C, Jamal F, Harden A, et al. (2013) Systematic review of the effects of schools and school environment interventions on health: evidence mapping and synthesis. *Public Health Research* 1(1).
- Bonell C, Prost A, Melendez-Torres GJ, et al. (2021) Will it work here? A realist approach to local decisions about implementing interventions evaluated as effective elsewhere. *J Epidemiol Community Health* 75(1): 46-50.
- Breuer E, Lee L, De Silva M, et al. (2016) Using theory of change to design and evaluate public health interventions: a systematic review. *Implementation science* 11(1): 63.
- Burchett HED, Kneale D, Blanchard L, et al. (2020) When assessing generalisability, focusing on differences in population or setting alone is insufficient. *Trials* 21(1): 286.
- Carey RN, Connell LE, Johnston M, et al. (2019) Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. *Annals of behavioral medicine* 53(8): 693-707.
- Charmaz K (2014) *Constructing Grounded Theory*. New York: Sage.
- Cohn S, Clinch M, Bunn C, et al. (2013) Entangled complexity: Why complex interventions are just not complicated enough. *Journal of Health Services Research & Policy* 18(1): 40-43.
- Davey C, Hassan S, Cartwright N, et al. (2019) *Designing Evaluations to Provide Evidence to Inform Action in New Settings*. London: Department for International Development.
- Falleti TG and Lynch JF (2009) Context and causal mechanisms in political analysis. *Comparative Political Studies* 42(9): 1143-1166.
- Gardner F, Burton J and Klimes I (2006) Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: outcomes and mechanisms of change. *Journal of Child Psychology and Psychiatry* 47(11): 1123-1132.
- Giddens A (1984) *The Constitution of Society*. Cambridge: Polity Press.
- Hawe P, Shiell A and Riley T (2004) Complex interventions: how "out of control" can a randomised controlled trial be? *British Medical Journal* 328: 1561-1563.
- Hedstrom P and Swedberg R (1998) *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge, UK: Cambridge University Press.
- Hoddy E (2018) Critical realism in empirical research: employing techniques from Grounded theory methodology. *International Journal of Social Research Methodology* 22: 111-124.

- Lacouture A, Breton E, Guichard A, et al. (2015) The concept of mechanism from a realist approach: a scoping review to facilitate its operationalization in public health program evaluation. *Implementation science* 10: 153.
- Lemire S, Kwako AJ, Nielsen SB, et al. (2020) What is this thing called a mechanism? Findings from a review of published realist evaluations. *New Directions for Evaluation* 167: 73-86.
- Manzano A (2016) The craft of interviewing in realist evaluation. *Evaluation* 22(3): 342-360.
- Marchal B, van Belle S, van Olmen J, et al. (2012) Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems. *Evaluation* 18(2): 192-212.
- Markham WA and Aveyard P (2003) A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice. *Social Science & Medicine* 56(6): 1209-1220.
- Marmot MG (2004) *Status Syndrome: How Your Social Standing Directly Affects Your Health and Life Expectancy*. London: Bloomsbury.
- May C (2013) Towards a general theory of implementation. *Implementation science* 8: 18.
- Melendez-Torres GJ, Warren E, Viner R, et al. (2021) Moderated mediation analyses to assess intervention mechanisms for impacts on victimisation, psycho-social problems and mental wellbeing: Evidence from the INCLUSIVE realist randomized trial. *Social Science and Medicine* 279: 113984.
- Moore G, Audrey S, Barker M, et al. (2014) *Process evaluation of complex interventions: UK Medical Research Council (MRC) guidance*. London: MRC Population Health Science Research Network.
- Moore GF and Evans RE (2017) What theory, for whom and in which context? Reflections on the application of theory in the development and evaluation of complex population health interventions. *Social Science and Medicine Population Health* 3: 132-135.
- Oliver C (2011) Critical realist grounded theory: A new approach for social work research. *British Journal of Social Work* 42(2): 371-387.
- Pawson R (1996) Theorizing the interview. *British Journal of Sociology* 47(2): 295-314.
- Pawson R and Tilley N (1997) *Realistic evaluation*. Sage.
- Porter S (2015) The uncritical realism of realist evaluation. *Evaluation* 21(1): 65-82.
- Sayer A (2000) *Realism and Social Science*. London: Sage.
- Schatzman L (1991) Dimensional analysis: notes on an alternative approach to the grounding of theory in qualitative research. In: Maines D (ed) *Social organization and social process: Essays in honor of Anselm Strauss*. New York: Aldine, pp.303-314.
- Schütz A (1962) *The Problem of Social Reality: Collected Papers I*. The Hague: Martinus Nijhoff.
- Skivington K, Matthews L, Simpson S, et al. (2021) A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *British Medical Journal* 374: n2061
- Thirsk LM and Clark AM (2017) Using qualitative research for complex interventions: the contributions of hermeneutics. *International Journal of Qualitative Methods* 16: 1-10.

Unrau YA (2001) Using client exit interviews to illuminate outcomes in program logic models: a case example. *Evaluation and Program Planning* 24: 353-361.

van Urk F, Grant S and Bonell C (2016) Involving stakeholders in programme theory specification: discussion of a systematic, consensus-based approach. *Prevention Science, in preparation* 12(14): 541-557.

Warren E BL, Opondo C, Allen E, Mathiot A, West G, Jamal F, Viner R, Bonell C. (2019) Action groups as a participative strategy for leading whole-school health promotion: results on implementation from the INCLUSIVE trial in English secondary schools. *British Education Research Journal* 45(5): 748-762.

Warren E, Melendez-Torres GJ, Viner RM, et al. (2020) Using qualitative research within a realist trial to build theory about how context and mechanisms interact to generate outcomes: findings from the INCLUSIVE trial of a whole-school health intervention. *Trials* 21: 774.

Wilkinson R and Pickett K (2009) *The Spirit Level: Why More Equal Societies Almost Always Do Better*. London: Allen Lane.

Appendix 3: Are randomised controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services

Bonell *et al. Trials* (2018) 19:238
<https://doi.org/10.1186/s13063-018-2589-4>

Trials

METHODOLOGY

Open Access

Are randomised controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services



Chris Bonell^{1*}, Graham Moore², Emily Warren¹ and Laurence Moore³

Abstract

Background: We have previously proposed that trials of social interventions can be done within a “realist” research paradigm. Critics have countered that such trials are irredeemably positivist and asked us to explain our philosophical position.

Methods: We set out to explore what is meant by positivism and whether trials adhere to its tenets (of necessity or in practice) via a narrative literature review of social science and philosophical discussions of positivism, and of the trials literature and three case studies of trials.

Results: The philosophical literature described positivism as asserting: (1) the epistemic primacy of sensory information; (2) the requirement that theoretical terms equate with empirical terms; (3) the aim of developing universal laws; and (4) the unity of method between natural and social sciences. Regarding (1), it seems that rather than embodying the epistemic primacy of sensory data, randomised controlled trials (RCTs) of social interventions in health embrace an anti-positivist approach aiming to test hypotheses derived deductively from prior theory. Considering (2), while some RCTs of social interventions appear to limit theorisation to concepts with empirical analogues, others examine interventions underpinned by theories engaging with mechanisms and contextual contingencies not all of which can be measured. Regarding (3), while some trialists and reviewers in the health field do limit their role to estimating statistical trends as a mechanistic form of generalisation, this is not an inevitable feature of RCT-based research. Trials of social interventions can instead aim to generalise at the level of theory which specifies how mechanisms are contingent on context. In terms of (4), while RCTs are used to examine biomedical as well as social interventions in health, RCTs of social interventions are often distinctive in using qualitative analyses of data on participant accounts to examine questions of meaning and agency not pursued in the natural sciences.

Conclusion: We conclude that the most appropriate paradigm for RCTs of social interventions is realism not positivism.

Keywords: Randomised controlled trials, Positivism, Realism

* Correspondence: chris.bonell@lshtm.ac.uk

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

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



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Randomised controlled trials (RCTs) have been used for many decades to evaluate not only biomedical interventions but also social interventions in fields such as public health, health services, economics and education [1–4]. RCTs are used to generate some of the evidence intended to inform evidence-based practice and policy. Evidence-based policy has a long intellectual history in which authors such as Donald Campbell and Karl Popper have argued that social science experiments should provide evidence to inform and assess “piecemeal social engineering” [5, 6]. This process involves incremental changes to public policy, which are evaluated to assess whether they have achieved their intended objectives and whether there have been any unintended harmful consequences. But in fields such as public health and health services, evidence-based policy, and in particular the use of RCTs, has attracted criticism in terms of its ontological (i.e. concerning the nature of reality) and epistemological (i.e. concerning how we know about reality) assumptions. These are said to situate RCTs firmly within a “positivist” paradigm [7–11]. We will explore later what is meant by positivism.

We have previously proposed that RCTs can contribute to “realist” evaluation of social interventions in public health or health services [12]. Realist evaluators argue that rather than merely examining “what works”, evaluations should examine what works, for whom and under what conditions [13]. Informed by critical realist philosophy [14], they suggest that social interventions do not produce outcomes directly but rather that interventions introduce new resources into social settings (or redistribute or displace existing resources). Local actors might then draw on these resources to enact local actions, which may then in turn trigger mechanisms that generate the intended and unintended “outcomes” of the intervention. Realist evaluators propose that the focus of evaluation should be on these mechanisms, which may play out differently in different contexts to generate different outcomes.

Critical realists further suggest that we can think of the world in terms of an “empirical” realm consisting of experience, an “actual” realm of occurrences whether or not these are observed and a “real” realm consisting of unobservable causal mechanisms that generate events in the actual realm. Realists argue against understanding causality merely in terms of observed “constant conjunctions” of causes and effects. They argue that causal mechanisms are tendencies and whether effects are generated depends on other factors. Mechanisms may be triggered but cancelled out by other mechanisms or may not be triggered at all depending on local circumstances. Therefore, a lack of “constant conjunction” does not necessarily mean that our theories about mechanisms are wrong [15, 16]. Instead of

evaluations focusing principally on estimating overall associations between allocation or exposure to an intervention and measures of outcome (i.e. effect sizes), realist evaluations aim to develop and test hypotheses concerning context-mechanism-outcome configurations [13].

We have previously suggested that RCTs and systematic reviews of RCTs could contribute to a realist approach to the evaluation of social interventions by focusing not merely on overall effect sizes but also by building and testing hypotheses about context-mechanism-outcome configurations. We have suggested that the plausibility of context-mechanism-outcome hypotheses could be examined by assessing whether these are borne out by “data signatures” from trials. Within a single RCT, moderator analyses might examine how outcomes vary between subgroups of sites, clusters or individuals defined by varying contextual factors. Within a systematic review, meta-regressions might explore how study-level effect sizes are moderated by context. The aspect of context in question (for example, whether the setting was urban, suburban or rural) need to be measured in a comparable way across trials or such information needs to be available from other sources available to the reviewer.

However, these proposals have been criticized by some realist evaluators, who argue that RCTs are irredeemably positivist and therefore inimical to realist enquiry [8]. Our realist critics have requested that we discuss our views concerning whether RCTs are positivist [8]:

“Bonell et al... indicate that they do not necessarily agree that RCTs are based on a positivist ontological and epistemological foundation, but they opt not to discuss this further... This is a pity, because ... it is the ontological position and its epistemological consequences that limit the usefulness of RCTs when applied to complex interventions.” (p. 125)

We agree that it would be useful for us to address this question of whether or not trials are positivist. RCTs are often described as being positivist by social scientists and this view may be an important barrier to harnessing realist approaches to improve the conduct of trials of social interventions in public health and health services, and enabling deeper collaborations between trialists and social scientists.

Methods

The paper does not aim to repeat all our previous arguments in favour of realist RCTs but instead aims to focus on the question of what positivism means and whether RCTs are *of necessity* or *in practice* positivist. We first examined how the term positivism has been used in the social science literature in the fields of health and education that describes or criticises RCTs as positivist. We

then explored how positivism has been defined in the wider literature on the philosophy of science. The paper then goes on to consider whether public health and health services research RCTs appear to embody the various tenets of positivism, and whether this is a necessary or contingent feature. In doing so we refer to three exemplar trials of school-based health interventions (see Table 1).

Results

What is positivism?

Social science references to RCTs as a positivist design

RCTs are frequently described by social scientists working in the fields of health and education as positivist. Some of this literature is descriptive and some critical. A good example of the former is Green and Thorogood [7] who described RCTs thus:

“the ‘classic’ design of the positivist tradition, as it sets up a study capable of answering a question about cause and effect.” (p. 34)

Green and Thorogood identified several features of positivism:

“[positivism] assumes that there is a stable reality out there... human understandings may be flawed ... but there is a potential ‘right’ explanation that we are getting closer to as understanding of health and disease increases... There is a stress on *empiricism*, or studying only observable phenomena..., *a unity of method*, the idea that eventually, when mature, all sciences will share the same methods of enquiry. At this point of maturity, the proper object of scientific inquiry is the establishment of relationships of cause and effect and the generation of laws about the natural world. That many of the social sciences focus on other questions is, in this view, evidence of their immaturity.” (p. 12, italics as published)

Authors who criticize RCTs as positivist tend to offer less comprehensive definitions than above of what they mean by positivism, the term sometimes being used pejoratively and vaguely. The originators of realist evaluation [13] acknowledged this tendency themselves:

“Experimental evaluation has struggled because of a basically ‘positivist’ understanding of the nature of social causation. We hesitate to put it like this, since the term ‘positivism’ these days has been reduced to a crude term of abuse. It is used as an evil totem by those intent on musing about there being no place for scientism in understanding the rich, meaningful, emotional world of human intercourse.” (p. 30)

What features of RCTs are presented as positivist then in these critical accounts? Rowe and Oltmann [10] suggest RCTs are positivist because they aim to produce objective knowledge including about causality and, in doing so, aim to test hypotheses:

“The evidence upon which EBP [evidence based practice] is premised is usually derived from experimental research conducted in professional disciplines that are firmly rooted in the positivist paradigm; the research method most closely associated with this is the randomised controlled trial (RCT). RCTs are quantitative, controlled experiments in which the effect of an intervention can be determined more objectively than by observational studies... It seems clear that those who most strongly advocate the use of RCTs in education have an inherent bias against other methods of data collection, strongly positioning themselves within a positivist interpretation of reality... Positivist research maintains that knowledge is objective, that it involves hypothesis testing and identifies causality.” (p. 6-7)

We will explore whether this quest for objective knowledge and focus on causality is viewed as a distinguishing feature of positivism in the philosophical literature.

The argument that hypothesis testing implies a positivist approach has been made not only by Rowe and Oltmann [10] but also by Tones and Green [11], who similarly labelled RCTs as positivist and suggested that trials take what they describe as a “hypothetico-deductive approach” to generating scientific knowledge (p. 310). As we shall see in the next section, the philosophy of science literature takes a very different view about positivism and the hypothetico-deductive approach.

Other critics have focused on different aspects of what they see as RCTs’ positivist approach. For example, Pearce and Raman [9], in their critique of the positivistic application of RCTs to informing public policy, focused on positivism aiming to develop generalisable conclusions devoid of context:

“When RCTs are presented as offering generalizable evidence of what works, the conditions and assumptions built into their doing and interpretation are erased from the story... The wider context in which an intervention works is ignored, and it is implied that success in one context can simply be transferred to another.” (p. 35)

This concern with generalisable knowledge and whether it implies a lack of concern with context is something that we will return to in the next section.

Table 1 Three exemplar trials of school-based health interventions

Child Development Program (CDP) RCT [21, 22, 38, 39]
<i>Timing:</i> 1982–89
<i>Setting and sample:</i> three elementary schools in the intervention group and three elementary schools in the control group, in northern California.
<i>Intervention:</i> this aimed to “encourage pro-social behaviour by providing children with several types of experience which serve to engender a sense of community and a climate of mutual respect and concern in the classroom and school” [22] (p.149). Activities included cooperative learning, and involvement of children in rule setting, discussions and helping activities.
<i>Outcome evaluation:</i> this reported positive intervention effects on interview-assessed cognitive problem-solving and conflict resolution skills, increased questionnaire-reported peer acceptance, reduced loneliness and anxiety and increased observer-rated prosocial behaviours. There were no effects on questionnaire-reported measures of self-esteem, liking of school, perceived social competence or popularity.
<i>Process evaluation:</i> observations indicated that intervention classrooms were more likely to use strategies promoted by the intervention, particular where teachers were rated as of high competence.
<i>Rationale for including as an example:</i> this RCT is quite old and did not employ any qualitative research despite focusing on a social intervention. It is therefore a good case study to assess whether in practice some trials might be positivist in approach.
A Stop Smoking in Schools Trial (ASSIST) RCT [27, 31, 32]
<i>Timing:</i> 2001–4
<i>Setting and sample:</i> Twenty-nine secondary schools allocated to intervention and 30 to be controls in western England and south-east Wales.
<i>Intervention:</i> Secondary school-based peer education outside classrooms focused on smoking prevention.
<i>Outcome evaluation:</i> This reported a reduction in the prevalence of smoking in the past week overall and among those who had smoked at baseline.
<i>Process evaluation:</i> Quantitative and qualitative research found that teachers generally supported the intervention despite concerns about some aspects such as the possibility that students might nominate some individuals as peer educators who teachers did not see as good representatives of the school. The evaluation found that peer educators themselves tended to focus messages on information more than persuasion and primarily targeted non-smoking friends.
<i>Rationale for including as an example:</i> This was a trial led by one of the authors of this paper (Initials withheld for blind-reviewing) which though not explicitly realist or anti-positivist in orientation, nonetheless focused on questions of how, for whom and under what circumstances the intervention worked. It therefore offers a promising case study to assess whether or not in practice a modern trial of a social intervention has positivist tendencies.
Initiating change locally in bullying and aggression through the school environment (INCLUSIVE) RCT [24, 46, 60]
<i>Timing:</i> 2014–17
<i>Setting and sample:</i> 20 secondary schools in the intervention group and 20 secondary schools in the control group, all in south-eastern England.
<i>Intervention:</i> a whole-school intervention to reduce bullying, aggression via training staff in restorative practice, provision of local data and a facilitator to enable local needs-led decisions involving staff and students and a social and emotional learning curriculum.
<i>Outcome evaluation:</i> this evaluated effects on student questionnaire-reported bullying and aggression (primary outcomes) plus secondary

Table 1 Three exemplar trials of school-based health interventions (*Continued*)

outcomes including student substance use, mental and sexual health and quality of life as well as staff attendance, quality of life and burnout.
<i>Process evaluation:</i> ongoing quantitative and qualitative research on intervention implementation, reach, acceptability and mechanisms, and how these varied by context.
<i>Rationale for including as an example:</i> this was a trial involving some of the authors of this paper (initials withheld for blind-reviewing) and was explicitly realist and anti-positivist in orientation. It is therefore a good case study to assess whether trials can avoid the various tenets of positivism or whether this is unavoidable.

In their own critique of RCTs, the realist evaluators Marchal et al. [8] focused on positivism as involving a concern with observable phenomena without considering the ways in which causation actually operates. They described RCTs as being:

“built upon objectivist (or ‘positivist’) assumptions, which hold that causality cannot be observed and that the best we can do is to demonstrate regularity between a particular intervention and a particular outcome” (p. 125)

This point will be considered further under point 2 of the next section.

Taken together, these descriptive and critical accounts of trials as positivist enable us to start to develop a sense of what positivism is and why RCTs might be viewed as a positivist strategy of research. But to get a more systematic sense of what are the distinguishing tenets of positivism, we need to examine how positivism has been defined in the wider literature on the philosophy of science.

Descriptions of positivism in the philosophy of science literature

The philosophical literature has described positivism’s long history and multiple schools ranging across different academic disciplines [17–19]. This literature has systematically mapped out several key tenets concerning how positivist social enquiry should proceed, some but not all of which also appear in the trials literature reviewed above [16, 18, 19].

The epistemic primacy of direct sensory information as the basis for scientific knowledge The philosophical literature has not identified objectivity as a distinguishing feature of positivism. Philosophers hold that positivism implies a belief not merely in a stable reality that exists independently of our senses (which critical realists also accept), but that knowledge of this world must derive entirely from our senses. In describing positivism, Blaikie [16] suggested:

“That which is to count as knowledge must be based on experience, on what an observer can perceive by

his or her senses..., it must be 'pure experience' with an empty consciousness" (p. 14)

This view suggests that both our informal knowledge as individuals and our more formal theories as social scientists about how phenomena relate to one another can be derived directly from sensory information. This view has its roots in the "empiricist" philosophy, for example of John Locke, which regarded the mind as a "blank slate" upon which knowledge is written, purely by the actions of generic logical mental processes applied to information from the senses [18]. The philosophy literature offers clarity on this point whereas the social science literature on trials does not. Although Green and Thorogood correctly suggested that positivism is based on an empiricist approach to knowledge [7], Tones and Green as well as Rowe and Oltmann suggested incorrectly that positivists embrace a hypothetico-deductive approach to producing knowledge [10, 11]. How RCTs actually engage with these questions will be explored later in this paper.

The requirement that theoretical terms must equate with empirical terms Blaikie argued that positivism holds that, to be meaningful, theoretical concepts must be able to be translated directly into empirically measurable elements. Thus, it is not merely that positivists focus on questions of cause and effect (indeed, realists also clearly focus on such questions) but rather that positivists believe that research should examine causal links between observable phenomena rather than speculating about underlying, unobservable mechanisms that might generate such causal links [16]. This echoes the argument made by Marchal et al. that RCTs are positivist because they do this [8]. Whether they do or not in practice will be considered later in this paper.

The aim of developing universally applicable laws For positivists, the aim of both natural sciences, such as biology, chemistry and physics, and social sciences, such as sociology, is to produce laws that apply universally, a point raised above in relation to RCTs by Green and Thorogood [7] and Pearce and Ramen [9]. Blaikie [16] suggested that for positivists: "laws summarise observations by specifying simple relations or constant conjunctions between phenomena" (p. 15). Hacking [18] argued that positivists advocate that science should understand causality not as a thing in itself but solely in terms of the constant conjunctions of observable phenomena. Bhaskar [14] wrote:

"Positivism pivots on the ... theory of constant conjunctions of atomistic events or states of affairs, interpreted as the objects of actual or possible experience." (p. 158)

However, there is no suggestion in this literature that the development of general laws implies a lack of interest in contextual contingencies. Later in the paper, we will explore how Karl Popper argued that science, including social science, should concern itself with developing general laws of cause and effect but that these should include within them consideration of how contextual contingencies will influence causation. Later in this paper, we will explore how our case study RCTs address questions of general conclusions.

A unity of method between the natural and social sciences This unity of method, referred to above by Green and Thorogood [7], concerns the overall approach to doing science: the exclusive focus on identifying regularities using researcher-controlled experiments. It does not refer to the specific methods that each branch of science uses because these will vary depending on the phenomena under investigation. This goal of a unified approach contrasts with the view that the social sciences need a totally different approach to the natural sciences because the "objects" of social scientific enquiry are quite different and not natural phenomena, such as atoms and antelopes. Humans are themselves subjects who have their own interpretations of the world and engage in willed, meaningful action. The classic anti-positivistic approach to social science is exemplified in the hermeneutic tradition of Max Weber which aims to *interpret and understand*, rather than *predict*, action based on the meanings conferred on it and the agency underpinning it on the part of social actors [20]. We shall explore later whether trials of social interventions in the health sector take an exclusively natural science approach or whether they engage with more Weberian approaches.

Drawing on the philosophy literature, we have identified a systematic set of tenets that should distinguish a positivist approach to research. The next section examines whether RCTs conducted in the field of public health and health services actually embody these tenets and, if so, whether this is a necessary or merely a contingent feature. We make these assessments based on a review of RCT research in public health and health services, and in particular of the RCTs of school-based health interventions described earlier.

Are randomised trials of social interventions in health positivist?

Do trials give primacy to sensory information in building knowledge?

There is no evidence that those undertaking RCTs of social interventions in health assume that all knowledge is derived from sensory experience. Medical Research Council guidance for RCTs of complex interventions has

highlighted the importance of developing coherent and explicit theory of intervention mechanisms prior to, not as a result of, evaluation [1].

The hypotheses that RCTs test certainly appear to be derived deductively from prior theories of change, whether or not these are explicitly stated. For example, even in the case of our apparently positivist RCT of the Child Development Program (CDP) intervention, where there was no formal theory of change for the intervention, the trial reports did nonetheless discuss the mechanisms by which the intervention was intended to work, grounded in descriptions of previous theory and empirical research [21]. The research reports located the outcomes to be examined in terms of gaps in previous literature and of theory on children's prosocial development. These were not worded as formal hypotheses but as expectations [22]. The ASSIST RCT prospectively identified a primary outcome of recent smoking and was explicitly informed by theory concerning the diffusion of prevention messages within a school social network. The INCLUSIVE RCT explicitly aimed to test hypotheses derived from a sociological theory of change concerning how changes to the school environment might promote student engagement and health [23–25].

Tones and Green rightly cite Karl Popper as making the case for science proceeding via the empirical testing of hypotheses derived deductively from theory but incorrectly view this as a positivist strategy. In fact, Popper argued for the hypothetico-deductive approach as an alternative to the naïve inductive empiricism of positivism. Popper himself was very clear that theories should direct empirical social research rather than being inductively built from it [6]:

“The fact that I have discussed the problem of social experiments before discussing ... the problem of sociological ... theories ... does not mean that I think observation and experiments are ... logically prior to theories. On the contrary I believe that theories are prior to observations as well as experiments, in the sense that the latter are significant only in relation to theoretical problems.” (p. 89-90)

“[I]n the social sciences it is even more obvious than in the natural sciences that we cannot see and observe our objects before we have thoughts about them. For most of the objects of social science, if not all of them, are abstract objects: they are theoretical constructions.” (p. 125)

Popper's approach was a “post-positivist” one of ontological realism, accepting that a world exists independent of our senses but avoiding the naïve empiricism that saw human knowledge being constructed only from sensory

information. Popper recommended the pursuit of objective truth but the recognition that this can only occur via attempts to test our cognitively derived theories. Theories will influence the questions we ask, what is observed and how it will be measured.

Do trials require that theoretical concepts must translate into empirical measures?

Most RCTs performed in the fields of public health and health services research focus on statistical measures of the association between quantitative measures of allocation or exposure to interventions and quantitative measures of health or risk states [1, 2]. This does appear at first to suggest a positivist approach in that understanding of cause and effect is apparently reduced to knowledge of constant conjunctions between empirical measures. However, such an approach is not particular to RCT research. Furthermore, in using statistics to estimate associations between interventions and outcomes, trialists are not searching for constant conjunctions. Indeed, an assumption that different individuals allocated to the same interventions will report different outcomes (i.e. that interventions and outcomes are not constantly conjoined) is built into trial statistics. An odds ratio, for example, presents the relative odds of a particular outcome in a group of individuals allocated to an intervention compared to a group of individuals not thus exposed. If intervention and outcome were constantly conjoined (i.e. if every single individual exposed to the intervention were to experience the same outcome) the odds ratio would be infinity. A focus on aggregate effects therefore does not imply that a trialist is thinking of cause and effect in terms of simple constant conjunctions. Rather, it is an attempt to estimate the extent to which the net effect of an intervention on an overall population for a particular outcome would be harmful or beneficial if widely used instead of, or in addition to, usual practice.

We would argue that while this statistical estimate of overall harms and benefits should not be the only information on causality that RCTs provide, it is nevertheless a valid and useful question for informing decisions. A primary focus on whole-population effects is appropriate for example when considering the effects of public health interventions informed by the Rose hypothesis, since here the focus is on population-wide and not subgroup effects [26]. For example, the ASSIST RCT reported an overall effect of the intervention in reducing smoking, not because the authors believed the intervention would have the same effect on every individual or in every school but because in judging the success of public health interventions, it is important to estimate the potential of the intervention to contribute towards population-level reductions in risk [27]:

“... if implemented on a UK-wide basis [ASSIST] could potentially reduce the number of 14–15-year-old school students taking up regular smoking by 43,289” (p. 1601)

This brings us to consider how interested trialists are in understanding causality beyond statistical associations of interventions and outcomes. It must be acknowledged that many trials have been conducted which have not theorised or empirically examined the intervening mechanisms or impacts that connect an intervention and its endpoints [28]. Even where RCTs do include a theory of change, in many cases this is little more than a string of empirical measures with arrows denoting lines of causation from intervention to mediating factors to proximal and distal outcomes, which is then sometimes empirically tested using mediation analyses [28]. Such theories rarely describe the real mechanisms that underlie causation and generate outcomes, or how causal such mechanisms might play out variably in different contexts [29]. Analyses of mediation simply add links to the “if x then y” thinking commonly attributed to RCTs [30]. In this sense, perhaps many RCTs have, as Marchal et al. suggest, restricted themselves to identifying conjunctions between observable phenomena and have only engaged with theoretical concepts where these have empirical analogues.

However, this tendency is not universal. In the case of the ASSIST RCT, which did not explicitly embrace realist approaches, the use of statistical data as part of a hypothetico-deductive approach within trials did not preclude using other forms of evidence to assess the plausibility of theories about mechanisms. The embedded process evaluation drew on a range of data including qualitative research on teachers’ and students’ accounts of their own observations about how implementation processes occur and how outcomes might be generated [31, 32]. In the explicitly realist INCLUSIVE RCT, the intervention theory of change centred on how the intervention might enable an erosion of “boundaries” between staff and students and between students’ academic and broader learning, which then encourages more students to exert agency to commit to school and avoid engaging in risk behaviours such as violence that function as symbolic markers of anti-school identity. The theory thus included elements that were not open to quantitative measurement but which were nonetheless included in the theory of change to give a fuller account of the way in which the intervention was intended to work. Such work clearly does not fit with a positivist focus only on constant conjunctions, and will be considered in more detail below in our consideration of whether RCTs necessarily imply a unity of method.

It is also worth highlighting that it is not only RCTs that shed light on causality partly using statistical analyses of overall associations between exposure to interventions and outcomes. For example, the originators of realist evaluation positively cited an evaluation of the effect of prisoner education on reoffending rates, where the analysis compared rates of recidivism between the intervention group and a non-randomised historical comparison group made up of a cohort of individuals imprisoned prior to implementation of the intervention [13]. It is not clear why using statistical association data from randomised experiments as one way to assess the plausibility of causal mechanisms should be considered positivist, whereas drawing on evidence of statistical associations from natural experiments is not.

Do trials aim to produce universally applicable laws?

A central feature of positivism lies in its attempt to identify law-like regularities. Indeed Marchal et al. argued that RCTs are underpinned by Humean notions of constant conjunction, directed toward identifying interventions that are essentially linked to particular outcomes. We disagree that this is a necessary feature of trials and think that current practices among trialists instead suggest a mixed, and arguably inconsistent, set of beliefs.

As discussed above, trialists do not have an expectation of identifying constant conjunctions and hence universally applicable laws at the level of the individual. No-one who understands trial statistics could possibly believe that any intervention is expected by trialists to produce the same effects in different people. Furthermore, nearly all trial reports draw attention to the uncertain generalisability of RCT evidence across groups of individuals. Guidance for undertaking health RCTs [33] explicitly has acknowledged that results from a trial may be an uncertain guide to wider effects:

“External validity is a matter of judgment and depends on the characteristics of the participants included in the trial, the trial setting, the treatment regimens tested, and the outcomes assessed.” (p. 20-21)

Furthermore, when social interventions in public health or health services are transported from one setting or population to another, they are commonly subjected to a new RCT in the new situation prior to wider use. This suggests that those involved accept that evidence of effect in one context cannot unproblematically be accepted as evidence that the intervention will work in the same way in a new time and place. The Family Nurse Partnership demonstrated benefits when evaluated in the USA, but in England had no effect on smoking cessation, birthweight, rates of second pregnancies or emergency hospital visits for the child [34, 35].

It is also instructive to explore how systematic reviews approach the question of generalisability because such reviews bring together evidence from different settings. That systematic reviewers also are aware of the far from unproblematic generalisability of trial evidence is evidenced by their common practice of defining a priori inclusion criteria for reviews not only in terms of interventions and evaluation methods but also in terms of the populations and settings involved in studies [36]. Assessment tools used by systematic reviewers include judgements of issues such as “directness”, which refers to the extent to which the evidence within the review provides evidence of direct or indirect relevance to the context of interest [37].

However, we acknowledge that the picture is mixed regarding whether those doing and synthesizing RCTs believe their results to be relevant universally or only relevant context-specifically. Many RCTs have confined themselves to examining overall effects and have not explored how these effects are moderated by the characteristics of individuals receiving the intervention or settings in which the intervention is delivered. In the case of our most potentially positivist RCT case study, that of the CDP intervention, the trial assessed the intervention in terms of its overall effects, finding evidence of various benefits including students being more accepting of other students, less lonely or anxious, with increased problem-solving and resolution skills and prosocial behaviours [21, 22, 38, 39]. The research did not examine how outcomes varied other than by age [21, 22] not even assessing whether effects varied by sex other than in the case of one outcome measure of between-sex friendship nominations by sex [39]. The trial reports did not explicitly claim that the intervention would be effective in all populations and settings but did consider the implications of the trial results for theories of child prosocial development in a way that implies an assumption that the results were generalisable [21, 22]. The only reference to context was in the discussion of one paper [22] where there was reference to the intervention being effective despite being delivered in schools in middle/upper-class neighbourhoods where children may “not have exceptional problems with peer relations” (p.166). However, in the case of the more recent and much less positivistically inclined ASSIST RCT, as we have seen above, despite estimating the potential population impact of the intervention were it to be scaled up, the authors also reported how intervention effects might have varied for example with the structure of local communities, acknowledging that RCT results are not mechanistically generalisable across populations [27].

The way that many systematic reviews are conducted does suggest that their authors expect interventions to have broadly similar effects across quite widely differing

populations and settings. Most systematic reviews of social interventions in the fields of public health and health services, such as those done within the Cochrane Collaboration, have as their main focus general questions such as “do health promoting schools interventions promote children and young people’s health?” [40]. In such cases, although the research question defines a specific population (such as children and young people) and a specific setting (such as schools), there is often wide diversity within these populations and settings. Even when their research questions refer to more specific populations or settings, these are usually broad in scope, such as students in schools in low-income countries [41]. Such questions do not refer to the detail of contextual contingencies as realist evaluators would understand them. Systematic reviews often pool effect estimates from studies conducted across these defined but diverse populations and settings and use fixed effect models [42] implying the assumption that the pattern of cause and effect is the same across studies, with any differences in effect sizes being largely the result of chance. Evidence that this assumption might be unwarranted comes from recent research that has demonstrated that systematic reviews of complex interventions rarely provide a high level of certainty in effect estimates of complex interventions, in large part due to high levels of heterogeneity in effects [43]. Thus, we acknowledge that the current picture is mixed, with many systematic reviews in particular acknowledging that generalisability is uncertain but proceeding as though it is not.

However, this does not mean that this is the only or the best use of RCTs. A more productive alternative was obliquely suggested in the work of Karl Popper, one of the original key influences on evidence-based policy and an opponent of positivism, in his critique of “historicist” social science. By historicism, Popper was referring to theorists such as Hegel and Marx [44], who aimed to develop general laws explaining the historical evolution of society and thus to predict future developments. Popper argued that such theories focus superficially on trends and mistake these for laws of general determination:

“[historicists] overlook the dependence of trends on initial conditions. They operate with trends as if they were unconditional, like laws. Their confusion of laws with trends makes them believe in trends which are unconditional (and therefore general)” (p. 118)

Although the subject matter is very different, precisely the same criticism could be applied to how information on statistical trends from RCTs is often currently mistaken for laws of generalisation. Systematic reviews as they are generally conducted try to identify overall statistical trends, often failing to do so because findings are

heterogeneous [13]. But even when they do find consistent evidence of effect sizes [45], this is not an adequate form of generalisation because, like the historicists cited by Popper, statistical trends alone say nothing about the contextual contingencies that are likely to affect whether similar trends might be expected in other populations, times and places.

Popper argued instead that social science should aim to identify general mechanisms of causation but should theorise and then explore empirically how the consequences of these will be influenced by contextual contingencies. He was also clear that ultimately all such generalisations will be quite tentative because they are limited by the potential of humans to make their own decisions. From a thinker sometimes mistaken for a positivist [11], this is a remarkably similar approach to the realist focus on generative mechanisms and context-mechanism-outcome configurations. As Popper argued, while generalisations from social science will always be less definite than those from natural science because of human agency, the use of theories that include contextual contingencies has the potential to enable social scientists to develop more informed and more precisely worded forms of generalisation.

The INCLUSIVE RCT examined such questions. It focused on testing various a priori hypotheses about context-mechanism-outcome configurations informed by theory, such as whether intervention effects were greater in schools with more socioeconomically disadvantaged students (because the theory of change suggests that boundary erosion will have more impact on the engagement and hence the health outcomes of such students). It also drew on qualitative data collected as part of the process evaluation to develop new configurations, to be tested in post hoc analyses where pertinent quantitative data allow. Because it was a pragmatic effectiveness trial of how the intervention worked in a group of schools under real-world conditions, the INCLUSIVE trial should have included sufficient diversity in terms of intervention delivery and school settings and populations to examine a range of context-mechanism-outcome configurations [46]. Similarly, although not explicitly realist in its aims, the ASSIST RCT aimed to identify factors external to the intervention, which might affect its implementation and effectiveness [47]. Trial papers hypothesised how outcomes might vary by context on the basis of its theory and confirmed this to be the case in statistical analysis:

“Interventions for health promotion based on diffusing new behavioural norms might work best in clearly defined, fairly close-knit communities, such as those assumed to exist in the ex-coalfield communities of the Welsh valleys, since peer supporters are in

very regular contact with members of a community whose membership is well defined and stable. Analysis showed this notion to be true, with a substantially greater effect in students from valley schools than in those from other areas” [27]. (pp. 1599-1600)

Any single study will lack the statistical power and heterogeneity of context to explore every single context-mechanism-outcome configuration but there is no reason why this is more the case in experimental than quasi-experimental or before-and-after research. The extent to which every individual study should attempt to investigate all potential mechanisms and contextual contingencies is also highly questionable. Some analyses of how mechanisms interact with wide variations in context might be best left to evidence synthesis rather than each individual evaluation study [48, 49].

Do trials embody a unity of methods between the natural and social sciences?

RCTs may appear vulnerable to this charge because they are a design also used in natural sciences such as agriculture and pharmacology [50]. However, as we saw from the philosophy of science literature on positivism, unity of method applies not at the level of a particular research design but at the level of an overall approach to science. So the question should be, do RCTs serve a form of social science that is exclusively focused on statistical associations like agricultural or pharmacological trials, or can social science RCTs include distinctive elements?

RCTs, as we have already discussed, clearly do examine statistical associations between measured phenomena as one way of considering the plausibility of theories of causation. As mentioned earlier we chose the RCT of the CDP as an example because of its potential for adhering to some positivist tenets. The CDP trial involved no qualitative research aiming to understand the perspectives, motivations or agency of those involved in delivering or receiving the intervention. Although the trial involved interviews with the children participating in the programme, these interviews focused solely on structured assessments of their cognitive social problem solving and prosocial resolution skills [21].

However, many RCTs of social interventions such as those of the ASSIST and INCLUSIVE interventions also collect qualitative data [31, 32]. Qualitative analyses ongoing in the INCLUSIVE trial draw on interviews and focus groups to explore how those involved with the intervention described the context of implementation, the meaning of the intervention for them, their agency and decisions in delivering or receiving the intervention and the consequences of these decisions [51]. Like many contemporary process evaluations, this was guided by a

sociological framework which sensitised evaluators to the ways in which local actors make sense of interventions, commit to using them, work collaboratively with others to draw on intervention resources to act and then reflect critically on these processes to inform choices about subsequent actions [52, 53].

Trials like those of ASSIST and INCLUSIVE that include such components are thus not merely aiming to generate information about statistical associations but are also aiming to understand action in terms of meanings and agency, very much in the hermeneutic tradition of Max Weber. Findings from qualitative research can be used in different ways within RCTs [54]. They might be compared with quantitative results to contribute towards assessing the plausibility of causation or, as is the case in the INCLUSIVE trial, be used to refine theories delineating context-mechanism-outcome configurations prior to hypotheses arising from these being tested using quantitative evidence from RCTs [24, 55]. Or qualitative research might be analysed separately to develop a deeper view of people's experiences [51]. An example of this comes from the ASSIST RCT, in which qualitative data from teachers were used to understand some of the institutional barriers to implementation [32]. Qualitative data from students were used to explore how peer educators actively reinterpreted and reconstructed the intervention from one focused on prevention messages targeting the overall peer group, encompassing smokers and non-smokers, to one often restricted to providing information and targeting friends and predominantly those who have never smoked [31]. Thus, RCTs of social interventions can and do employ a multi-faceted approach that is distinctive from that of field trials of the biological effects of agricultural or pharmacological interventions.

While we have drawn on the work of Popper in the course of this paper to make the general case that an anti-positivist strand of thought has from the outset permeated social experimentation, Popper was in fact himself aligned with positivism on the specific question of the unity of method. He dismissed qualitative research as hopelessly unfocused in comparison to hypothetico-deductive science [56]. But here we depart from Popper in that we believe that qualitative research, which collects data in the form of participants' own accounts of their understandings and actions, and the consequences of these, can be crucial in helping social scientists refine their theories of how social mechanisms operate [57]. Adopting a realist rather than a positivist approach provides an appropriate framework for drawing on both quantitative and qualitative research because realist social science aims to explore cause and effect but also meaning and agency. As Blaikie suggests:

"Social objects cannot be studied in the same way as natural objects, but they can be studied 'scientifically' as social objects... social reality is pre-interpreted, ... society is both produced and reproduced by its members and is therefore both a condition, and an outcome of their activity. The social sciences have a subject-subject relationship to their subject matter rather than a subject-object one characteristic of the natural sciences... [W]hile sharing Positivism's desire for producing causal explanations and Interpretivism's view on the nature of social reality, Realism argues for a view of science that is very different from either of these approaches." (p. 59)

Discussion

The literature describing or criticizing RCTs of social interventions as being positivist was not comprehensive, and was in fact frequently inconsistent in defining what is meant by positivism. The literature on the philosophy of science provided a more consistent and comprehensive definition of positivism. It depicted this as involving a number of tenets: the epistemic primacy of direct sensory information as the basis for scientific knowledge; the requirement that theoretical terms must equate with empirical terms; the aim of developing universally applicable laws; and the unity of method between the natural and social sciences.

Our review of current practice in RCT research focused on social interventions in public health and health services suggests a mixed picture. It is very difficult to see how RCTs embody the epistemic primacy of sensory information. Instead RCTs appear to embrace Karl Popper's anti-positivist hypothetico-deductive approach to enquiry. Many RCTs appear to accept implicitly the requirement that theoretical terms are limited to those that are empirically measurable, for example by having logic models that are little more than strings of variables. However, RCTs are also now being done that employ more sophisticated theories of change, engaging with the deeper sociological mechanisms by which social interventions operate. Even if not all aspects of these mechanisms are directly measured in the realm of the "actual" they are nonetheless useful in formulating how causality actually operates in the realm of the real and therefore in informing more nuanced hypotheses to examine empirically. The picture is mixed as to whether those doing and synthesizing RCTs see their role as producing universal or contextually contingent generalisations. Although there is a tendency among many evidence producers and synthesisers to view their role as limited to the production of statistical trends as a form of generalisation, this is not an inevitable feature of all RCT-based research. The work of Karl Popper is again instructive,

suggesting the need to generalise on the basis not of trends but of theory that specifies what contextual contingencies will influence the way in which mechanisms generate outcomes. While RCTs might appear to embody a unity of scientific method in that they are also applied in some natural sciences, there are in fact important divergences. In many cases, trials of social interventions use distinctive methods that would never be used in the natural sciences, such as hermeneutically inclined qualitative research aiming to understand how interventions are interpreted and enacted locally.

We have provided suggestions for how RCTs can move beyond residual features of a positivist paradigm by focusing on the refinement and testing of theory concerning intervention mechanisms and the contextual contingencies affecting how these generate outcomes; examining not only overall effect sizes but also how these vary by context in order to test the plausibility of theory; accepting that generalisations are tentative and in the form of theories not merely statistical trends; and taking a distinctly social science approach to trials, which embraces qualitative data on participants' meanings and experiences alongside quantitative data on statistical associations.

Our suggestions for the non-positivist conduct of social experiments draw heavily on the work of Karl Popper. Popper's work is useful not only in providing suggestions for how to do social experimentation better, but also in illustrating that a non-positivist approach is not a recent attempt to redeem trials but in fact permeates the intellectual roots of RCTs. However, in our enthusiasm for using qualitative research within RCTs, we depart from Popper who dismissed the value of open-ended research and inductive analysis in building or refining scientific or social scientific theory. We believe that the most appropriate paradigm for RCTs of social interventions is realism. Karl Popper's brand of post-positivism and critical realism is united in viewing a world replete with causal mechanisms independent of our perceptions. Both also view human knowledge as an indirect representation of the world, and one that is infused with theory, fallible and provisional. Both recognise that positivism is redundant but reject the relativistic suggestions that no reality exists independent of our senses or that we cannot rationally judge between competing truth claims [58]. Although some realists appear to view realism and "post-positivism" as rival paradigms [29], like the co-formulator of the principles of realist evaluation, we regard realism as the pre-eminent post-positivist paradigm, and one that rejects both the crude empiricism and determinism of positivism while maintaining a commitment to developing empirically informed accounts of causal processes in a real world [59].

Conclusion

We hope we have demonstrated, as requested by our realist critics, that RCTs are not inherently aligned with a positivist philosophical position.

Abbreviations

ASSIST: A Stop Smoking in Schools Trial; CDP: Child Development Program; INCLUSIVE: Initiating change locally in bullying and aggression through the school environment; RCT: Randomised controlled trial

Acknowledgements

We would like to thank the other co-investigators working on the INCLUSIVE trial and the students and staff of the participating schools.

Funding

The INCLUSIVE trial was funded from a grant by the National Institute for Health Research Public Health Research Programme (grant 12/153/60). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the National Institute for Health Research Public Health Research Programme or the Department of Health. The work was undertaken with the support of The Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement (DECIPHer), a UKCRC Public Health Research Centre of Excellence. Joint funding (MR/K023233/1) from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, the Welsh Government and the Wellcome Trust, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

Availability of data and materials

All literature reviewed is publicly available.

Authors' contributions

CB drafted the paper and reviewed the literature. GM, EW and LM all contributed to editing the paper. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The paper was based on a literature review only and required no ethical review or any participant consent.

Consent for publication

No participant data are included.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹London School of Hygiene & Tropical Medicine, Faculty of Public Health and Policy, 15-17 Tavistock Place, London WC1H 9SH, UK. ²Cardiff University, School of Social Sciences, Cardiff, UK. ³University of Glasgow, Institute of Health & Wellbeing, Glasgow, UK.

Received: 19 January 2018 Accepted: 12 March 2018

Published online: 19 April 2018

References

1. 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.
2. Weinberger M, Oddone EZ, Henderson WG, Smith DM, Huey J, Giobbie-Hurder A, Feussner JR. Multisite randomized controlled trials in health services research: scientific challenges and operational issues. *Med Care*. 2001;39(6):627-34.
3. Banerjee AV, Dufo E. The experimental approach to development economics. *Annual Review of Economics*. 2009;1:151-78.
4. Torgerson CJ, Torgerson DJ, Birks YF, Porthouse J. A comparison of randomised controlled trials in health and education. *Br Educ Res J*. 2005;31:761-85.

5. Campbell DT. Methods for the experimenting society. *Am J Eval*. 1991; 12(3):223–60.
6. Popper K. The poverty of historicism. London: Routledge & Kegan Paul; 1957.
7. Green J, Thorogood N. Qualitative methods for health research. London: Sage; 2004.
8. Marchal B, Westhorp G, Wong G, Van Belle S, Greenhalgh T, Kegels G, Pawson R. Realist RCTs of complex interventions - an oxymoron. *Soc Sci Med*. 2013;94:124–8.
9. Pearce W, Raman S. The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance. *Policy Sci*. 2014;47:387–402.
10. Rowe M, Oltmann C. Randomised controlled trials in educational research: ontological and epistemological limitations. *Afr J Health Prof Educ*. 2016;8(1):6–8.
11. Tones K, Green J. Health Promotion: Planning and Strategies. London: Sage; 2004.
12. Bonell C, Fletcher A, Morton M, Lorenc T. Realist Randomised Controlled Trials: a new approach to evaluating complex public health interventions. *Soc Sci Med*. 2012;75(12):2299–306.
13. Pawson R, Tilley N. Realistic evaluation. London: Sage; 1997.
14. Bhaskar R. The possibility of naturalism: a philosophical critique of the contemporary human sciences. Brighton: Harvester; 1979.
15. Bhaskar R. A realist theory of science. Leeds: Leeds Books; 1975.
16. Blaikie N. Approaches to social enquiry. Cambridge: Polity Press; 1993.
17. Russell B. A history of Western philosophy. London: Simon & Schuster, Inc; 1945.
18. Hacking I. Representing and intervening: introductory topics in the philosophy of natural science. Cambridge: Cambridge University Press; 1983.
19. Kolakowski L. Positivist philosophy: from Hume to the Vienna circle. Harmondsworth: Penguin; 1972.
20. Weber M. The methodology of the social sciences. Glencoe: Free Press; 1949.
21. Solomon D, Watson MS, Delucchi KL, Schaps E, Battistich V. Enhancing children's prosocial behavior in the classroom. *Am Educ Res J*. 1988; 25(4):527–54.
22. Battistich V, Solomon D, Watson M, Solomon J, Schaps E. Effects of an elementary school program to enhance prosocial behavior on children's cognitive-social problem solving skills and strategies. *J Appl Dev Psychol*. 1989;10:147–69.
23. Bernstein B. Class, codes and control, vol. 3: towards a theory of educational transmission. London: Routledge; 1975.
24. Jamal F, Fletcher A, Shackleton A, Elbourne D, Viner R, Bonell C. The three stages of building and testing mid-level theories in a realist RCT: a case-example. *Trials*. 2015;16:466.
25. Markham WA, Aveyard P. A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice. *Soc Sci Med*. 2003;56(6):1209–20.
26. Rose G. The strategy of preventive medicine. Oxford: Oxford university Press; 1992.
27. Campbell R, Starkey F, Holliday J, Audrey S, Bloor M, Parry-Langdon N, Hughes R, Moore L. An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): a cluster randomised trial. *Lancet*. 2008;371:1595–602.
28. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, Moore L, O' Cathain A, Tinati T, Wight D. Process evaluation of complex interventions: Medical Research Council guidance. *Br Med J*. 2015;350:h1258.
29. Van Belle S, Wong G, Westhorp G, Pearson M, Emmel N, Manzano A, Marchal B. Can "realist" randomised controlled trials be genuinely realist? *Trials*. 2016;17:313.
30. Steckler L, Linnan A, editors. Process evaluation for public health interventions and research. San Francisco: Jossey-Bass; 2004.
31. Audrey S, Holliday J, Campbell R. It's good to talk: adolescent perspectives of an informal, peer-led intervention to reduce smoking. *Soc Sci Med*. 2006; 63(2):320–34.
32. Audrey S, Holliday J, Campbell R. Commitment and compatibility: teachers' perspectives on the implementation of an effective school-based, peer-led smoking intervention. *Health Educ J*. 2008;67(2):74–90.
33. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Br Med J*. 2010;340:c869.
34. Olds DL, Robinson J, O'Brien R, Luckey DW, Pettitt LM, Henderson CR, Ng RK, Sheff KL, Korfmacher J, Hiatt S, et al. Home visiting by paraprofessionals and by nurses: a randomized, controlled trial. *Pediatrics*. 2002;110(3):486–96.
35. Robling M, Bekkers MJ, Bell K, Butler CC, Cannings-John R, Channon S, Corbacho Martin B, Gregory JW, Hood K, Kemp A, et al. Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial. *Lancet*. 2016;387(10014):146–55.
36. Littell JH, Corcoran J, Pillai V. Systematic reviews and meta-analysis. Oxford: Oxford University Press; 2008.
37. Schünemann H, Brozek J, Guyatt G, Oxman AD. Introduction to GRADE handbook. In: Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Oxford: Cochrane Collaboration; 2013.
38. Battistich V, Solomon D, Delucchi KL. Effects of a program to enhance prosocial development on adjustment. Boston: Annual Meeting of the American Psychological Association; 1990.
39. Battistich V. Effects of a school-based program to enhance prosocial development on children's peer relations and social adjustment. *J Res Character Educ*. 2003;1:1–7.
40. Langford R, Campbell R, Magnus D, Bonell CP, Murphy SM, Waters E, Komro KA, Gibbs LF. The WHO health promoting school framework for improving the health and well-being of students and staff. *Cochrane Database Syst Rev*. 2011;1:CD008958. 2014.
41. Barry MM, Clarke AM, Jenkins R, Patel V. A systematic review of the effectiveness of mental health promotion interventions for young people in low and middle income countries. *BMC Public Health*. 2013;13:835.
42. Thomas RE, McLellan J, Perera R. Effectiveness of school-based smoking prevention curricula: systematic review and meta-analysis. *BMJ Open*. 2015;5:e006976.
43. Movsisyan A, Melendez-Torres GJ, Montgomery P. Outcomes in systematic reviews of complex interventions never reached "high" GRADE ratings when compared with those of simple interventions. *J Clin Epidemiol*. 2016;78:22–33.
44. Popper K. The open society and its enemies, Vol. 2: Hegel and Marx. London: Routledge & Kegan Paul; 1945.
45. Gardner F, Montgomery P, Knerr W. Transporting evidence-based parenting programs for child problem behavior (age 3–10) between countries: systematic review and meta-analysis. *J Clin Child Adolesc Psychol*. 2016; 45(6):749–62.
46. Bonell C, Allen E, Christie D, Elbourne D, Fletcher A, Grieve R, LeGood R, Mathiot A, Scott S, Wiggins M, et al. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial. *Trials*. 2014;15:381.
47. Starkey F, Moore L, Campbell R, Sidaway M, Bloor M. Rationale, design and conduct of a comprehensive evaluation of a school-based peer-led anti-smoking intervention in the UK: the ASSIST cluster randomised trial. *BMC Public Health*. 2005;5(1):43.
48. Ragin CC, Drass KA, Davey S. Fuzzy-set / qualitative comparative analysis 2.0. Tucson: Department of Sociology, University of Arizona; 2006.
49. Borenstein M, Hedges L, Higgins J, Rothstein H. Introduction to meta-analysis. Chichester: John Wiley & Sons Ltd; 2009.
50. Fisher R. The Design of experiments (5th edition). Edinburgh: Oliver & Boyd; 1949.
51. Oakley A, Strange V, Bonell C, Allen E, Stephenson J. Process evaluation in randomised controlled trials of complex interventions. *BMJ*. 2006;332(7538):413–6.
52. May C. Towards a general theory of implementation. *Implement Sci*. 2013;8:18.
53. Nilsen P. Making sense of implementation theories, models and frameworks. *Implement Sci*. 2015;10:53.
54. Snowdon C. Qualitative and mixed methods research in trials. *Trials*. 2015;16:558.
55. Oakley A, Strange V, Bonell C, Allen E, Stephenson J. Integrating process evaluation in the design of randomised controlled trials of complex interventions: the example of the RIPPLE Study. *Br Med J*. 2006;332:413–6.
56. Popper K. The logic of scientific discovery. London: Routledge & Kegan Paul; 1959.
57. Giddens A. The constitution of society. Cambridge: Polity Press; 1984.
58. Groff R. Critical realism, post-positivism, and the possibility of knowledge. New York: Routledge; 2004.
59. Pawson R. Evidence-based policy: a realist perspective. Thousand Oaks: Sage; 2006.
60. Bonell C, Fletcher A, Fitzgerald-Yau N, Hale D, Allen E, Elbourne D, Jones R, Bond L, Wiggins M, Miner A, et al. A pilot randomised controlled trial of the INCLUSIVE intervention for initiating change locally in bullying and aggression through the school environment: final report. *Health Technol Assess*. 2013; 19(53):1–109.

Learning Together Process Evaluation Protocol

Last updated: 1 April 2015

Table of contents

Executive summary	71
1. The Learning Together programme and its theory of change	73
Intervention theory	73
Intervention design and logic model	75
2. Existing process evaluations of similar interventions	79
Whole school interventions	79
Restorative practice in schools	80
3. Process evaluation frameworks and hypotheses	82
Implementation	82
Fidelity	82
Reach and acceptability	85
Mechanisms of action and contextual variations	86
Pre-hypothesised intervention mediators	88
Contextual factors that may affect implementation	89
Contextual factors that may modify intervention effects	90
Hypotheses of how the intervention might produce harms	90
Normalisation	92
Summary of research questions	92
4. Data collection methods for process evaluation	95
Appendix 1. Overview of data sources and how they pertain to each domain of evaluation	103
Appendix 2. Data analysis methods: Implementation	113
Appendix 3. Data analysis methods: Mechanisms of action	119
Appendix 4. Data analysis methods: Normalisation	130
References	131

Figures

Figure 1: Learning Together logic model	78
Figure 2: Evaluation Process	88

Tables

Table 1: Characteristics of case study schools	101
Table 2: Overview of all data sources for process evaluation	103

Table 3: Data sources for RQ on implementation.....	108
Table 4: Data sources for RQ on mechanisms of action	109
Table 5: Data sources for RQ on normalisation	111
Table 6: Data sources to assess: fidelity of form.....	113
Table 7: Needs assessment surveys.....	113
Table 8: Action group implementation across intervention schools	114
Table 9: Social and emotional curriculum implementation.....	114
Table 10: All-staff training implementation.....	115
Table 11: In-depth training implementation	115
Table 12: Data sources to assess: fidelity of function.....	116
Table 13: Restorative practices implemented	116
Table 14: Other locally decided actions.....	116
Table 15: Data sources to assess: reach and acceptability.....	117
Table 16: Students and staff involvement.....	118
Table 17: School-level mediation analysis	119
Table 18: Individual-level mediation analysis	122
Table 19: Data sources to assess: moderation (context)	124
Table 20: Data sources to assess potential harms and underlying mechanisms	127

Executive summary

Learning Together (LT) is a three year cluster randomised control trial (RCT) of a whole-school restorative approach to behaviour change, which aims to reduce bullying and aggression via: providing schools with facilitation for whole school organisational change directed by an action team of staff and students, as well as training for staff in restorative practice and the delivery of a social and emotional skills curriculum for students (at the end of year 7; age 11-12 years at baseline) in the trial cohort. The intervention combines pre-specified inputs, processes and outputs with the capacity for tailoring some elements to institutional needs and local ownership via decision making by staff and students on the action group.

The RCT will identify whether the actions of our intervention were effective in the time and place it was delivered, while the concurrent process evaluation, which this document reports, will allow us to interpret findings and understand how they might be applied elsewhere. By combining process evaluation with an RCT design we enable evaluators to limit biases in estimating effects, while developing the detailed understandings of causality that can support a policymaker, practitioner or systematic reviewer in interpreting effectiveness data (Craig et al. 2008; Moore et al, 2014). Thus, we recognise that effect sizes are important but alone are insufficient, and that process evaluation is necessary to understand implementation, causal mechanisms and contextual factors which shape outcomes (Craig et al. 2008, Moore et al. 2014).

Informed by MRC guidance (Moore et al, 2014; Craig et al 2008) and the wider implementation science literature (e.g. May and Finch 2009; Bumbarger and Perkins, 2009; Hawe, et al. 2009) this process evaluation investigates the following domains:

- implementation
- mechanisms of impact and context
- normalisation

Section 1 and 2 of the report describe the development of the intervention logic and assumptions. In **Section 1** we describe the theory of change. This includes inputs that the intervention involves the processes that these initiate and the mechanisms via which these are intended to realise positive outcomes. The theory of change sits at the heart of the evaluation, informing collection of data on likely causal pathways, and how these vary according to individual and contextual factors, which in turn feed back into refining our theory of change post evaluation. **Section 2** describes previous programmes that have informed the LT intervention, including the pilot trial. **Section 3** sets out the process evaluation framework and hypothesis. This is divided into three sections pertaining to the three domains of evaluation: implementation (fidelity of form and fidelity of function; and reach and acceptability); mechanisms of impact and context; and normalisation. Within each domain we describe our approach to evaluation, the hypotheses to be examined and the relevant data sources used for analysis. At the end of this section we collate a list of research questions pertaining to each domain of investigation. **Section 4** summarises the sampling and data collection we will conduct as part of the evaluation. We provide an overview of all data sources and how these map across

research questions in **Appendix 1** before describing the methods of analysis in **Appendices 3-5**.

1. The Learning Together programme and its theory of change

Learning Together (LT) is a three year whole-school restorative approach to behaviour change, which aims to reduce bullying and aggression via: 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 school and reduce aggression; whole-school staff training in restorative practices; and a new social and emotional skills curriculum for students in the trial cohort (at the end of year 7; age 11-12 years at baseline). The intervention combines pre-specified inputs, processes and outputs with the capacity for tailoring some elements to institutional needs and local ownership via decision making by staff and students on the action group.

All interventions can be described as ‘theories incarnate’ (Pawson and Tilley, 1997) in that they reflect assumptions regarding the causes of a problem and how actions will produce change. Complex interventions such as LT reflect many causal assumptions. We make our intervention assumptions clear via a theory of change presented in a (diagrammatic) logic model of inputs that the intervention involves, the processes that these initiate and the mechanisms via which these are intended to realise positive outcomes (see **Figure 1**). Our theory of change sits at the heart of the evaluation, informing collection of data on likely causal pathways, and how these vary according to individual and contextual factors, which in turn feed back into refining our theory of change post evaluation. First, we describe the theory that informs our intervention and how we have applied it to develop an intervention logic model.

Intervention theory

Our theory of change draws predominantly upon sociological theory, focusing on system level change. It starts from the theoretical position that schools have a wide-ranging influence on student behaviours. Informed by Markham and Aveyard’s (2003) theory of human functioning, school organisation and health promoting schools, our intervention theory suggests a person’s autonomy to make and enact good decisions is a necessary precondition for healthy behaviour so in order to promote health, schools must enable students to develop autonomy. To achieve autonomy, people have various needs which must be met and capacities which must be enabled. Enabling people to develop capacities for ‘practical reasoning’ and for ‘affiliation’ are most crucial since the fulfilment of all other needs and capacity will require a person to be able to think and form relationships.

These capacities, according to the theory, are facilitated by greater commitment to what Bernstein (1975) termed schools’ instructional and regulatory orders. The instructional order focuses on the relaying of knowledge and skills, and is concerned with students’ ability to contribute to future production through work. The regulatory order focuses on the relaying of values and is concerned with the conduct, character, and manner of students while they are at school and after they have left. If students accept and meet the demands of both the instructional and regulatory orders, they are termed ‘committed’. Committed students have the

greatest opportunity to use school to promote good human functioning and health. If students reject or cannot meet the demands of both the instructional and regulatory orders they are termed 'alienated'. Alienated students experience restricted opportunities to develop good human functioning and hence health.

However, Markham and Aveyard's theory (2003) has been critiqued for not sufficiently acknowledging that students may engage in risk behaviours not only because of deficits in practical reasoning and affiliation but also by students developing practical reasoning and affiliation developed via anti-school peer groups and directed towards anti-school actions which might include risk behaviours, such as verbal bullying and violence (Jamal et al. 2013). Thus, informed by qualitative studies (Fletcher 2009; Paulle 2013; Cousins 1997; Waldron 2009; Bourgois 1995; Dance 2002), our theory of change hypothesised that some students who are not committed to schools' instructional and regulatory orders may engage in risk behaviours either because they have deficits in pro-school practical reasoning and affiliation, or because they are committed to anti-school peer groups, and these commitments encourage and enable students to engage in risk behaviours in various ways. The risk behaviours thus ultimately reflect deficits in commitment to school, but this association may be mediated by agency within a context of structural constraint and not merely by a lack of informed agency.

In addition to students' own agency, the social background of the student and the culture of the school (e.g. the methods used to convey instructional and regulatory orders) also influence young people's responses to the instructional and regulatory orders. Students from middle class backgrounds are more likely to be committed, while those from working class backgrounds are more likely to be alienated. However, schools can influence the proportions of committed and alienated students. It is theorised that if schools reduce barriers between the school and the communities it serves, between students and teachers, between student groups, and between subjects (together termed 'boundaries'), and if they increase students' input and control over learning (termed 'framing'), proportionately more students can become committed rather than alienated, even when accounting for students' social class background (Markham and Aveyard 2003).

The theory suggests that commitment might be achieved by schools implementing organisational approaches, policies and practices which erode various 'boundaries' within the school between:

- staff so authority is distributed rather than concentrated among senior staff;
- staff and students so relationships are collaborative rather than authoritarian;
- between students so positive relationships are encouraged and students are treated equitably;
- different areas of students' life, so teachers focus on students' overall wellbeing and development rather than merely academic progress, and support is provided across the whole school rather than merely in the classroom; and
- the school and its local community.

For example, boundaries between teachers and students might prevent students from being involved in decision making (Markham and Aveyard 2003). Involvement of students in school decision making (e.g. via school council) might weaken these boundaries and promote greater insights into the realities of staff and young people's lives. This would facilitate the realisation of the capacity for practical reasoning. There may also be strong boundaries between students, fuelled by social hierarchies for example, which facilitate division and subordination (Bernstein 1975). These boundaries between students can be weakened through greater communication, shared tasks and greater co-operation, which facilitates the development of insights into multiple realities and hence, realisation of the capacity for practical reasoning (Markham and Aveyard 2003).

The LT intervention aims to help schools build student commitment to their instructional and regulatory orders by modifying schools' systems of 'boundaries' and 'framing'. However, previous evaluations of whole-school interventions (Bonell et al. 2010) suggests this intervention is unlikely to transform the ethos of the whole school since this is often strongly determined by the values and priorities of the senior leadership team (SLT). Therefore, we hypothesise that the intervention will enable the SLT to develop those aspects of the school ethos which they are already committed to developing, but haven't yet had the capacity to do so, and which are consistent with the ethos of the intervention.

Intervention design and logic model

Standardised activities leading to organisational change in schools include training in restorative practice, delivery of the social and emotional curriculum and implementation of action group meetings. However, the precise methods of delivery and resultant changes are intended to vary between schools (with the exception of restorative training which is intended to be standardised for all intervention schools). The extent to which implementation of the structures and processes triggers changes in schools' practices and ethos are presented as key pathways linking intervention inputs to student health outcomes.

We will facilitate school staff and students to participate collaboratively in a number of 'whole-school' activities. An action team, comprising staff and students will work to revise school policies and oversee all intervention activities. Staff and students will collaborate to re-write school policies and rules. These activities are intended to erode boundaries between students and staff by enabling them to work together to make decisions about their school. All staff will be involved in training to introduce them to concepts of restorative practices, and 8-10 staff per school will also be involved in in-depth restorative training focusing on methods and procedures for applying restorative practice with students in schools. Such activities will aim to transcend the boundaries between staff and students, and among students, and therefore build student commitment to the school's regulatory order. They will also aim to re-frame communication and decision-making regarding how schools are administered from being purely staff-led to involving students as well. This should in turn enable students (and staff) who participate to develop their practical reasoning (for example, understanding other people's perspectives and appreciating where they share values, and thinking through how to resolve

differences amicably where they differ) and sense of affiliation and, ultimately, the development of socially valued forms of pro-school identity.

We also anticipate that the products of these processes will also bring benefits. For example we hope that students in general will be more prepared to accept and abide by rules which students have been involved in developing and that this will increase commitment to the school's regulatory order. Similarly we anticipate that the use of restorative approaches will be more effective in preventing damage to student relationships, again enhancing the school's regulatory order and increasing student commitment to school.

We will train school staff to use restorative practices to re-frame how they manage their classrooms. This will again aim to reduce boundaries between staff and students and boundaries between students, promoting student commitment to the school's instructional and regulatory orders and development of capacities for practical reasoning and affiliation. We will also train teachers to enable them to use restorative practices to repair relationships between students and between staff and students where these are harmed. This should, as above, enable students to develop practical reasoning (for example, think more clearly about their own and other's perspectives, empathise with others and learn how to manage their own emotions and resolve differences amicably) and develop a stronger sense of affiliation with other members within the school community, eroding boundaries between staff and students and among staff.

We will also train some school staff to deliver a social/emotional skills curriculum which aims to enhance the school's instructional order and enable students to develop a better understanding of their own and others' emotions, think about others' feelings and better manage social relationships. We anticipate that as well as enabling students to develop their capacities for practical reasoning regarding social and emotional skills and affiliation, this training may also more generally facilitate learning and school engagement more broadly. This is especially the case in those schools that decide to provide our curriculum by integrating it across a number of subject areas such as English literature and personal, social and health education. This may therefore reduce boundaries between academic subjects, further encouraging student commitment to the school's instructional order and their capacity for practical reasoning.

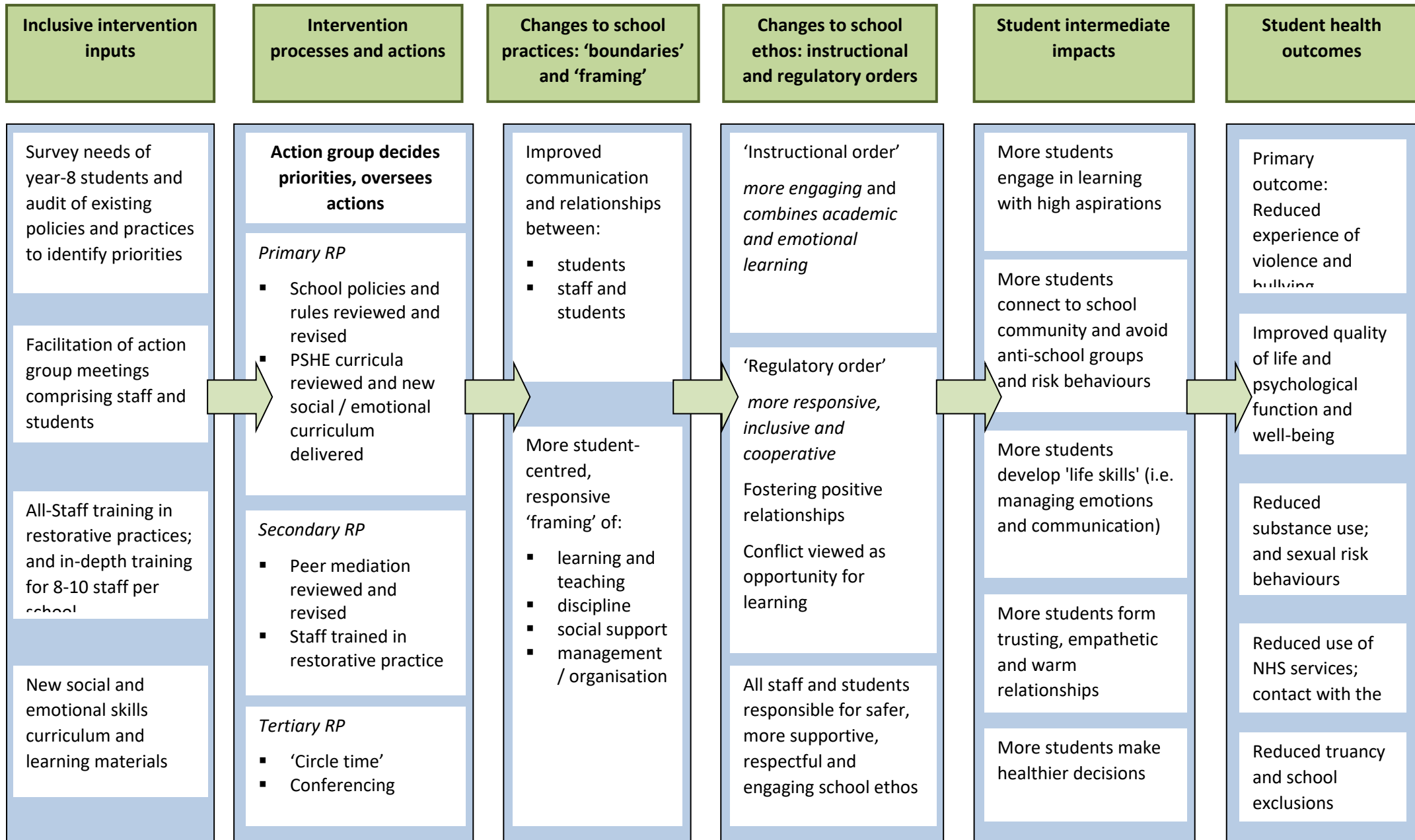
These components aim to operate synergistically. For example, the social/emotional skills component in enabling the development of practical reasoning in the realms of social and emotional skills will also empower students to participate more effectively in whole-school actions, while the improvements in classroom management will promote commitment to learning which will encourage students to be receptive to the curriculum and restorative interventions. Actions taken to align schools policies and practices with the principles of restorative practice should also support teacher application of these principles on a day-to-day basis.

In turn, by enabling students to develop their capacities for practical reasoning and affiliation, our intervention should increase students' abilities to avoid or manage conflicts with others

without recourse to aggression, manage their own emotional and mental health, develop more mutually supportive relations with other staff and students and avoid engaging in self-harming behaviours. It should then reduce the numbers of students who feel alienated or detached from school and instead invest in anti-school peer groups and anti-social behaviours such as aggression and substance use. It may also reduce the number of students who feel estranged from learning and have low self-esteem as a result of their failure to meet the challenges of the school's instructional order. It should increase the number of students feeling committed to school and safe and supported within it. These students are more likely to internalise pro-school values and norms, and develop high aspirations. Through their participation in our whole-school actions and more generally in school communal life they will develop socially valued identities, high self esteem and reduced levels of emotional distress, directly and also as a result of reduced experience of aggression and substance use.

These processes are set out diagrammatically in the logic model (**Figure 1**) for the LT intervention. It provides a general overview of the pathways involved, delineating key concepts. It attempts to put Markham and Aveyard's (2003) theory in language that is accessible to the staff and students of schools. The logic model shown in **Figure 1** presents our theory of change in a way that funders, providers and schools would understand. Therefore rather than discussing staff-student and student-student boundaries, it discusses improved communication among students and between students and staff. It doesn't describe at all the intervention's aim to erode boundaries between academic learning and broader development but does describe how the intervention aims to render learning and teaching, discipline, social support as well as management and organisation at schools in a more student-centred manner. Rather than discussing commitment to the instructional and regulatory orders it describes the intervention as aiming to engage more students with learning and to connect more students to the school community. Rather than engaging with the terms practical reasoning and affiliation it describes how the intervention aims not only to promote students' life skills and warm, trusting and empathetic relationships. It does not aim to provide an exhaustive set of all pathways, which will be multiple, potentially additive and involve complex feedback loops. The evaluation will refine this model and delineate the most important specific pathways and loops.

Figure 1: Learning Together logic model



2. Existing process evaluations of similar interventions

Whole school interventions

The LT intervention has been particularly informed by two previous programmes. First, the Aban Aya Youth Project (AAYP) is a multi-component intervention, enabling schools to modify their social environment as well as delivering a social skills curriculum. This approach was 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 in violence and aggressive behaviour (Flay et al 2004). The intervention also brought benefits in terms of reduced sexual risk behaviours and drug use, as well as provoking behaviour and school delinquency. A three-arm trial design established that the full intervention including environmental and curriculum components was more effective than intervention with the curriculum alone (Flay et al 2004). No process evaluation was conducted.

Second, the Gatehouse Project in Australia also aimed to reduce health problems via 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 randomised controlled trial (RCT) found consistent reductions in a composite measure of health risk behaviours, which included violence and anti-social behaviour (Bond et al 2004; Patton et al 2006). Process evaluation of the Gatehouse project found the various components (needs-survey, action-team, critical friend) functioned synergistically, and although specific actions varied between schools, these were well completed. Implementation was facilitated by supportive management and broad participation (Bond et al., 2001; Glover and Butler, 2004). However, this evaluation did not attempt to assess systematically how completeness of implementation might have been influenced by schools’ baseline social climate or “ethos”, i.e. the contextual characteristics specific to the school that distinguish it from other schools (Rutter et al., 1979; Gittelsohn et al., 2003).

The Healthy School Ethos intervention was piloted in English schools. Using a structured process modelled closely on the Gatehouse project, it aimed to enable each school to carry out locally determined actions to increase students’ security, positive self-regard and communication with staff and students. The intervention provided an external facilitator, data on student needs survey and training; and enabled schools to convene action-teams to determine priorities and ensure delivery. Students and staff coproduced rules for appropriate conduct and revised policies on bullying and student feedback. Staff were trained to improve classroom management. There was no curriculum component. Process evaluation (Bonell et al 2010b) reported that the intervention was delivered with good fidelity. Locally determined actions (e.g. peer-mediators) were generally more popular than mandatory actions. Implementation was more feasible where it built on aspects of schools’ baseline ethos and where senior staff led actions. Student awareness of the intervention was high. Student accounts

suggested benefits might arise as much from participation in intervention processes such as rewriting rules as from the effects of subsequent actions. Some processes could be made to reach significant number of students (Bonell et al 2010b).

Before this phase III trial, the LT intervention was piloted in four schools (Fletcher et al in press). Overall, school staff members were consistently supportive. Although some schools were already deploying some restorative approaches, it was nonetheless attractive because it enabled restorative practices to be delivered more coherently and consistently across the school. The adaptability of the intervention, in contrast to overly prescriptive, 'one-size-fits-all' interventions, was also a strong motivating force and source of acceptability to school managers. Staff valued the 'external push' which was provided by the external facilitator. The intervention was highly acceptable to school staff because of its fit with national policies and school metrics focused on attendance and exclusions. Some staff reported that it took time for them to understand how the various intervention components joined up and this could have been better explained from the outset. Staff were positive about sustainability, some reporting that activities would continue even after the pilot ceased.

Regarding particular components, staff reported that the needs-assessment survey allowed them to see the 'big picture' and identify priorities, but some suggested that the needs assessment could also feel too 'negative' at times, especially among established staff who could view this as a reflection on their work at the school. Negative aspects of the needs reports could also present problems for schools because if inspected by Ofsted they would be expected to share results with inspectors. As with the HSE evaluation (Bonell et al., 2010b), action groups were positively viewed, and it was suggested that student participation may be an active ingredient in improving relationships and engagement across the school, particularly when these involved students who might be less committed to school and involved in anti-school peer groups. Again, the presence of SLT on the group was seen as critical to driving actions. The training was more critically received, with many staff suggesting this was too didactic and contained too few examples from secondary schools. All schools successfully implemented the curriculum, welcoming its flexibility whereby modules could be implemented using the newly provided or existing materials.

The pilot did not examine causal mechanisms in terms of whether the intervention reduced school boundaries and increased student commitment, practical reasoning and affiliation. The pilot also lacked a large enough sample to examine how implementation and processes might vary across a range of different school contexts. The pilot also focused only on the first year of implementation and so could not examine the processes by which the intervention might become normalised within schools' institutional policies and practices and sustained once external facilitation is withdrawn.

Restorative practice in schools

Restorative practice is a key component of the intervention. The central tenet of restorative approaches is to repair the harms caused to relationships and communities rather than merely assign blame and enact punishment (Morrison 2005). 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. Restorative practice calls for a paradigm

shift away from punitive approaches to addressing poor behaviours (such as bullying, violence, and aggression) to an approach that restores relationships between victim, offender and the (school) community (Bazemore, 2001). Repairing harm involves active involvement of those most affected by the harmful action to come together and respond by developing a reparative plan. The restorative framework also includes consequences, sanctions, service and apologies, which are grounded in a commitment to change relationships and roles (Bazemore, 2001).

Restorative practices in schools encompass a continuum from using a restorative mindset, practicing affective language, conducting circle time to facilitated meetings and restorative conferencing. Affective language refers to the respectful use of language to challenge or support behaviour in a manner that preserves or enhances the relationship. Circle time involves classes coming together to check in, discussing their feelings and airing any problems so that these may be addressed before they escalate. Restorative conferences involve the parties to a conflict, dispute or crime being invited to a facilitated face-to-face meeting to discuss the facts that led up to the incident, for the harmed/aggrieved person or persons to explain what has happened to them as a result of the incident, and for the perpetrator(s) to take responsibility for their actions.

Most proponents of restorative practices in schools argue that restorative approaches work best when a positive ethos has been established, and when one-to-one problem-solving skills (such as listening and responsibility) have been introduced into the curriculum (YJB 2004). Evaluations suggest that this works best when the head teacher is fully supportive (Bitel 2001).

The theoretical basis for restorative approaches shares much in common with the theory of human functioning and school organisation. It is theorised that the process of students coming together, discussing the harm and working towards a reparative plan develops perpetrators' competency via accepting responsibility for the actions and contributing to a reparative solution, and develops offender understanding of the realities of others. Victims are also empowered in this process as they become an active participant in the decision-making process and the acknowledgment of the offenders' ability to offer some healing to the victim (e.g. via an apology or carrying out a sanction) gives dignity to both parties (Bazemore 2001; Pepi 1998). This resonates with the ideas of reducing boundaries, as well as promoting practical reasoning and sense of affiliation. By eliciting accountability for the harm caused to the victim and the school community and negotiating a plan for restitution or making right the wrong, the young person is encouraged in reclaiming an identity as a participant of the school community, not a peripheral outsider (Bazemore 2001). Through this process, the young people involved develop relational competency, reduce 'boundaries' between students and staff and students, as well as reduce 'framing'. Restorative approaches might indeed be particularly suitable for 'alienated' student offenders as they are given the opportunity to develop the necessary competencies to participate as a responsible member of the school community (towards restitution), which the student may have felt previously excluded (Pepi 1998). It may also be particularly helpful for female young people as gender theory suggests that female adolescent identity is often based within a framework of relationship and connection. Thus application of the principles of restorative approaches becomes a natural adjunct to the therapeutic process of self-identity and growth (Pepi 1998).

No RCTs of restorative practice in schools have been conducted but non-random studies suggest plausible benefits. A non-randomised national evaluation in England found no significant differences in student reported outcomes between intervention and comparison schools. However, teachers in programme schools reported better behaviour in intervention than control schools (YJB 2004). A non-randomised evaluation of restorative practices in Scottish schools (Kean et al 2007) concluded that implementation was better in schools: where the head teacher supported the work; where there was an agreed plan with clear objectives; where a broadly child-centred culture already existing; where restorative approaches joined up with other initiatives and policies; and where staff morale and interest in restorative approaches to discipline was already high. A non-randomised study in Bristol found that restorative practices were associated with improved attendance and reduced fixed-term exclusions, with staff reporting such changes would improve attainment through reduced disruption (Skinns et al 2009). A quasi-experimental evaluation of restorative approaches in two Durham secondary schools reported that staff and students preferred restorative practices to traditional forms of discipline (Kokotsaki et al 2013).

3. Process evaluation frameworks and hypotheses

RCTs identify whether interventions are effective in the time and place when they are delivered. Process evaluation linked to RCTs allows fuller interpretation of these findings and understanding of how they might be applied elsewhere. Combining process evaluations with an RCT design enables evaluators to limit biases in estimating effects, while developing the detailed understandings of context, implementation and mechanism that can support a policymaker, practitioner or systematic reviewer in interpreting effectiveness data (Craig et al. 2008; Moore et al. 2014). Thus, estimating effect sizes are important but alone are insufficient to inform future policy and practice. Process evaluation is necessary to understand the factors which shape outcomes (Craig et al. 2008, Moore et al. 2014).

Informed by the MRC guidance on process evaluation (Moore et al, 2014) and developing and evaluating complex interventions (Craig et al, 2008), as well as the wider implementation science literature (e.g. May and Finch 2009; Bumbarger and Perkins, 2009; Hawe, et al. 2009), our process evaluation investigates the following domains:

- implementation
- mechanisms of impact and context
- normalisation

Implementation

There are two broad areas of interest in terms of implementation: 1) fidelity; and 2) reach and acceptability.

Fidelity

An understanding of what is implemented and how is integral to explaining how an intervention works. The principal concern of early process evaluation frameworks was capturing what was delivered in practice in order to avoid type III error: dismissal of sound intervention theories due to a failure to implement them effectively (Steckler and Linnan 2002). Most frameworks focus on the precise 'form' of delivery in terms of whether this represents fidelity to what was intended to be delivered, as well as measuring the dose and reach of delivery. There is debate amongst prevention scientists and practitioners about whether adaptations in programme delivery, decided locally, enhance intervention effectiveness or lead to poorer outcomes (see for example, Bernal and Saez-Santiago 2006; Castro et al. 2004; Elliott and Mihalic 2004).

Advocates of strict fidelity (Mihalic 2004) argue that this is essential if effective interventions are to be replicated, especially when an intervention's 'active ingredient' may not be known. They present evidence which suggests that high fidelity is associated with greater impact for some interventions (Mihalic 2004). Advocates of local adaptation (Dane and Schneider 1998) argue that interventions need to be tailored to local circumstances. Durlak and DuPre (2008) propose a compromise whereby an intervention's 'core components' should be delivered in standard form but less central intervention components or features can be modified to fit local needs. They present research which suggests that a balance between fidelity and adaptation is likely to be most effective, with the precise balance dependant on the specific intervention. Hawe and colleagues (2004) go further to argue that with complex interventions characterised by synergistic interactions between multiple components, too much attention has been paid to standardizing the *form* of intervention components (what precise actions are delivered) and more attention should be given to standardizing the *function* of intervention components (what the aim of each component is and how it contributes to the overall theory of change of the intervention).

The LT intervention includes standardised inputs and processes alongside some degree of local tailoring. Where activities and outputs are locally tailored these are nonetheless intended to serve similar 'functions', i.e. to initiate a set of casual mechanisms that are pre-hypothesised within our theory of change. Therefore, our approach to evaluating implementation considers both: fidelity of form and fidelity of function.

Fidelity of form refers to the standardised features of our intervention. Standard structures and processes of the LT intervention include the development of an action group of school staff and students; delivery of a new social and emotional curriculum; and staff training in restorative practice. We hypothesise that, for the intervention to be optimally effective, the essential elements of each component should be delivered in all intervention schools. We will examine the extent to which the following standardised intervention components of LT were implemented with fidelity of form:

- needs assessment surveys;
- training;
- action group meetings;
- review of needs data;
- development of an action plan based on needs data;

- review of policies relevant to aggression and bullying;
- rewriting of school rules; and
- social and emotional skills curriculum.

We will draw on the intervention manual to define what fidelity of form involves. Where we note that deviations from fidelity of form occurred we will, where possible, assess whether these were:

- intentional adaptations (and if so what motivated it), unintentional drift or simple omission;
- whether the adaptation runs with or against the logic of our theory of change.

It is unlikely that we will be able to examine every deviation from fidelity of form but where possible we will assess these in interviews with facilitators and school staff, particularly in the six intervention schools that have been selected for in-depth 'case studies' (see Section 4, Table 1 for details).

Fidelity of function refers to the extent to which locally decided actions are consistent with a theory of change. Many of the activities which are decided and implemented by the action group will differ across intervention schools but should contribute towards our theory of change. For example, action group meetings in each school will locally decide what actions and priorities to implement for school organisational change suited to the particular needs, desires and context of each school; schools will also be able to choose which curriculum units to deliver (with the exception of unit 1 which is mandatory) based on local student needs and the methods of delivering these (e.g. subject area, materials used etc).

While the form of these activities varies across schools, they are meant to serve similar functions. Bumbarger and Perkins (2008) argue that evaluators need to distinguish between intended tailoring and unintended drift. In assessing the fidelity of function of intervention activities that are intended to be locally tailored, we will assess whether these retain fidelity of function or whether they depart from our theory of change so that they are best considered as 'drift'. Furthermore, we will apply this distinction to cases where elements of our intervention that we stipulate to be standardised are nonetheless locally adapted. In such cases we will assess whether in departing from fidelity of form these nonetheless retain fidelity of function or whether they drift from the theory of change so that they represent fidelity neither of form nor function.

We will examine the extent to which the following locally tailored intervention components of LT were implemented with fidelity of function:

- local actions included in action plan; and
- locally decided means of implementing restorative approaches.

We will monitor the extent to which these locally decided actions are implemented. We will scrutinise their logic to determine whether they run with or against the logic of our theory of change to determine fidelity of function. Where possible we will also assess the underlying logic of these local decisions in interviews with facilitators and school staff, particularly in case studies.

We will examine how fidelity varied across schools and what contextual factors appeared to affect this. In the section below on mechanism of action and context we develop a priori hypotheses to guide this.

Reach and acceptability

Reach is the extent to which the target audience come into contact with the intervention (Moore et al 2014). Some process evaluation frameworks incorporate 'reach' or 'coverage' as an integral part of implementation (Baranowski and Stables 2000; Carroll et al 2007; Steckler and Linnan 2002), while others see it as a separate dimension (Glasgow et al. 1999). Moore et al (2014) recommend that process evaluations should include quantitative assessments of reach, in terms of, for example, proportions of the target audience who came into contact with the intervention.

Acceptability refers to how intervention participants, providers or other stakeholders received or engaged with the intervention (Bonell et al 2006). Within Steckler and Linnan's (2002) framework, participant responses to an intervention are largely discussed in terms of 'dose received'. However, this incorrectly suggests that participants passively receive interventions, when in most cases they exercise agency in interacting with them. The term 'dose' also implies a privileging of quantitative measurements whereas acceptability can be examined quantitatively and qualitatively. Acceptability also needs to be considered as a dynamic characteristic. Interventions may initially raise resistance which in some cases dissipates with skilful delivery (Moore et al 2014; Grant et al 2013a).

We will examine reach and acceptability of the following aspects of LT:

- needs survey;
- Action Groups;
- rewriting school rules;
- social and emotional skills curriculum;
- training in restorative approaches;
- implementation of restorative approaches; and
- locally decided actions.

We will examine how reach and acceptability varied across schools and what contextual factors appeared to affect this. In the section below on mechanism of action and context we develop a priori hypotheses to guide this.

Mechanisms of action and contextual variations

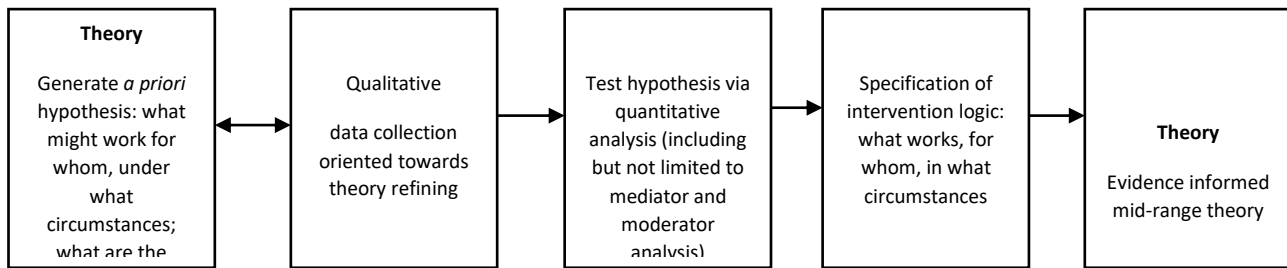
We draw on realist evaluation principles posited by Pawson and Tilley (1997) to investigate how the LT intervention works. Realist evaluation views interventions as ‘working’ by introducing *mechanisms* that are sufficiently suited to their *context* to produce *outcomes* (Pawson and Tilley 1997). Realists assume that interventions ‘work’ by enabling participants to make different choices. This might involve changes to participants’ cognitions or the opportunities and constraints that are present in their environment. A key aspect of realist evaluation is to anticipate the diversity of potential intervention mechanisms, to present this in a theory of change and to assess empirically whether and how these mechanisms are ‘enabled’ or ‘disabled’ in the varying contexts in which the intervention is delivered. Context refers to the pre-existing set of social situations, norms, values and inter-relationships (e.g. organisational structure, geographic location, demographics of participants) within which an intervention is implemented. Thus, the evaluator needs to hypothesise both the intervention theory of change and how this interacts with context to enable (or disable) implementation and outcomes. Pawson and Tilley (1997) suggest such hypothesis building proceeds via defining ‘context-mechanism-outcome’ (CMO) configurations.

Building on realist methods, Bonell et al (2014) have argued for the need for intervention theory to pre-hypothesise potential harms alongside potential benefits. Mechanisms of harm are not obvious and are not necessarily merely the converse of intended intervention mechanisms of action. There is a need for a priori theorisation of *potential harms* and their underlying mechanisms (Bonell et al. 2014). Our definition of harms includes what pharmacologists term “paradoxical effects”, i.e. intervention increasing the adverse outcomes they seek to prevent (Smith 2012) as well as “harmful externalities” where interventions produce harms in outcomes they are not aiming to prevent.

It is useful to think of the evaluation of mechanisms of action (underlying benefits and harms) as a series of steps. First, we will build hypotheses about how intervention mechanisms play out differently in different contexts. This will be informed by our intervention theory of change, the intervention pilot study (Fletcher et al. in press) and existing research (Moore et al, 2014; Craig et al 2008; May and Finch 2009; Bumbarger and Perkins, 2009; Hawe, et al. 2009). Currently this is an entirely linear model with no consideration of how the theorised pathways will vary by context or how they might produce harms. We will develop hypotheses about the influence of context as well as potential harms informed by the theory of human functioning and school organisation, existing process evaluations reviewed earlier, as well as from the wider sociological literature on schools. Second, we will use emerging data on the implementation and receipt of LT (primarily but not exclusively qualitative) to add to or amend the hypotheses developed in step 1. Qualitative research captures a sense of research participants’ own meanings, their sense of agency and how this inter-relates with the social structure of intervention context, and thus is useful for identifying pathways that were not originally anticipated in our theory of change. Third, we will test the hypotheses developed in steps 1 and 2 via quantitative analysis drawing on process and outcome data, for example of effect moderation and mediation. Fourth, informed by these analyses, we will draw conclusions about the contexts (person and place) under which the various intervention mechanisms appear to

produce benefits or harms. Finally, we will refine our theory of change in the light of these conclusions. Our aim thus is not merely to assess whether LT is an effective intervention or not, but to develop empirically informed mid-range theory (Merton 1949; Bonell et al. 2012) about school processes and how these may be modified by intervention, and the extent to which LT may be transferable to a range of contexts (Bonell 2006).

Figure 2: Evaluation Process



By putting theory of change at the heart of our evaluation, we aim to maximize the usefulness and generalisability of our evaluation findings. As well developing new hypotheses during the evaluation via the process described above, we have listed below our *a priori* hypotheses about the factors we think will mediate intervention effects on our primary and secondary outcome measures, how contextual factors may affect implementation and intervention effects, and how these contextual variations may directly modify the strength or direction of effects. Finally, we also pre-specify the mechanisms via which the intervention may produce harmful effects in order to also examine these *a priori*.

Pre-hypothesised intervention mediators

The main trial analyses of intervention effects on primary and secondary outcomes provide strong evidence of causality of effects but do not on their own tell us how such an effect arises. This is an important limitation because the identification of causal mechanisms is required to test competing theoretical explanations of the same causal effects, as well as to inform assessment of the potential generalisability of the evidence. Causal mediation analysis plays an essential role in potentially overcoming this limitation by helping to identify intermediate variables (or mediators) that lie in the causal pathways between the treatment and the outcome.

Our current logic model is linear and does not engage with how mechanisms vary with school context. This is a limitation which we will aim to address in the course of the evaluation, and we start this process in the next section by developing some hypotheses about how context will moderate intervention mechanisms and outcomes. Nonetheless, the linear logic model is useful in setting out the main mechanisms which we hope will occur in most schools which will enable intervention benefits to be realised. To examine empirically whether this is the case we will explore the validity of mediation hypotheses suggested by the linear logic model:

Hypothesis 1: LT schools will report reduced student-student, student-staff and academic-broader learning boundaries and increase student-centred framing at follow up 1 and 2.

Hypothesis 2: LT schools will report higher rates of student commitment to the schools' instructional and regulatory orders by follow up 1 and 2.

Hypothesis 3: LT schools will report higher rates of student life skills and warm, trusting and empathetic relationships and lower rates of student involvement in anti-school peer groups by follow up 1 and 2.

Hypothesis 4: Intervention beneficial effects on primary and secondary trial outcomes will be mediated by these reductions in boundaries, increases in student-centred framing, student commitment, skills and relationships, and decrease in involvement in anti-school peer groups.

Contextual factors that may affect implementation

We develop some hypothesis regarding likely moderators of implementation and effectiveness organised in CMO configurations. These will be refined in the light of emerging data on intervention implementation and receipt as discussed above. In line with MRC guidance, we did not attempt to consider every possible external factor with which the intervention might interact (Craig et al. 2014).

Hypothesis 1: The intervention will be more acceptable and be implemented with better fidelity when it is line with existing school institutional approaches and teacher practices which aim to erode the boundaries which the intervention is addressing (Bonell et al 2010a,b).

Hypothesis 2: The intervention will be implemented with better fidelity when the school has the capacity to implement it properly, in terms of: the action team being chaired or otherwise led by a person with real authority in the school; the action team involving other individuals which means it is taken seriously both by staff and students; the action team being formally linked in to the school decision making structures (e.g. SLT); teachers with varying levels of experience (early career teachers and experienced teachers) involved in implementation; and the school being a generally functional institution e.g. stable staffing, not in crisis with respect to targets and inspections (Bonell et al 2010a,b; 2015).

Hypothesis 3: The intervention will be implemented with better fidelity in schools that include students with varying degrees of educational engagement in its activities (e.g. action groups), including students who have a history of, or considered likely to be involved in bullying behaviours (Bonell et al 2015).

Hypothesis 4: We hypothesise that schools will implemented restorative approaches with less fidelity of function in schools with higher numbers of African Caribbean or minority ethnic students (as reported in cross sectional research in US schools) because staff stereotype these students as too challenging, unruly or aggressive to benefit from restorative practice (Payne and Welch 2013).

Contextual factors that may modify intervention effects

Hypothesis 5: The intervention will be more effective in schools with more students of low SES backgrounds since eroding boundaries is hypothesised as more important for these students (Markham and Aveyard 2003).

Regarding gender, Flay et al (2004) reported a range of benefits for boys but not girls. However, in the absence of process evaluation, the reasons for these differential effects are unclear. Therefore before developing hypotheses, we will examine emerging data from our process evaluation on whether and how the intervention may be implemented for and received by girls and boys.

Hypotheses of how the intervention might produce harms

Below we outline hypothesised harms and uncertainties of our intervention and describe the potential mechanisms underlying these, and the data sources to assess these.

Perverse effects of public sector targets/ measures: Intervention providers may respond creatively to structural conditions (intervention implementation) in ways that enable them to meet their (government) monitoring targets (van Thiel and Leeuw 2002). Some schools may focus too narrowly on developing school actions that they perceive will most positively impact on the attainment of a subgroup of students on the threshold of achieving 5 good GCSEs (as this remains the key metric which schools are judged on) or on the attainment of all student but using techniques such as teaching to the test. Such actions might amplify any existing trend towards increasing boundaries between attainment and broader development rather than eroding them. These actions might make health worse by: further displacing actions addressing broader health and well-being; by alienating students not in the 'threshold' group or increasing academic related anxiety among students (Bonell et al 2013).

Perverse effects of reducing boundaries: it is possible that the intervention does succeed in eroding boundaries, but that this has harmful consequences. It may be that Markham and Aveyard's (2003) theory of human functioning and school organisation, on which our

intervention is based, is wrong and in fact a school ethos characterised by staff authority and a firm priority on academic achievement may in fact be responsible for better attainment and reduced risk behaviours of students rather than via eroding boundaries. This view receives partial support from educational research (Mortimore et al. 1988; Sammons 2007; Sammons 2012; Sammons et al. 2011) that suggests that an orderly environment and priority for academic engagement are key for attainment (though the studies do not suggest that schools achieve these in ways that run counter to the logic of our intervention). According to this theorisation, reducing boundaries may produce increased student risk behaviours, truancy and worse attainment than control schools at follow-up.

Perverse effects of actions developed from Action Group meetings: it is possible that the actions developed and implemented for school organisational change resulting from the Action Group result in harmful consequences. Students (and potentially staff) may creatively suggest activities that run against the logic of the intervention and which potentially cause harm. This is avoided by having an external facilitator present at all Action Group meetings (as evidence in the pilot), but may still be plausible.

Perverse effects of focusing on health and well-being: Related to the point above, a primary focus on student health and well-being (as opposed to attainment) may lead to changes in school composition that exacerbate negative outcomes. There is some suggestion that some health and wellbeing interventions might sometimes cause schools to lose more academically able students and attract more challenging students. Interventions which either are so stressful as to lead to staff burnout or which equip staff with marketable news skills may both lead to greater staff turnover (Aber, personal communication). These mechanisms are plausible in our intervention and we hypothesise this may lead to increased aggregation of anti-school young people and thus worse behaviours; as well as decrease in staff well-being and retention.

Perverse effects of restorative practices: restorative practices may simply be less effective than traditional discipline when applied in a school context. There is no previous RCT in schools examining the effects of restorative practices, so we are uncertain. It is plausible that exclusions and punitive discipline are more effective in promoting the wellbeing of students who remain in the school. It is also plausible that offenders (or potential offenders) may view restorative approaches and informal communication processes as 'easy', reinforcing their belief that the aggression is acceptable, trivial or justified. On the other hand, our intervention may lead to a more pervasive system of discipline in schools. Restorative practices are considered more benevolent than traditional discipline, so may result in more teachers disciplining more students. In schools with the lowest rates of exclusion, we may see higher rates of bullying and/or violence as a result of using an *overly* restorative approach, as the few persistent offenders are not removed from school or are too slowly excluded, which may then affect other students. This might then lead to more students labelled as problematic leading to increased risk behaviours.

We believe that the likelihood of these hypotheses being confirmed is low but will examine these empirically. We will also use qualitative data examine the possibility of other unanticipated harms arising, where possible also then examining these quantitatively.

Normalisation

A key question for process evaluators is considering whether and how interventions become part of the policies and systems of the institutions in which they are delivered. This is important if complex interventions that are trialled are to become scaled up as part of routine practice. Understanding this process by which interventions become part of routine practice, often described in terms of ‘normalisation’ or ‘sustainability’, is another component of this process evaluation.

There are a variety of frameworks which help evaluators assess intervention normalisation such as the RE-AIM framework (Glasgow et al., 2001; Glasgow et al., 1999) and normalisation process theory (NPT; May and Finch 2009). The RE-AIM framework argues that in order to sustainably benefit a population, an intervention needs to be not merely effective but also reach the targeted population of beneficiaries, be adopted by those organisations that are to deliver it, be implemented with fidelity, and have all of these factors be maintained over time. While this usefully provides a set of critical factors to be achieved when scaling up an intervention, it does not aim to provide insights into the processes through which this occurs in order to ensure these factors are in place.

NPT is concerned with how implementers can enable the embedding of interventions within institutions and social contexts so that they become an integrated part of these (May and Finch 2009). The theory suggests that whether this happens depends on the following four “generative mechanisms” that those working on implementation engage in individually and collectively:

- coherence (how people make sense of a new practice);
- cognitive participation (the willingness of people to sign-up and commit to the new practice);
- collective action (their ability to take on the work required of the practice); and
- reflexive monitoring (activity undertaken to monitor and review the practice).

Summary of research questions

As previously outlined, our process evaluation, informed by MRC guidance (Moore et al, 2014; Craig et al 2008) and the wider implementation science literature (e.g. May and Finch 2009; Bumbarger and Perkins, 2009; Hawe, et al. 2009) investigates the following domains:

- implementation
- mechanisms of impact and context
- normalisation

Below we outline the research questions we investigate within each of these domains. Refer to tables 3-5 for a list of data sources which will be used to analyse each question.

Implementation

1. Are standardised intervention components implemented with fidelity of form?
2. Where there are deviations from fidelity of form, do these reflect intentional adaptation (and if so with what motivation), unintentional drift or simple omission, and do adaptations run with or against the logic of our theory of change?
3. Are locally tailored intervention components implemented with fidelity of function?
4. How does fidelity of form and function vary across schools and what contextual factors appear to affect this?
5. What is the reach and acceptability of each intervention component?
6. How do reach and acceptability vary across schools and what contextual factors appear to affect this?

Mechanisms of action

1. Does the intervention reduce student-student, student-staff and academic-broader learning boundaries and increase student-centred framing?
2. Does the intervention increase student commitment to schools' instructional and regulatory orders?
3. Does the intervention increase student life skills and warm, trusting and empathetic relationship, and reduce involvement in anti-school peer groups?
4. Are intervention beneficial effects on primary and secondary trial outcomes mediated by reductions in boundaries, increases in student-centred framing, student commitment, skills and relationships, and decreases in involvement in anti-school peer groups?
5. What refinements to our theory of change are suggested by qualitative data?
6. Is the intervention more acceptable and implemented with better fidelity when it is line with existing school institutional approaches and teacher practices which aim to erode school boundaries?
7. Is the intervention implemented with better fidelity when the school has the capacity to implement it properly?
8. Is the intervention implemented with better fidelity in schools that include students with varying degrees of educational engagement in its activities?
9. Do schools with higher numbers of African Caribbean students implement restorative approaches with less fidelity?
10. Is the intervention more effective in schools with more students of low SES backgrounds?
11. Are there gender differences in intervention effects?
12. Do intervention harms arise because: some schools focus on increasing attainment of some students rather than increasing the wellbeing of all students; eroding boundaries is associated with increased risk; the intervention leads to increased staff and student

mobility; restorative approaches are less effective than traditional discipline; the intervention promotes peer deviancy training?

13. Are other harms suggested by qualitative data?

Normalisation

1. Is the intervention more sustained in year 3 in those schools in which staff and students:
 - a. view the intervention as coherent;
 - b. commit to participation;
 - c. collectively take on the work arising from the intervention; and
 - d. review progress implementing it?

4. Data collection methods for process evaluation

Below we describe our methods of sampling and data collection. Where these are already outlined in sufficient detail in the main trial protocol (Bonell et al, 2014) we present a summary, reserving more space for discussing areas where we need to expand or deviate from what is currently in the main trial protocol. Areas where we have deviated from the main trial protocol are marked with an asterisk (*). Appendix 1, Table 2 provides further details on deviations, including rationales for changes made. Most of the data collection tools outlined below were informed by tools developed and piloted in the pilot trial (Fletcher et al, in press; Bonell et al, in press) and found to be feasible and appropriate. All individuals (including students) will give their informed written consent to participate. In the case of students, parents will also be sent information about the research and have the right to withdraw their children from the research.

The way in which specific measures and data will be used to assess our hypotheses will be considered in a later section on analysis.

Baseline and follow-up staff and student surveys

These survey all teaching staff and students in intervention and control schools at baseline (end of students' year 7), 24 months (end of student year 9) and 36 months (end of student year 10), sampling all staff and students who are part of the school at that point in time. As well as providing data for the evaluation, the baseline and 24 month surveys, as well as a 12 month survey in intervention schools, provide needs assessment data to schools and would remain a part of the intervention were it to be scaled up.

Facilitator diary forms

Facilitators are required to complete a diary form for each action group meeting they conduct in the schools in which they work. Diary forms report: general meeting information such as duration, date, number of attendees, chair and minute taker names; members' (staff and student) role, year group and gender; how and what data are used to inform setting up school actions (e.g. needs assessment reports, other school data); priorities set by the school and actions for whole-school change stemming from these; actions concerning the revision of school rules and school policies; identification of which parts of the LT curriculum are to be implemented and how this was decided; and comments on responsiveness of action group members. The diary form is informed by the same tool used in the pilot study.

Facilitators will be asked to report any alterations made. When alterations are made, they will be asked whether these related to: procedure (e.g. timing, location); participants (e.g. attendees); and/or content. Intervention deliverers will also select reasons for alterations which could include: logistical (e.g. related to capacity, resource, time); locally appropriate (e.g. made

sense based on culture, environment and/or participants); other (free text provided); and/or 'don't know'. Free text boxes are also provided for deliverers to describe the alteration itself.

Action Group meeting minutes

Facilitators are required to return meeting minutes from each action group meeting held. The minutes from a random sample of 10 schools will be used to triangulate the validity of facilitator diary forms.

Trainer diary forms and attendance sheet

LT restorative practice trainers complete a diary form for each session of all-staff awareness training they deliver. Trainers rate the extent to which they covered topics/materials as they had intended in the training session; and the materials and activities (e.g. power point slides, small group or paired activities) used. Free text is provided to comment on participant responsiveness to the training. Trainers will be asked to report any alterations made. When alterations are made, they will be asked whether these related to: procedure (e.g. timing, location); participants (e.g. attendees); and/or content. Intervention deliverers will also select reasons for alterations which could include: logistical (e.g. related to capacity, resource, time); locally appropriate (e.g. made sense based on culture, environment and/or participants); other (free text provided); and/or 'don't know'. Free text boxes are also provided for deliverers to describe the alteration itself. The diary form is informed by the same tool used in the pilot study. An attendance sheets for each all-staff awareness training session is completed by an attending staff member at the school. Attendance sheets are also circulated to staff attending the in-depth three day restorative practice course.

Participant satisfaction survey for in-depth training

A satisfaction survey is to be completed by all staff members attending the in-depth three day training on restorative practice (8-10 staff members per school). Questions assess whether participants felt the training was useful, if they feel confident about putting into practice the skills learnt; if they would recommend the training; and overall how they would rate the training provided. The survey is informed by the same tool used in the pilot study.

Curriculum implementation log*

The implementation log is completed by the teacher acting as LT curriculum co-ordinator in each intervention school (or in some cases all teaching staff delivering the curriculum in the school will complete a form individually, as each school prefers). Curriculum coordinators will be sent a curriculum log by a member of the evaluation team via email during each term in years 1-2 and in Term 3 in year 3 and asked to complete this and return it to the evaluation team via email. It reports what units and lessons were delivered, when, in which subjects, for how many hours and what LT materials (e.g. power points, lesson plans), if any, were used to deliver the content. Curriculum coordinators will be asked to report if any amendments were made to the

lesson plans and materials. Where amendments were made, they will be asked whether these were major or minor amendments. The log is informed by the tool used in the pilot study.

Staff telephone interviews*

Telephone interviews will be conducted with one senior leadership team (SLT) member and two teaching staff across all intervention and control schools (n=40) at the beginning of year 1. The key contact liaising with the trial team at each of the 40 schools will be contacted via email and/or telephone and asked to identify three staff members to participate. Staff members will thus be chosen based on convenience sample either identified by the key contact or volunteering in response to a request from the contact. A member of the research team will then contact selected staff via email to schedule a telephone interview and obtain informed written consent to participate. No interviews will be conducted in Year 2. In year 3 interviews will be conducted with one SLT member across all intervention and control schools. Control schools will be interviewed in Term 1 and intervention schools will be interviewed in the Term 3.

The first part of all interviews (year 1 and 3) will be structured, collecting quantitative data on staff perceptions of their schools' organisational approaches and practices. The survey asks a series of questions which respondent's rate on a likert scale of 1 to 4. The questions are aimed at identifying whether schools use approaches which erode 'boundaries': (a) between staff so that authority is distributed across staff versus concentrating authority among senior staff; (b) between students so that positive relationships are encouraged and students treated equitably versus students being divided into groups which may not be treated equitably; (c) between different areas of school life, so teachers focus on students' overall wellbeing and development not merely academic progress; and (d) between the school and its community so that the school is integrated as opposed to separated from its local community. Responses will be recorded in writing by the interviewee.

The second part will be semi-structured, collecting qualitative data on staff views about being involved in the trial; the contexts of schools including, current management priorities, policies and practices towards bullying and aggressive behaviours; how PSHE and social and emotional curriculum are delivered; student participation in decision-making; what other relevant services and practices; staff training; and existing programmes related to health in the school. These will be audio-recorded.

Telephone interviews with curriculum deliverers (LT curriculum teaching staff)

Telephone interviews will be held in years 1-3 in all intervention schools with one teaching staff member who is responsible for delivery of the social and emotional curriculum. The curriculum coordinator at each school will be contacted via email and/or telephone and asked to identify a member of staff delivering the curriculum to participate. Staff members will thus be chosen based on convenience sample either identified by the curriculum coordinator or volunteering in response to a request from the coordinator. The interviews will gather views on the fidelity,

reach and acceptability of the curriculum, which materials are used, delivery methods, student responsiveness and contextual barriers or facilitators of delivery. Interviews will also explore alterations/deviations that may have been made and the reasons for these. These will be audio-recorded.

Interviews with intervention deliverers

One to one interviews with LT facilitators (n=6) and trainers (n=2) will be conducted in years 1 and 2. Interviews aim to gather views on school responsiveness to intervention activities; adaptations and deviations made to intended delivery and reasons for this; and barriers and facilitators to delivery.

Interviews with Action Group members

We will interview two action group members per school annually. A member of the evaluation team will contact intervention school Learning Together lead co-ordinators and ask them to identify 2 action group members (staff and/or student) to interview. Identified participants will be contacted via email and/or phone to schedule a telephone interview (or if possible for researcher, in person). These will be semi-structured to elicit views on the acceptability of facilitators, what they like/dislike or barriers/facilitators of action group meetings, how they might be improved, extent to which actions arising from meetings are cascaded in the school and general usefulness of meetings in reducing boundaries between staff-students. Interviews will also explore alterations/deviations that may have been made and the reasons for these. These will be audio-recorded.

Action group members' survey

A survey will be circulated to all action group members (year 1-3) in intervention schools to gather quantitative data on the acceptability, functioning and composition of the action groups. It asks questions, for example, related to the diversity of staff and students on the action group, the usefulness of the needs-assessment report; and an empowerment scale. This was informed by the tool used in the pilot study.

Researcher observations*

We will conduct structured observations of: a number of training sessions so that all schools (n=20) have been observed at least once in year 1; and one action group meeting at 10 randomly selected intervention schools in years 1-3. Observation records are informed by the same tools used in the pilot study. Researchers will also keep unstructured diary notes of observations and discussions when visiting intervention schools.

Restorative monitoring survey

The purpose of the survey is to examine the extent to which restorative practices are implemented within the school. The survey is to be completed once per year in term 3 by all staff who attended in-depth training at each of the intervention schools (8-10 staff per school). Staff who attended in-depth training will be sent an invitation to complete the survey. We will collect data on whether those trained in depth in restorative approaches have used restorative practice to inform their practice in the classroom and elsewhere in school including the use of affective language, circles, mediation, restorative conferencing, family group conferencing, community conferencing.

Routinely collected school data (intervention and control)

The following routine individual and school level data will be collected in all intervention, as well as control schools:

- school attendance will be measured via routine school data on each student expressed as number of half days absent; for which we will seek students' informed consent to access;
- staff attendance will be measured via routine school data on each staff-member expressed as number of half days absent, for which we will seek staff-members' informed consent to access;
- school-level rates of temporary and permanent exclusions;
- total number of staff (by job types) and students;
- mean pay of teachers;
- number of days lost to staff sickness; and
- student educational attainment: this will be assessed by an independent team based at the University of Manchester drawing on routine data to which we will have access.

These data on schools will also be used to get a more complete picture of school level characteristics (e.g. staff turnover; staff-student ratio; student attainment figures).

Data collected in case study schools*

Six intervention schools will be selected as 'case study schools' to gather in-depth qualitative data on intervention processes and school context. We intend to conduct focus groups with staff and students in case study schools, but will also collect other data via interviews with students for example and school observations. To capture the range of different types of schools in our overall sample, case study schools were purposively sampled for diversity in terms of: percentage of free school meals (schools identified as either above or below national average in 2012 for secondary schools, 16.3 %), type of school; the external intervention facilitator

assigned to the school; and on the extent to which the school was responsive (highly responsive, somewhat responsive; poorly responsive) to intervention activities, as rated by the intervention facilitators, three months into the intervention start date.

Table 1: Characteristics of case study schools

Case study schools* (N=6)	Type of school	FSM	Intervention facilitator	School responsiveness
SCHOOL 1	Voluntary Aided School	Above average (56.7%)	JH	Somewhat
SCHOOL 3	Academy - Converter Mainstream	Above average (56.3%)	RM	Highly
SCHOOL 19	Community School	Above average (56.9%)	TM	Somewhat
SCHOOL 22	Academy - Converter Mainstream	Below average (3%)	JH	Highly
SCHOOL 25	Foundation School	Above average (74.5%)	SS	Highly
SCHOOL 30	Academy - Converter Mainstream	Below average (5.8%)	JG	Somewhat

*Identifiers are used so that school names remain anonymous

Staff focus groups in case study schools

Each year 1-3, we will undertake one staff focus group in each case study schools, each involving 4-6 members of staff. Staff will be purposively selected by the LT lead contact in each school to include diversity according to participation in the LT intervention and roles (including senior leaders, pastoral staff and classroom teachers). Focus group discussions aim to: gather views and perspectives around intervention pathways/mechanisms (i.e. to get a sense of social change in the school that has been enabled or constrained by the LT intervention); and to gain an understanding of how restorative practices are applied and barriers to facilitators to its use.

Student focus groups in case study schools

Each year, we will conduct two focus groups with students in each case study school, comprising 4-6 students each. Students will be purposively selected by the LT lead contact in each school based on the following guidance: one of the focus groups should include 4-6 students who have been directly involved in LT activities (e.g. action group members). The other focus group should include 4-6 students who are not directly involved in LT. Both groups should involve students from the study cohort. Each group should reflect the diversity of the school in term of boys and girls, different ethnic groups and students of varying degrees of educational engagement. Focus group discussions with students will aim to gather views on intervention pathways/mechanisms. Focus group discussions will also aim to get a deeper understanding of the nature of bullying in schools and views on the acceptability and appropriateness of restorative approaches.

One to one interviews with students involved in restorative practices

One to one interviews will take place with two students at each case study school that have been involved in restorative practices (e.g. conferencing). The lead contact or pastoral coordinator in each school will identify students for participation but will be asked to identify a male and female student where possible. The purpose is to understand processes of restorative practice and assess acceptability of the approach.

Appendix 1. Overview of data sources and how they pertain to each domain of evaluation

Overview of data sources for the process evaluation

Table 2: Overview of all data sources for process evaluation

Evaluation data	Completed by	Timing	No. of participants	No. of schools (intervention/control)	Evaluation domain addressed	Deviation from LT protocol and/or notes
Baseline and follow-up surveys	Staff and students	Baseline (year 1); 24 month; 36 month	Approx. N=180 per school	N=40 (Int: N=20/Cont: N=20)	Student outcomes Mechanisms Context	None
Routinely collected school data (intervention and control)	Routinely collected	End of year 1, 2, 3	-	N=40 (Int: N=20/Cont: N=20)	Context	None
Staff telephone interviews	Staff (Senior leaders and teaching staff)	Year 1, 3	Year 1 N=1 SLT; N=2 staff (Total: N=3) Year 2 None Year 3	N=40 (Int: N=20/Cont: N=20)	Implementation fidelity acceptability Context	Deviation: We originally planned to interview 1 SLT and 2 teaching staff across all 40 schools in all three years but no longer intend to do these interviews in year 2 and only have interviews with 1 SLT in year 3.

			N=1 SLT			Rationale for deviation: Interviews in year 2 were considered unnecessary since we are already collecting other data (e.g. via interviews with action team members, curriculum surveys, focus groups) on how the intervention is progressing in intervention schools. Some control schools have also reported overburden following Year 1 interviews, so we have reduced this in Year 3. Resources are being re-directed to in-depth case studies of intervention schools (and away from superficial data collection across all schools).
Facilitator diary forms	Facilitators	Approx. half-termly	N=6 meetings per year	N=20 (I)	Implementation fidelity	None
Action Group meeting minutes	Staff member in AGM	Approx. half-termly	N=6 meetings per year	N=20 (I)	Implementation fidelity	None.
Trainer diary forms and attendance sheet	Trainers	Year 1	N=2	N=20 (I)	Implementation fidelity	None
Participant satisfaction	Staff attending in-depth training	Year 1	Approx N=160-200 in total	N=20 (I)	Implementation fidelity acceptability	None

survey for in-depth training						
Curriculum implementation log	School curriculum co-ordinator	Years 1-3	N=1 per school	N=20 (I)	Implementation fidelity	Deviation: We originally intended to observe N=1 curriculum session in each school but are now logging delivery and interviewing curriculum leads instead.
Telephone interviews with curriculum deliverers (LT curriculum teaching staff)	Staff	Years 1-3	N=1	N=20 (I)	Implementation fidelity Acceptability	<p>Rationale: the lead intervention facilitator advised us that observations would create an excessive administrative burden for schools.</p> <p>Deviation 2: We originally planned to collect data termly on curriculum delivery but are in year 3 we are collecting it only at the end of the year</p> <p>Rationale 2: to increase response rate and decrease participant fatigue.</p>
Restorative monitoring survey	Staff attending in-depth training	Years 1-3	Approx N=160-200	N=20 (I)	Implementation fidelity Acceptability	Deviation: We originally conducted the survey termly but because of poor response rates changed to an

						annual (summer term) survey for year 3.
Interviews with Action Group members	Staff and students	Years 1-3	N=2	N=20 (I)	Implementation fidelity Acceptability Context Mechanisms	None
Action group members' survey	Staff/Students on action group	Year 1-3	Approx N=10	N=20 (I)	Implementation fidelity Acceptability	None
Researcher observations	Researcher	Years 1-3	-	N=10 (I)	Implementation fidelity	Deviation: Observations to be done in n=10 schools at random rather than n=20. Rationale: resources being redirected to in-depth case studies
Staff focus group in case study schools	Staff	Years 1-3	N=1 FGD (comprising N=4-6 staff each)	N=6 (I)	Acceptability Context Mechanisms	Deviation: Protocol originally specified case studies in n=4 control schools; n=4 intervention schools.
Student focus groups in case study schools	Students	Years 1-3	N=2 FGD (comprising N=4-6)	N=6 (I)	Acceptability Context Mechanisms	Rationale: The main purpose of the case studies is to capture data on intervention mechanisms. Case studies of control schools will not be informative. Control schools have

			students each)			also complained about overburden. We have re-directed resources so that we are doing more work in intervention schools (n=6 schools as case study sites; conducting 1 focus group with staff; and 2 focus groups with students in each school. Conducting 8 case studies as originally intended would also provide too much data to do in-depth analysis.
One to one interviews with students involved in restorative practices in case study schools	Students	Years 1-3	N=12	N=6 (I)	Acceptability Context Mechanisms	
Interviews with intervention deliverers	Facilitators and trainers	Facilitators: years 1-2; Trainers: year 1	N=6 facilitators; N=2 trainers	-	Implementation fidelity	None

Overview of data sources pertaining to each component of evaluation

Implementation

Table 3: Data sources for RQ on implementation

Data collected	Whose perspective do the data address?	Implementation			
		<i>Fidelity of form/function</i>	<i>Adaptation, drift or omission</i>	<i>Reach and acceptability</i>	<i>Implementation variation by context</i>
12 month survey	Students			✓	✓
24 month survey	Students			✓	✓
36 month survey	Students			✓	✓
Staff telephone interviews	Staff			✓	
Facilitator diaries	AGM members	✓	✓		
AGM attendance sheet	AGM members	✓		✓	
Interviews with AGM	AGM members	✓	✓	✓	✓
AGM members survey	AGM members	✓	✓	✓	✓
Trainers' diaries	Trainers	✓	✓		
Training attendance sheet	School staff	✓		✓	

In-depth training satisfaction survey	School staff			✓	✓
Curriculum implementation log	School staff	✓	✓	✓	
Restorative practice monitoring tool	School staff	✓		✓	
Facilitator interviews	Facilitators	✓	✓	✓	✓
Case study school in-depth data	Students and staff			✓	✓
Interviews with intervention deliverers	Facilitators, trainers and curriculum teachers	✓	✓	✓	✓

Mechanisms of action

Table 4: Data sources for RQ on mechanisms of action

Data collected	Whose perspective do the data address?	Mechanisms		
		<i>Pre-hypothesised intervention mediators</i>	<i>Contextual factors that may affect implementation</i>	<i>Contextual factors that modify intervention effects</i>
12 month survey	Students	✓		✓

24 month survey	Students	✓		✓
36 month survey	Students	✓		✓
Staff telephone interviews	School staff		✓	
Facilitator diaries	AGM members		✓	✓
AGM attendance sheet	AGM members			
Interviews with AGM	AGM members	✓	✓	✓
Trainers' diaries	Trainers		✓	✓
Training attendance sheet	School staff			
In-depth training satisfaction survey	School staff			
Curriculum implementation log	School staff			
Restorative practice monitoring tool	School staff			
Facilitator interviews	Facilitators	✓	✓	✓
Case study school in-depth data	Students and staff	✓	✓	✓
Interviews with intervention deliverers	Facilitators, trainers and curriculum teachers	✓	✓	✓

Normalisation

Table 5: Data sources for RQ on normalisation

Data collected	Whose perspective do the data address?	Normalisation			
		<i>Coherence</i>	<i>Commitment</i>	<i>Collective action</i>	<i>Review progress</i>
12 month survey	Students				
24 month survey	Students				
36 month survey	Students				
Staff telephone interviews		✓	✓	✓	✓
Facilitator diaries	AGM members				
AGM attendance sheet	AGM members				
Interviews with AGM	AGM members	✓	✓	✓	✓
Trainers' diaries	Trainers				
Training attendance sheet	School staff				
In-depth training satisfaction survey	School staff				
Curriculum implementation log	School staff				
Restorative practice monitoring tool	School staff	✓	✓	✓	✓

Facilitator interviews	Facilitators	✓	✓	✓	✓
Case study school in-depth data	Students and staff	✓	✓	✓	✓
Interviews with intervention deliverers	Facilitators, trainers and curriculum teachers	✓	✓	✓	✓

Appendix 2. Data analysis methods: Implementation

Standardised components: fidelity of form

We will use the following data sources to examine the extent to which the following standardised components of LT were implemented with fidelity of form:

Table 6: Data sources to assess: fidelity of form

Intervention component	Data source to assess fidelity of form
Needs assessment surveys	Research team documentation
Training	Trainer diaries, observations
Action Group meetings	Facilitator diaries, observations, minutes, interviews with facilitator and Action Group members
Review of needs data	Facilitator diaries, observations, minutes, interviews with facilitator and Action Group members
Development of an action plan based on needs data	Facilitator diaries, observations, copy of plan, interviews with facilitator and Action Group members
Review of policies relevant to aggression and bullying	Facilitator diaries, observations, minutes, copy of revised documents, interviews with facilitator and Action Group members
Rewriting of school rules	Facilitator diaries, observations, minutes, copy of revised documents, interviews with facilitator, Action Group members, other students and staff
Social and emotional skills curriculum	Monitoring data collated by school curriculum coordinator, interviews with other students and staff

Data from facilitator and trainer diaries and curriculum implementation log will be entered into an Excel database that contains fields for standardised intervention features which we examine (i.e. action groups, social and emotional curriculum, restorative training; revision of school policies and rules) by school. This will provide a comparative overview of implementation fidelity.

Table 7: Needs assessment surveys

Needs assessment surveys	Number (and %) of schools		
	Year 1	Year 2	Year 3
Conducted			
Response rate <70%, 70-79%, 80-89%, 90+%			

Table 8: Action group implementation across intervention schools

<i>Action group implementation</i>	Number (and %) of schools			
	Year 1	Year 2	Year 3	Overall
Size of action group <8, 8-10, >10				
Broad cross section of students represented – yes, somewhat, no				
At least six action groups conducted y/n				
Attended by senior leader with authority – y/n				
Action plan drafted				
NAR informed priorities/actions – y/n				
School policies have been revised y/n		-	-	
School rules have been revised y/n		-	-	

Table 9: Social and emotional curriculum implementation

Social and emotional curriculum implementation	Number (and %) of schools			
	Year 1	Year 2	Year 3	Overall
NAR informed curriculum delivery – y/n				
Usage of LT materials -always, sometimes, never				
Curriculum delivered to all class groups in years 8-10 in years 1-3				
Number of hours delivered/year – average				
Number of units delivered – total number				

Number of lessons delivered – total number				
--	--	--	--	--

Table 10: All-staff training implementation

Training implementation	Number (and %) of schools
<i>Topics covered as self reported in trainer diary- as intended, less than, not at all):</i>	Year 1
What is restorative justice/practice	
The importance of language	
What we do to challenge bad behaviour/ nature of challenge	
The importance of emotions	
The importance of listening	
Maintaining the relationship after difficult conversation	
<i>Training materials and activities used</i>	
Power point slides	
DVD	
Paired activity	
Small group activity	
Number of staff attended (as proportion of total number of school staff)	

Table 11: In-depth training implementation

Training implementation	Number (and %) of schools
<i>Topics covered as reported by participants attending training - as intended, less than, not at all):</i>	Year 1
TBA	
TBA	
Etc.	

Locally tailored components: fidelity of function

We will use the following data sources to examine the extent to which the following locally tailored intervention components of LT were implemented with fidelity of function:

Table 12: Data sources to assess: fidelity of function

Intervention component	Data source to assess fidelity of form
Other local actions included in Action Plan	Facilitator diaries, observations, minutes, interviews with facilitator and Action Group members
Locally decided means of implementing restorative approaches	Monitoring data on use of restorative approaches, interviews with facilitator and Action Group members, staff trained to deliver restorative approaches

Table 13: Restorative practices implemented

Locally decided implementation of restorative approaches	Number (and %) of schools			
	Intervention (n=20 schools)			
	Year 1	Year 2	Year 3	Overall implementation
Use of restorative approaches to inform classroom management (many times; a few of times; hardly ever; none)				
Circle time/equivalent (many times; a few of times; hardly ever; none)				
Facilitated meetings internal to the school *(many times; a few of times; hardly ever; none)				
Conferences with participation of external agencies * (many times; a few of times; hardly ever; none)				

*Respond only if there was conflict in your school to address

Table 14: Other locally decided actions

Implementation of locally decided actions	Number (and %) of schools
	Intervention (n=20 schools)

	Year 1	Year 2	Year 3	Overall implementation
Locally decided actions included in action plan				
Locally decided actions informed by NAR data				
All locally decided actions implemented as planned				

As above, we will then draw on qualitative data to explore whether locally decided implementation of restorative approaches were intentionally tailored to be consistent with the logic of our theory of change or whether implementation was not carefully thought through or deviated from the logic of our theory of change.

Reach and acceptability

We will use the following data sources to examine the reach and acceptability of the following aspects of LT

Table 15: Data sources to assess: reach and acceptability

Intervention component	Reach	Acceptability	
	<i>Quantitative data</i>	<i>Quantitative data</i>	<i>Qualitative data</i>
Needs survey	Response rates	Question on 2 nd needs assessment survey – TBC	Focus groups with students
Action Groups	Minutes, facilitator diaries	Survey of Action Group members	Interviews with Action Group members
Rewriting school rules	Action Group minutes	Follow up student survey	Focus groups with staff and students
Social and emotional skills curriculum	Monitoring data collated by school	-	Focus groups with

	curriculum coordinator		staff and students
Training in restorative approaches	Attendance sheets	Satisfaction survey for in-depth training	Focus groups with staff
Implementation of restorative approaches	Monitoring data on use of restorative approaches	-	Focus groups with staff and students
Other locally decided actions	Action Group minutes	-	Focus groups with staff and students

We will report quantitative data on the reach (actual c.f. intended) and acceptability of each intervention component in each school. We will analyse qualitative data exploring acceptability as well as factors which promoted or detracted from intervention reach and acceptability.

Table 16: Students and staff involvement

Data	Number (and %) of schools			
	Intervention (n=20 schools)			
	Year 1	Year 2	Year 3	Difference
Students report awareness of intervention to reduce bullying and aggression in their school conferences – yes/no/unsure				
Students report involvement in rewriting school rules – yes/no/unsure				
Students report involvement in circle time or conferences – yes/no/unsure				
Students report receiving social and emotional skills curriculum – yes/no/unsure				
Staff report involvement in rewriting school rules – yes/no/unsure				
Staff report use of restorative approaches to resolve conflict – yes/no/unsure				

Appendix 3. Data analysis methods: Mechanisms of action

Mediation analyses

We will conduct mediation analyses (Baron and Kenny 1986). These will assess whether the intervention reduces student-student, student-staff and academic-broader learning boundaries, increases student-centred framing, student commitment to schools’ instructional and regulatory orders, student life skills and affiliation, and decreases involvement in anti-school peer groups. We will then assess whether these appear to mediate intervention beneficial effects on primary and secondary trial outcomes.

Table 17: School-level mediation analysis

Learning Together theory of change constructs	Originator scales	Items
Aggregate student perception of staff-student boundaries	Beyond Blue - Student Teacher relationships 10 item subscale	My teachers are fair in dealing with students There’s at least one teacher or other adult in this school I can talk to if I have a problem I feel I can go to my teacher with the things that are on my mind In this school, teachers believe all students can learn In this school, students’ ideas are listened to and valued In this school, teachers and students really trust one another In this school, teachers treat students with respect This school really cares about students as individuals Most of my teachers really listen to what I have to say Thinking of my teachers this term, I really like:
Aggregate student perception of student-centred framing of school	Beyond Blue – Participation 6 item sub scale	There are lots of chances for students at my school to get involved in sports, clubs and other activities outside class Teachers notice when students are doing a good job and let them know about it At my school, students have a lot of chances to help decide and plan

			<p>things like school activities, events and policies</p> <p>Student activities at this school offer something for everyone</p> <p>Students have a say in decisions affecting them at this school</p> <p>Students at this school are encouraged to take part in activities, programs and special events</p>
Aggregate staff perception of staff-student boundaries	Teacher authority or teacher-student collaboration	School policies and practices survey	<p>In my school students participate in decision making</p> <p>Teachers in this school always show respect towards students</p> <p>Students' views are listened to and taken seriously by staff in this school</p> <p>Teaching strategies at this school enable students to build their own knowledge</p> <p>There are opportunities for students to take responsibilities for their own learning in school ²</p> <p>In this school the senior leadership team makes decisions without consulting students</p>
	Teacher support for students across school or restriction to classroom	School policies and practices survey	<p>Teachers at this school are often involved in extracurricular activities</p> <p>In my school teachers mix with students at break times</p> <p>In my school teachers mix with students at lunch time</p> <p>In my school, a lot of student pastoral care is delegated to non teachers</p> <p>In my school, teachers avoid intervening in students disputes outside the classroom</p>

<p>Aggregate staff perception of academic/broader development boundaries</p>	<p>School policies and practices survey</p>	<p>The school has a system for rewarding students who achieve in non academic areas e.g. sport, arts</p> <p>Our school provides a broad range of extracurricular activities for students (e.g. plays, athletics, music, dance)</p> <p>The school development/improvement plan has targets related to student health and wellbeing</p> <p>School INSET/training days often focus on student health</p> <p>The school has a comprehensive written policy to address student smoking, drugs or alcohol use</p> <p>The school teaches a social and emotional learning curriculum</p> <p>My school offers a range of non traditional subjects for students in years 10 and 11</p>
<p>Aggregate staff perception of student-student boundaries</p>	<p>Dividing up or bringing together students (learning)</p>	<p>School policies and practices survey</p> <p>Teachers at this school are more interested in the students with potential to do well in tests and examinations</p> <p>The school has a system for rewarding students who work hard and/or make good progress even if they do not reach high standards</p> <p>Students of similar academic ability are grouped together for teaching in most subject areas</p> <p>This school targets resources on the students on the borderline of achieving 5 good GCSEs</p>

	Dividing up or bringing together students (discipline and pastoral)	School policies and practices survey	<p>My school mixes together students who are of different ages e.g. through tutor groups or extracurricular activities</p> <p>Certain students in my school are repeatedly isolated from other students in response to misbehaviour</p> <p>My school has a strong system of peer mentoring or peer buddying</p> <p>My school runs conflict resolution programmes for students</p>
--	---	--------------------------------------	---

Table 18: Individual-level mediation analysis

Learning Together theory of change constructs	Originator scales	Items on questionnaire
Student commitment to school regulatory order	Beyond Blue - Sense of belonging 8 item subscale	<p>I feel very different from most students here</p> <p>I can really be myself at this school</p> <p>Other students in this school take my opinions seriously</p> <p>I am encouraged to express my own views in my class(es)</p> <p>Most of the students in my class(es) enjoy being together</p> <p>Most of the students in my class(es) are kind and helpful</p> <p>Most other students accept me as I am</p> <p>I feel I belong at this school</p>
Student commitment to school instructional order	Beyond Blue - student commitment to academic values 4 item subscale	<p>I try hard in school</p> <p>Doing well in school is important to me</p> <p>Continuing or completing my education is important to me</p> <p>I feel like I am successful at this school</p>

Student capacity for affiliation		SDQ item 4	I usually share with others (food, games, pens etc.)
		SDQ item 9	I am helpful if someone is hurt, upset or feeling ill.
		SDQ item 17	I am kind to younger children.
		SDQ item 20	I often volunteer to help others (parents, teachers, children)
		SDQ item 12	I fight a lot. I can make other people do what I want.
Student capacity for practical reasoning	Empathy with others	SDQ item 1	I try to be nice to other people. I care about their feelings
	Ability to manage own emotions	SDQ item 21	I think before I do things
		WEMWBS_short item 4	I've been dealing with problems well
		WEMWBS_short item 5	I've been thinking clearly
	WEMWBS_short item 7	I've been able to make up my own mind about things	
	Ability to manage/not manage conflict	SDQ item 5	I get very angry and often lose my temper
Involvement with anti-school peer groups		SRD questions (adapted)	During the last 3 months at school, have you/any of your friends skipped or skived off at this school? Have you/any of your friends ever been temporarily or permanently excluded from this school?
		YDP	Please think about your best friends who are the same age as you. How many of them have been told off, stopped or picked

		up by the police in the last 12 months?
--	--	---

Before undertaking mediator analyses we will assess the reliability of the subscales used. Baseline analyses suggest that the staff reported scales are not reliable so we will explore alternative data sources with a view to including these on the staff 24 month follow up survey.

In addition, to allow identification of unintended pathways, unanticipated pathways and in-depth exploration of pathways which are too complex and extended to be captured quantitatively we will draw on interviews with school staff, focus group discussions with students and staff, interviews with action group members, interviews with facilitators and researcher observations from case study schools. Note that in some cases, where suitable, qualitative data may be coded to be assessed via quantitative analysis.

The qualitative data will be analysed using Framework Analysis. Framework analysis allows a large amount of diverse data to be analysed systematically and is more transparent than most other qualitative data analysis methods, and supports comparative case analysis. All qualitative transcripts will be uploaded to NVivo10 and synthesised in matrices. Data will be organised both by case (individuals clustered by school) and by category. Categories will be developed a priori by the process evaluation team and informed by hypothesis developed in section 1: theory of change (see table below for draft categories). The categories may also be refined and expanded with evidence emerging from the evaluation. Completed matrices thus synthesise our varied data within broad categories, in preparation for more interpretive analysis. Interpretive analysis will be guided by principles of ‘grounded theory’ and we will make constant comparisons and examine deviant cases to refine our analysis.

Examining CMO configuration hypotheses

We will use the following data sources from case study schools to examine the following CMO Configurations concerned with school and individual characteristics that moderate implementation

Table 19: Data sources to assess: moderation (context)

Context (moderator)	Enables or constraints mechanisms:	Outcome	Data collection
---------------------	------------------------------------	---------	-----------------

School already aiming to erode staff-student boundaries	Implementation of actions to change school boundaries	<p>Students engage in learning with high aspirations; more students connect to school and avoid risk; more students develop life skills; more students form trusting relations.</p> <p>This leads to the following outcomes:</p> <ol style="list-style-type: none"> 1) Reduced bullying and aggression 2) Improved quality of life and emotional and mental health 3) Reduced substance use and sexual risk 4) Reduced truancy and school exclusions 	School policy and practice survey. Fidelity and staff acceptability data.
School already aiming to erode staff-staff boundaries	Implementation of actions to change school boundaries		School policy and practice survey. Fidelity and staff acceptability data.
School already aiming to erode boundaries between academic and broader development	Implementation of actions to change school boundaries		School policy and practice survey. Fidelity and staff acceptability data.
School organisational capacity; staff turnover; stability	Implementation of any intervention activities.		Routine monitoring data; School policy and practice survey. Fidelity data.
Inclusion of students with varying degrees of educational engagement in LT activities including bullies.	Implementation of actions to change school boundaries. Increased commitment of disengaged students.		Facilitator diary forms; interviews with AGM. Fidelity data.

Racial composition of students	Implementation of restorative approaches.		Student survey. Fidelity of restorative approaches data.
Socio-economic status of students	Increased commitment of critical mass of disengaged students.		Routinely collected data; student survey

Context (moderator): The conditions in which the intervention is introduced that is relevant to the operation of the mechanism

Mechanism: What is it about the interventions or interventions that bring about effects “active ingredients”

Outcome: Intended outcomes resulting from the activation of different mechanisms (in different contexts).

Baseline analyses suggest that the staff reported scales are not reliable so we will explore using qualitative data from staff and facilitator interviews to categorise schools according to whether they are already aiming to erode boundaries at baseline.

With only 20 intervention schools we are unlikely to be able to test statistically whether implementation and effectiveness is significantly better in the types of school contexts set out in the above CMO configurations, so that these analyses will be qualitative oriented towards hypothesis building rather than testing.

It is plausible that schools’ capacity to implement standard components of the LT intervention are also affected by differences their in attitudes and acceptability towards the intervention. So you might expect for example, that where there is poor action group implementation, schools are also more likely to disagree with: “the time required by LT is well worth it in improved student behaviour”. We plan to assess whether standard components of the intervention were satisfactorily developed and examine this against the list of measures of attitudes/acceptability towards Learning Together.

Assessing potential harms and underlying mechanisms

We will examine whether the intervention causes harm both in terms of paradoxical effects on primary and secondary outcomes as well as in terms of any harmful externalities suggested by interim qualitative analyses from case study schools.

We will use the following sources of data to examine pre-hypothesised harms

Table 20: Data sources to assess potential harms and underlying mechanisms

Hypothesised harm	Mechanisms	Outcomes*	Data on mechanisms and outcomes
<i>Perverse effects of public sector targets/measures</i>	Increased boundaries between attainment and broader development to address attainment metrics (rather than eroding them).	Adverse emotional and mental health outcomes, increased risk behaviours.	Facilitator diary forms (reporting priorities/actions from AGM). School policies and practices survey. Follow up survey measures.
<i>Perverse effects of reducing boundaries</i>	Eroding boundaries and increasing student-centred learning via intervention practices	Increase in student risk behaviours, truancy and worse attainment results	School policies and practices survey. Routine data on attainment. Interviews with AGM Staff focus groups Follow up survey measures.
<i>Perverse effects of focusing on</i>	Secondary focus on attainment (primary is student health) leads to loss of pro-	Increase student risk behaviours	Routinely collected school

<i>health and well-being</i>	school/high achieving students and aggregation of anti-school students; increase stress for teachers to implement health related practices; newly acquired skills from intervention training	Increase staff turnover Decrease staff well-being	data on staff and student mobility. Interviews with AGM Staff focus groups Follow up survey measures.
<i>Perverse effects of restorative practices</i>	Restorative process considered 'easy', reinforcing bad behaviours; More teachers delivering restorative practice to more students (and at higher doses).	Increased risk behaviours (bullying and aggressive). More students labelled as problematic leading to (self-fulfilling prophecy) increased risk behaviours.	Monitoring data on use of restorative approaches. Interview with AGM; focus groups with students. Follow up survey measures.
<i>Perverse effects of anti-school peer influences/ aggregating students</i>	Reinforcing or cascading risk behaviours (via group aggregation)	Increase in risk behaviours including bullying and aggression	Facilitator diary: AGM membership data; Interviews with AGM. Monitoring data on use of restorative approaches. Follow up survey measures.

Where we find evidence of harms, we will examine whether these might have arisen through the mechanisms hypothesised above by assessing whether these harms occur more frequently in schools characterised by the above processes, as well as via mediator analyses for example examining student-level involvement in intervention activities.

Appendix 4. Data analysis methods: Normalisation

We will use qualitative data primarily from our six case study schools in the first two years of the trial to develop theories and hypotheses about whether and how the above mechanisms are enabling the normalisation of LT in schools. Thus, as recommended by May and Finch (2009), we will examine processes of normalisation from the outset of the intervention implementation. We will then aim to test this emerging theory by using it to predict which schools will most successfully continue to implement the intervention in the third year of the trial, when schools are asked to continue using the manual, needs data, action group and curriculum materials to deliver Learning Together but without access to our external facilitators. Thus the third year provides a unique opportunity for evaluators to assess normalisation under somewhat 'natural' conditions.

References

Aber, personal communication to C Bonell

Baranowski, T. and Stables, G. (2000) Process Evaluations of the 5-a-Day Projects, *Health Educ Behav*, 27, 2, 157-166.

Baranowski, T., & Stables, G. (2000). Process evaluations of the 5-a-day projects. *Health Education & Behavior*, 27(2), 157-166.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6), 1173.

Bazemore, G. (2001). Young People, Trouble, and Crime Restorative Justice as a Normative Theory of Informal Social Control and Social Support. *Youth & Society*, 33(2), 199-226.

Bernal, G., & Sáez-Santiago, E. (2006). Culturally centered psychosocial interventions. *Journal of Community Psychology*, 34(2), 121-132.

Bitel, Mark (2001) *The evaluation of Year 3 of Leicestershire Mediation Services Conflict Resolution in Schools programme*. London: Charities Evaluation Services.

Bond, L., Patton, G., Glover, S., Carlin, J. B., Butler, H., Thomas, L., & Bowes, G. (2004). The Gatehouse Project: can a multilevel school intervention affect emotional wellbeing and health risk behaviours?. *Journal of Epidemiology and Community Health*, 58(12), 997-1003.

Bond, L., Glover, S., Godfrey, C., Butler, H., & Patton, G. C. (2001). Building capacity for system-level change in schools: lessons from the Gatehouse Project. *Health education & behavior*, 28(3), 368-383.

Bond, L., Thomas, L., Coffey, C., Glover, S., Butler, H., Carlin, J. B., & Patton, G. (2004). Long-term impact of the gatehouse project on cannabis use of 16-year-olds in Australia. *Journal of School Health*, 74(1), 23-29.

Bonell C, Allen L, Christie D, Elbourne D, Fletcher A, Grieve R, Legood R, Mathiot A, Scott S, Wiggins M, Viner R (2014) Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for cluster randomised controlled trial. *Trials* 15: 381.

Bonell C, Fletcher A, Fitzgerald-Yau N, Hale D, Allen E, Elbourne D, Jones R, Bond L, Wiggins M, Miner A, Legood R, Scott S, Christie D, Viner R. A pilot randomised controlled trial of the INCLUSIVE intervention for initiating change locally in bullying and aggression through the school environment: final report. *Health Technology Assessment* (in press).

Bonell, C., Farah, J., Harden, A., Wells, H., Parry, W., Fletcher, A., ... & Moore, L. (2013). Systematic review of the effects of schools and school environment interventions on health: evidence mapping and synthesis. *Public Health Research, 1*(1).

Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. (2006). Trials of health interventions and empirical assessment of generalizability: suggested framework and systematic review. *British Medical Journal, 333*: 346-349.

Bonell C, Sorhaindo A, Allen E, Strange V, Wiggins M, Fletcher A, Oakley A, Bond L, Flay B, Patton GC, Rhodes T. (2010a) Pilot multi-method trial of a school-ethos intervention to reduce substance use: building hypotheses about upstream pathways to prevention *Journal of Adolescent Health, 47*(6): 555-563.

Bonell C, Sorhaindo A, Strange V, Wiggins M, Allen E, Fletcher A, Oakley A, Bond L, Flay B, Patton G, Rhodes T. (2010b). A pilot whole-school intervention to improve school ethos and reduce substance use. *Health Education, 110*(4): 252-272.

Bourgois, P. (2003). *In search of respect: Selling crack in El Barrio* (Vol. 10). Cambridge University Press.

Bumbarger, B., & Perkins, D. (2009). After randomized trials: Issues related to dissemination of evidence-based interventions. *Journal of Children's Services, 3*(2).

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. and Balain, S. (2007). A conceptual framework for implementation fidelity, *Implementation Science, 2*, 1, 40.

Castro, F. G., Barrera Jr, M., & Martinez Jr, C. R. (2004). The cultural adaptation of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science, 5*(1), 41-45.

Cousins, L. H. (1997). Toward a sociocultural context for understanding violence and disruption in Black urban schools and communities. *J. Soc. & Soc. Welfare, 24*, 41.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *Bmj*, 337.

Dance, L. J. (2002). *Tough fronts: The impact of street culture on schooling*. Psychology Press.

Dane, A.V. and Schneider, B.H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review*, 18, 23–45.

Dishion, T.J., J. McCord, and Poulin F. (1999), *When interventions harm*. *American Psychologist*, 54(9): p. 755-764.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American journal of community psychology*, 41(3-4), 327-350.

Durlak, J.A. and Dupre, E.P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–50.

Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science*, 5(1), 47-53.

Flay BR, Graumlich S, Segawa E, Burns JL, Holliday MY (2004). Effects of 2 prevention programs on high-risk behaviors among African American youth: a randomized trial. *Arch Pediatr Adolesc Med*, 158(4):377.

Fletcher, A., Bonell, C., Sorhaindo, A., & Rhodes, T. (2009). Cannabis use and 'safe' identities in an inner-city school risk environment. *International Journal of Drug Policy*, 20(3), 244-250.

Gittelsohn, J., Merkle, S., Story, M., Stone, E. J., Steckler, A., Noel, J., ... & Ethelbah, B. (2003). School climate and implementation of the Pathways study. *Preventive medicine*, 37, S97-S106.

Glasgow, R.E., Klesges, L.M., Dzewaltowski, D.A., Estabrooks, P.A. and Vogt, T.M. (2006) Evaluating the impact of health promotion programs: using the RE-AIM framework to form summary measures for decision making involving complex issues, *Health Education Research*, 21, 5, 688-694.

Glasgow, R.E., McKay, H.G., Piette, J.D. and Reynolds, K.D. (2001) The RE-AIM framework for evaluating interventions: what can it tell us about approaches to chronic illness management?, *Patient Education and Counseling*, 44, 2, 119-127.

Glasgow, R.E., Vogt, T.M. and Boles, S.M. (1999) Evaluating the public health impact of health promotion interventions: the RE-AIM framework, *Am J Public Health*, 89, 9, 1322-1327.

Glover S, et al. (2001) Gatehouse Project: Teaching Resources for Emotional Well-being. Victoria, Australia: the Centre for Adolescent Health.

Glover, S. and H. Butler, *Facilitating health promotion within school communities*, in *Hands-on Health Promotion*, R. Moodie and A. Hulme, Editors. 2004, IP Communications: Melbourne.

Grant, A., Treweek, S., Dreischulte, T., Foy, R. and Guthrie, B. (2013) Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting, *Trials*, 14, 1, 15.

Hawe, P., Shiell, A., & Riley, T. (2009). Theorising interventions as events in systems. *American journal of community psychology*, 43(3-4), 267-276.

Jamal F, Fletcher A, Harden A, Wells H, Thomas J, Bonell C (2013). The school environment and student health: a meta-ethnography of qualitative research. *BMC Public Health.*, 13: 798. DOI: 10.1186/1471-2458-13-798.

Kean, J, Lloyd, G., McCluskey, G., Riddell, S., Stead, J., & Weedon, E. (2007). *Restorative Practices in Three Scottish Councils: Final Report of the Evaluation of the First Two Years of the Pilot Projects 2004-2006*. Edinburgh: Scottish Executive.

Kokotsaki, D (2013) A review of the Implementation of Restorative Approaches and its Outcomes within Durham Local Authority. Centre for Evaluation and Monitoring, Durham University.

Markham, W. A., & Aveyard, P. (2003). A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice. *Social science & medicine*, 56(6), 1209-1220.

May, C., & Finch, T. (2009). Implementing, embedding, and integrating practices: an outline of normalization process theory. *Sociology*, 43(3), 535-554.

Merton, R. K. (Ed.). (1968). *Social theory and social structure*. Simon and Schuster.

Mihalic, S. (2004). The importance of implementation fidelity. *Emotional & Behavioral Disorders in Youth*, 4, 83–105. Patton, G., et al., *Promoting social inclusion in schools: group-randomized trial of effects on student health risk behaviour and well-being*. *Am J Public Health*, 2006. 96(9): p. 1582-1587.

Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, Moore L, O’Cathain A, Tinati T, Wight D, Baird J. *Process evaluation of complex interventions: Medical Research Council guidance*. MRC Population Health Science Research Network, London, 2014

Morrison, B. (2005). Restorative justice in schools. *New directions in restorative justice*, 26-52.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years*. Open Books.

Patton GC, Bond L, Carlin JB, Thomas L, Butler H, Glover S, Catalano R, Bowes G (2006). Promoting social inclusion in schools: a group-randomized trial of effects on student health risk behavior and well-being. *Am J Public Health*, 96(9):1582.

Paulle, B. (2013). *Toxic Schools: High-poverty Education in New York and Amsterdam*. University of Chicago Press.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.

Pepi, C. L. O. H. (1998). Children without childhoods: A feminist intervention strategy utilizing systems theory and restorative justice in treating female adolescent offenders. *Women & Therapy*, 20(4), 85-101.

Rutter, M., et al., *Fifteen Thousand Hours: Secondary Schools and their Effects on Children*. 1979, London: Open Books.

Rutter, M., Maugham, B., Mortimore, P., Ouston, J., 1979. *Fifteen Thousand Hours: Secondary Schools and their Effects on Children*. Open Books, London.

Sammons, P., Day, C., Kington, A., Gu, Q., Stobart, G., & Smees, R. (2007). Exploring variations in teachers' work, lives and their effects on pupils: key findings and implications from a longitudinal mixed-method study. *British Educational Research Journal*, 33(5), 681-701.

Sammons, P., Gu, Q., Day, C., & Ko, J. (2011). Exploring the impact of school leadership on pupil outcomes: Results from a study of academically improved and effective schools in England. *International Journal of Educational Management*, 25(1), 83-101.

Sammons, P., Sylva, K., Melhuish, E. C., Siraj-Blatchford, I., Taggart, B., Smees, R., ... & Toth, K. (2012). Effective Pre-school, Primary and Secondary Education 3-14 Project (EPPSE 3-14): influences on students' dispositions in Key Stage 3: exploring enjoyment of school, popularity, anxiety, citizenship values and academic self-concept in Year 9.

Skinns, L., Du Rose, N., & Hough, M. (2009). An evaluation of Bristol RAiS. *London: ICPR, King's College London*.

Steckler, A. and Linnan, L. (2002) *Process evaluation for public health interventions and research*, San Francisco: Jossey-Bass.

Steckler, A. B., Linnan, L., & Israel, B. A. (2002). *Process evaluation for public health interventions and research* (pp. 1-23). San Francisco: Jossey-Bass.

van Thiel, S. and F.L. Leeuw, *The performance paradox in the public sector*. *Public Performance & Management Review*, 2002. 25(3): p. 267-281.

Waldron, L. M. (2005). The Messy Nature of Discipline and Zero Tolerance Policies: Negotiating Safe School Environments among Inconsistencies, Structural Constraints and the Complex Lives of Youth. *Sociological Studies of Children and Youth*, 11, 81-114.

Youth Justice Board for England and Wales, *National Evaluation of the Restorative Justice in Schools Programme*. 2004, London: Youth Justice Board for England and Wales

Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial

Chris Bonell¹, Elizabeth Allen², Deborah Christie³, Diana Elbourne², Adam Fletcher⁴, Richard Grieve², Rosa LeGood², Anne Mathiot⁵, Stephen Scott⁶, Meg Wiggins¹ and Russell M Viner^{5*}

Abstract

Background: Systematic reviews suggest that interventions that address school organisation are effective in reducing victimisation and bullying. We successfully piloted a school environment intervention modified from international studies to incorporate 'restorative justice' approaches. This trial aims to establish the effectiveness and cost-effectiveness of the INCLUSIVE intervention in reducing aggression and bullying in English secondary schools.

Methods: Design: cluster randomised trial.

Participants: 40 state-supported secondary schools. Outcomes assessed among the cohort of students in year 8 (n = approximately 6,000) in intervention year 1.

Intervention: INCLUSIVE is a school-led intervention which 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 and staff supported by an external facilitator to review local data on needs, determine priorities, and develop and implement an action plan for revising relevant school policies/rules and other actions to improve relationships at school and reduce aggression; staff training in restorative practices; and a new social and emotional skills curriculum. The intervention will be delivered by schools supported in the first two years by educational facilitators independent of the research team, with a third locally facilitated intervention year.

Comparator: normal practice.

Outcomes: primary: 2 primary outcomes at student level assessed at baseline and at 36 months:

1. Aggressive behaviours in school: Edinburgh Study of Youth Transitions and Crime school misbehaviour subscale (ESYTC)
2. Bullying and victimisation: Gatehouse Bullying Scale (GBS)

Secondary outcomes assessed at baseline, 24 and 36 months will include measures relating to the economic evaluation, psychosocial outcomes in students and staff and school-level truancy and exclusion rates.

Sample size: 20 schools per arm will provide 90% power to identify an effect size of 0.25 SD with a 5% significance level.

Randomisation: eligible consenting schools will be randomised stratified for single sex versus mixed sex schools, school-level deprivation and measures of school attainment.

(Continued on next page)

* Correspondence: r.viner@ucl.ac.uk

⁵UCL Institute of Child Health, 30 Guilford St, London WC1N 1EH, UK
Full list of author information is available at the end of the article



© 2014 Bonell et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

(Continued from previous page)

Discussion: The trial will be run by independent research and intervention teams and supervised by a Trial Steering Committee and a Data Monitoring Committee (DMC).

Trial registration: Current Controlled Trials ISRCTN10751359 (Registered 11 March 2014)

Keywords: Bullying, Cluster randomised trial, School intervention, Violence prevention, Adolescent

Background

The prevalence and harms of aggressive behaviours among youth make addressing them a public health priority [1-4]. The World Health Organisation considers bullying to be a major adolescent health problem, defining this to include the intentional use of physical or psychological force against others [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 reporting that they have been subjected to serious peer bullying [9]. There are marked social gradients, with both family deprivation and school-level deprivation increasing the risk of experiencing bullying [10]. Bullying most commonly occurs in schools [11,12] and prevalence varies significantly between schools [13-16].

Being a victim of peer bullying is associated with an increased risk of: physical health problems [17]; engaging in health risk behaviours such as substance use [18-20]; long-term emotional, behavioural and mental health problems [21-23]; self-harm and suicide [24]; and poorer educational attainment [25,26]. Students who experience physical, verbal and relational bullying on a regular basis tend to experience the most adverse health outcomes [27]. There is also evidence suggesting that childhood exposure to bullying and aggression may also influence life-long health through biological mechanisms [28]. The perpetrators of peer bullying are also at greater risk of a range of adverse emotional and mental health outcomes, including depression and anxiety [8,13].

Bullying is also often a precursor to more serious violent behaviours commonly reported by British youth. One UK study of 14,000 students found that 1 in 10 young people aged 11 to 12 reported carrying a weapon and 8% of this age group admitted they had attacked another with the intention to hurt them seriously [29]. By age 15 to 16, 24% of students report that they have carried a weapon and 19% reported attacking someone with the intention to hurt them seriously [29]. Inter-personal 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 anti-social behaviours in youth and violent crime in adulthood [30,31]. There is increasing concern because low-level provocation and aggressive behaviours in secondary schools are educationally

disruptive, emotionally harmful, reduce educational attainments and later life-chances, and can lead to more overt physical aggression over time [32-34]. The economic costs to society as a whole due to 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 £60 billion a year in England and Wales [35].

School-based interventions

Reducing aggression, bullying and violence in British schools has been a consistent priority within recent public health and education policies [36-38]. The 2009 Steer Review concluded that schools' approaches to discipline, behaviour management and bullying prevention vary widely and are rarely evidence-based, and that further resources and research are urgently needed to combat aggressive behaviours and other conduct problems [34]. There is, therefore, a pressing 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.

A number of systematic reviews assess school-based interventions to address bullying and aggression. Interventions that promote change across school systems and addressed different levels of school organisation, that is 'whole-school' or 'school environment' interventions, are particularly effective in reducing victimisation and bullying than curriculum interventions [39-41]. The effectiveness of such interventions may be because they address bullying as a systemic problem meriting an 'environmental solution' [39]. Whole-school interventions are also inherently universal in reach and likely to provide a cost-effective and non-stigmatising approach to preventing bullying [40]. 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 [42,43].

These school environment interventions thus take a 'socio-ecological' [44] or 'structural' [45] approach to promoting health whereby behaviours are understood to be influenced not only by characteristics of individuals, but also the wider social context. A recent National Institute of 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' (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 [43]. Other outcomes that are improved by school environment interventions include mental health and physical activity and reduced substance use including alcohol, tobacco and drugs [43].

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, due to their potential for modifying population-level risk as well as their reach and sustainability [43]. Multiple risk behaviours in adolescence are subject to socio-economic stratification, and are strongly associated with poor health outcomes, social exclusion, educational failure and poor mental health in adult life [46]. 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 [47].

The INCLUSIVE intervention under trial here has been particularly informed by two international evidence-based school environment programmes. First, the Aban Aya Youth Project (AAYP) is a multi-component intervention, enabling schools to modify their social environment as well as delivering a social skills curriculum. This approach was 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 in violence and aggressive behaviour [48]. The intervention also brought benefits in terms of reduced 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 via 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 randomised controlled trial (RCT) found consistent reductions in a composite measure of

health risk behaviours, which included violence and anti-social behaviour [49,50].

INCLUSIVE extends the AAYP and Gatehouse interventions by including 'restorative justice' approaches. The Steer Review in 2009 called for English schools to consider adopting more restorative approaches to prevent bullying and other aggressive behaviour to help minimise the harms associated with such problems [34]. The central tenet of such approaches 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 practice currently employed in schools is the use of 'circle time' to develop and maintain good communication and relationships [51]. Restorative 'conferencing' can also be used in schools to deal with more serious incidents [51].

Restorative approaches have only been evaluated using non-random designs, although such studies do suggest that the restorative approach is a promising one in the UK [52-54] and internationally, particularly when implemented at the whole-school level [55-57]. For example, in England and Wales, the Youth Justice Board evaluated the use of restorative approaches at twenty 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. Restorative approaches thus appear to have the potential to complement school-environment interventions such as Aban Aya and the Gatehouse Project. They offer a highly promising way forward for reducing aggressive behaviours among British youth. A recent Cochrane review found no RCTs of interventions employing restorative approaches to reduce bullying in schools and recommended that this should be a priority for future research [58]. 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.

Findings from the INCLUSIVE pilot study

The evidence above demonstrates that bullying and aggression are highly prevalent in English schools, and generate health harms and inequalities, educational and other harms, and economic costs. While existing systematic reviews suggest 'whole-school' interventions are an effective approach to addressing these problems, the recent Cochrane review [58] recommends further trials in this area examining restorative practices. The INCLUSIVE intervention addresses

these points and has been successfully piloted, funded for 20 months (July 2011 to February 2013) through a commissioned funding call from the UK NIHR Health Technology Assessment (HTA). Criteria were agreed for progression to a full trial, with further funding for a phase III trial of a three-year intervention being dependent on a new funding application. Intervention funding was provided by the Paul Hamlyn Foundation, the Big Lottery Fund, and the Coultts Charitable Trust.

We undertook a cluster RCT in eight mixed-sex secondary schools in London and south-east England, purposively sampled to ensure diversity with regard to Ofsted rating and rate of eligibility for free school meals (four intervention, four comparison) with integral process evaluation. The aim was to assess the feasibility and acceptability of the INCLUSIVE intervention and trial methods over one academic year (whereas INCLUSIVE was designed as a three-year intervention). The objectives of the study were to: (1) examine the feasibility and acceptability of delivering and trialling the intervention according to pre-specified criteria agreed with the HTA; (2) explore participants' experiences of implementing and trialling the intervention and how this varied according to school context to refine the intervention and trial methods; and (3) pilot indicative primary outcomes (aggressive behaviour measures), other outcomes and economic evaluation methods.

All pre-specified feasibility and acceptability criteria were met (objective 1) and the process data indicated that all intervention components, the trial design and methods were feasible and acceptable (objective 2). Qualitative data suggested that student participation may be a core component in improving relationships and engagement across the school. Appropriate outcome measures and economic methods were identified (objective 3): the Gatehouse Bullying Scale (GBS) and the Edinburgh Study of Youth Transitions and Crime (ESTYC) school misbehaviour subscale were acceptable, discriminating and reliable measures of bullying and aggression in this context. Pilot economic analyses support the use of the Child Health Utility 9D (CHU9D) scale with this population and the feasibility of cost-utility analysis. Analysis of outcomes in the pilot showed that confidence intervals encompassed potential intervention benefits. There was no evidence of harm.

We were then successful in obtaining further NIHR funding from the Public Health Research programme (PHR) to undertake a large-scale cluster RCT to examine the effectiveness and cost-effectiveness of the INCLUSIVE intervention. Intervention funding was obtained from the Educational Endowment Fund (EEF), which also funded an independent evaluation of effects on educational attainment to be conducted by the University of Manchester.

Research questions

RQ1. Is the INCLUSIVE intervention implemented over three school years more effective and cost-effective than standard practice in reducing bullying and aggression among 12- to 15-year olds in English secondary schools?

RQ2. Is the INCLUSIVE intervention more effective than standard practice in improving students' quality of life (QoL), well-being, psychological function and attainments, and reducing school exclusion and truancy, substance use, sexual risk, National Health Service (NHS) use, police contacts among students, and improving staff QoL and attendance and reducing burn-out?

RQ3. What pre-hypothesised factors moderate and mediate the effectiveness of the INCLUSIVE intervention; including, do effects vary by socio-economic status and sex?

Methods

The trial is a 3-year cluster randomised controlled trial with integral economic evaluation and process evaluation in 40 schools across south-east England, with schools as the unit of allocation.

Study population

INCLUSIVE is a universal intervention, aimed at all 11- to 16-year olds in participating secondary schools in England. While the intervention will have effects on the whole school, our study population of students will be those at the end of year 7 (age 11 to 12 years) at baseline and at the end of year 10 at 36-month follow-up (age 14 to 15), as well as all school teaching and teaching assistant staff. All students in the school in that year and all teaching staff will be surveyed at each time-point, not only those who participated at baseline.

Inclusion/Exclusion criteria

Eligible schools are those:

- (i). Secondary schools within the state education system (including community, academy or free schools, and mixed or single sex) in south-east England. We will take the widest definition of a 'state school' and will only exclude private schools, schools exclusively for those with learning disabilities and pupil referral units. The latter two will be excluded as it is unlikely that INCLUSIVE will be appropriate for their populations.
- (ii). Ofsted rating (most recent) of 'requires improvement'/'satisfactory' or better; we will exclude schools with an 'inadequate'/'poor' Ofsted rating because such schools are subject to special measures which are likely to impede INCLUSIVE delivery.

Note there are no inclusion/exclusion criteria for students.

Recruitment

Schools will be recruited from secondary schools in Greater London and the surrounding counties (Surrey, Kent, Essex, Hertfordshire, Buckinghamshire, and Berkshire) with a maximum travel time of one hour from the study centres in London. To aid recruitment, we will partner with existing schools networks such as the UCL Partners Schools Network, the Institute of Education Teaching Schools and schools that are part of our collaborating schools network, Challenge Partners. We will approach approximately 500 eligible schools, initially by letter and Email with a telephone follow-up, complying with good practice and research governance for undertaking studies within the education system.

Randomisation

Eligible schools whose head-teacher gives informed written consent to participate will be allocated with a 1:1 ratio between intervention and control arms. Stratified randomisation will be undertaken remotely by the Clinical Trials Unit (CTU) at the London School of Hygiene & Tropical Medicine (LSHTM). To promote baseline equivalence, we will stratify by key school-level determinants of violence:

- a. Single sex versus mixed sex school.
- b. School-level deprivation, as measured by percentage of students eligible for free school meals (low/moderate 0 to 23%; high >23%, with 23% being the median for England).
- c. School 'best eight value added' in GCSE exams (above and below median for England of 1,000). Value added (VA) score is a school-level measure of students' attainment in public exams adjusting for their attainment on entry to the school. We use VA rather than Ofsted ratings for schools as there is better evidence for VA being associated with violence rates [59].

Schools will be allocated randomly within each of these eight strata.

Protecting against selection bias:

- (1) School level: the randomisation schedule will be drawn up once the schools have consented and after the baseline survey, thus guarding against selection biases at entry of clusters to the trial. The randomisation may occur sequentially in groups of 10 schools, should there be any delays with baseline surveys in some schools. As with most social intervention trials, schools, their students, teachers and other staff cannot be 'blinded' to allocation status. However, fieldwork staff will be blinded to allocation as will data-input staff. Analysis of follow-

up quantitative data will be undertaken blind to allocation.

Retention of control schools will be maximised by ensuring regular senior liaison and provision of participation incentives (£500 per school).

- (2) Student level: we had very high student participation in our pilot study: 96% of eligible at baseline and 93% at follow-up. To minimise bias, we will use in-school, mail and telephone contacts to try to include all enrolled students absent at either baseline or follow-up questionnaires. Note we will not attempt to follow-up students who have left the school.

A flow chart of recruitment and intervention and control treatment is shown in Figure 1.

Intervention and comparison groups:

1. Intervention:

The INCLUSIVE intervention is intended principally to augment rather than to replace existing activities (for example, training, curricula, and so on) in intervention schools. However, it is intended to replace existing non-restorative disciplinary school policies and practices where restorative approaches are deemed by the action group to be more appropriate.

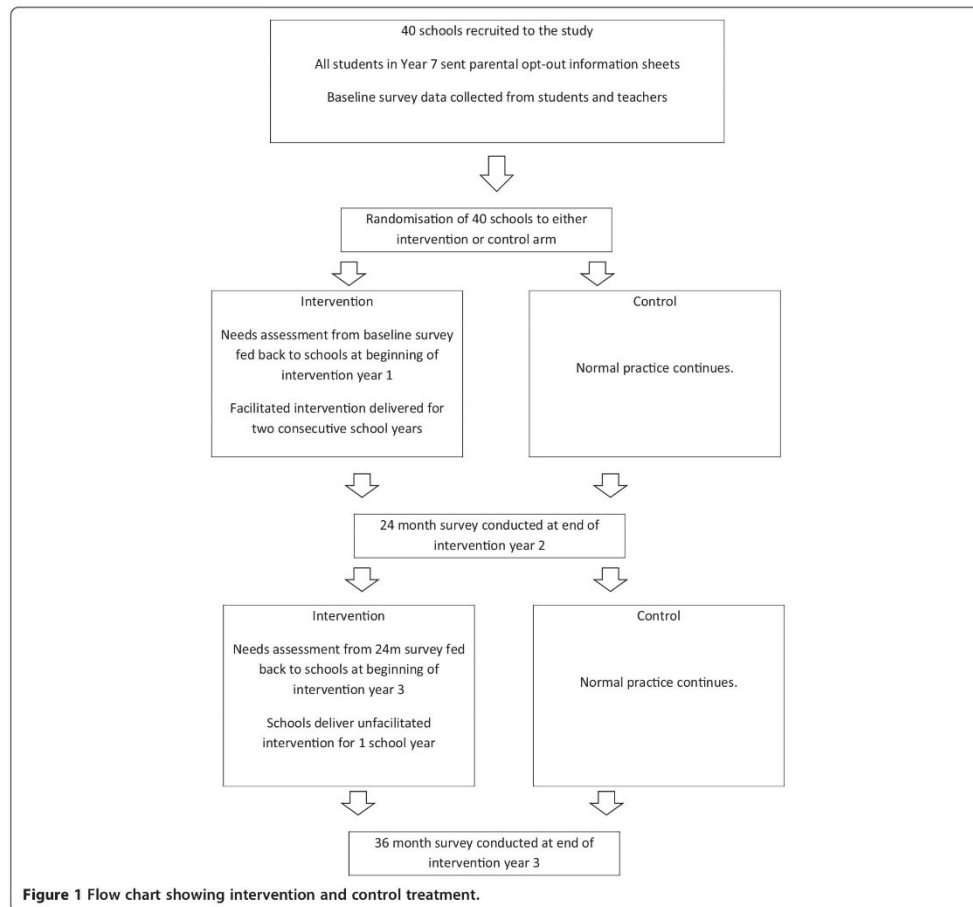
The facilitated phase provides the following inputs:

- i). Annual surveys of local needs and assets (including bullying, aggression, prevalence and determinants) and progress in addressing these.
- ii). Support from an external expert education facilitator trained in facilitating INCLUSIVE.
- iii). Social and emotional learning curriculum resources.
- iv). Staff training in restorative practices provided by the education facilitators and comprising a short introduction and subsequent half day for all staff (focused on introducing them to restorative practices, such as 'circle-time', to promote positive relationships and communication, plus enhanced three-day training course in restorative practices targeting five to ten staff at each school, including training in formal 'conferencing' to deal with more serious incidents via bringing together students, parents and/or staff.)

These inputs will enable schools during all three years to convene an action group, which comprises (at a minimum):

1. Six students
2. Six staff, including at least one Senior Management Team (SMT) member and one member of each of the teaching, pastoral and support staff

Membership from specialist health staff, such as the school nurse and/or local child and adolescent mental



health services staff, are desirable but optional. The action group must meet at least six times per school year (that is approximately once every half-term).

The action group develops an action plan that coordinates delivery of the following intervention outputs:

- i). Reviewing and revising *school rules and policies* relating to discipline, behaviour management and staff-student communication.
- ii). Implementing *restorative practices* throughout the school. Restorative practices include 'circle-time' (which brings students together with their teacher during registration periods or other lessons to maintain good relationships, or be used to deal with specific problems) and 'conferencing' (used to

deal with more serious incidents and brings together relevant staff, students, parents and, where necessary, external agencies).

- iii). Additional *tailored actions* to address local priorities.
- iv). Delivering the *social and emotional skills curriculum* for years eight to ten. The curriculum targets students in years eight to ten who receive five to ten hours teaching and learning per year on restorative practices, relationships, and social and emotional skills based on the Gatehouse Project curriculum. The curriculum is designed as a set of learning modules which schools can address using our own or existing materials if these aligned with our curriculum. Modules cover: 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 will tailor the curriculum to their needs and could deliver modules either as 'stand-alone' lessons, for example within Personal, Social and Health Education (PSHE), and/or integrated into various subject lessons (for example, English).

The intervention enables local tailoring, informed by the needs survey and other local data sources. These locally adaptable actions occurred within a standardised overall process with various core standardised intervention elements, such as the staff training in restorative practices; review and revision of school rules and policies; and the social and emotional skills curriculum. This balance of standardisation and flexibility is a common practice in complex interventions, enabling a balance between fidelity of the core components with local adaptation [60]. This allows schools to build on their current good practice, and also encourages students and staff to develop ownership of the work, which may be a key factor in intervention effects. To support this, the facilitator works with schools to ensure all members of the action group are supported to identify and undertake locally determined actions to improve the school environment.

Internally facilitated intervention year: the third intervention year will be identical to the externally facilitated intervention described above, with the exception that there will be no provision of external facilitation. One of the roles of the external facilitator over the two facilitated years will be to ensure the school action group and SMT develop the capacity to undertake this internal facilitation in the third year.

2. Comparator - control schools:

Schools randomised to the control group will continue with normal practice for the school in question and receive no additional input. They will be provided with £500 (to cover administrative costs and/or provide cover for staff involvement in organising data collection) and at the end of the study be offered a brief report of the survey data collected at the school. Control schools are free to engage in actions to reduce bullying and aggression but the contract signed with head-teachers will preclude their engaging in a facilitated whole-school programme similar to INCLUSIVE during the period of the trial. We will examine control schools' policies and practices related to bullying and aggression.

Endpoints of the study

Primary outcomes

The primary outcome will be an assessment of experience of violence and bullying measured using 2 scales at

36 months through student survey self-reports. As is conventional in trials of interventions addressing violence and aggression in schools, we will rely on self-reports from students, rather than observations or teacher reports, because of the impracticality and greater likelihood of bias respectively of the latter two. The primary outcomes measures include one measure of bullying victimisation and one measure of perpetration of aggressive behaviours that were shown to be reliable and valid in our pilot study:

- a. *Gatehouse Bullying Scale (GBS)*. The GBS [49] is a short, reliable tool to measure the occurrence of bullying victimisation in schools. This measure was designed by one of our collaborators (LB) and has been shown to be related to other measures of social attachments, school engagement, and anxiety and depressive symptoms. The scale has 12 items, and asks about being the subject of recent teasing, name-calling, rumours, being left out of things and physical threats or actual violence from other students in the last 3 months. Each section asks about the recent experience of that type of bullying ('yes' or 'no'), how often it occurred, and how upset the student was by each type of bullying [49,61].
- b. *Edinburgh Study of Youth Transitions and Crime (ESYTC) school misbehaviour subscale*. The ESYTC measures several domains of violence and aggression at school [62].

Secondary outcomes

These will include our aggression/bullying measures (GBS and ESYTC) measured at 24 months and other outcomes measured at both 24 and 36 months:

- (i) Student-self-report outcomes: these will be measured through student survey self-reports:
 1. Paediatric quality of life inventory (PedsQL) version 4.0 will be used to assess overall QoL. The 30-item PedsQL [63] has been shown to be a reliable and valid measure of QoL in normative adolescent populations. It consists of 30 items representing five functional domains: physical, emotional, social, school and well-being, and yields a total QoL score, two summary scores for 'Physical Health' and 'Psychosocial Health' and three subscale scores for 'Emotional', 'Social', and 'School' functioning.
 2. Psychological function and well-being:
 - a. The Strengths and Difficulties Questionnaire (SDQ) [61] is a brief screening instrument for detecting behavioural, emotional and peer problems and pro-social strengths in children and adolescents. It is

brief, quick to complete, and validated in national UK samples.

- b. Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) [64] is a seven-item scale designed to capture a broad concept of positive emotional well-being including psychological functioning, cognitive-evaluative dimensions and affective-emotional aspects, with a total 'Well-Being Index' generated.

3. Risk behaviours;

- a. Substance use. Validated age-appropriate questions taken from national surveys and/or previous trials will be used to assess smoking (smoking in previous week; ever smoked regularly), alcohol use (use in previous week; number of times really drunk; binge drinking) and illicit drug use (last month; lifetime use).
- b. Sexual risk behaviours: age of sexual debut and use of contraception at first sex may be examined by measures used in the Ripple trial [65]. We will consult with schools about the acceptability of asking these questions at follow-up (year ten).

4. Use of NHS services: self-report use of primary care, accident and emergency, other service in past 12 months.

5. Contact with police will be assessed using the Young People's Development Programme (YPDP) evaluation measure [66], which asks whether the young person has been stopped, told off, or picked up by the police in the last 12 months.

(ii) Student-level data collected from schools:

1. School attendance will be measured via routine school data on each student expressed as number of half days absent; for which we will seek students' informed consent to access.
2. Educational attainment: this will be assessed by an independent team based at the University of Manchester drawing on routine data.

(iii) Individual staff-level outcomes. We will measure the following secondary outcomes through survey self-reports from teachers and teaching assistants:

1. Staff attendance will be measured via routine school data on each staff-member expressed as number of half days absent; for which we will seek staff-members' informed consent to access.
2. Staff QoL will be measured using the Short Form (SF)-12 version 2 Health Survey [67], a brief well-validated measure of adult health-related QoL.
3. Staff stress and burnout will be measured using the Maslach Burnout Inventory [68], an established scale

which uses a three-dimensional description of exhaustion, cynicism, and inefficacy.

(iv) School-level outcomes: routinely-collected data on school rates of temporary and permanent exclusions.

Student surveys will be conducted in exam conditions in schools, maximizing privacy. All students in the school in that year and all teaching and teaching assistant staff will be surveyed at each time-point, not only those who participated at baseline. Paper-based questionnaires will be completed confidentially in a 45-minute class session devoted to the purpose. Field workers will supervise the class completing the questionnaire, with the teacher present (for disciplinary purposes) but unable to see the questionnaires. The field-workers will assist students with questions that they do not understand and ensure students complete as much of the questionnaire as possible. Note that students with mild learning difficulties or with limited command of written English will be supported to complete the questionnaires by fieldworkers.

We will ask students in intervention schools involved in qualitative interviews whether their reporting (as opposed to their experience) of bullying and aggression might have been affected by the intervention.

Power and sample size

The average English school has approximately 190 students per year, although this varies across schools. A systematic review of school-based secondary preventive interventions to prevent violence [69] reported a pooled effect size of 0.41 on measures of aggressive behaviour. Effect sizes for aggressive behaviour from similar interventions approximate 0.3 to 0.4 SDs in males. Recent data from three large UK school cohorts [70] suggest that intra-cluster correlation coefficients (ICC) for aggression and bullying outcomes vary between 0.01 and 0.03

We propose to recruit sufficient participants to detect a difference between groups of 0.25 SD with 90% power and a 5% level of significance. This is considered to represent a moderate size of effect and in line with the effect sizes seen in the literature.

Conservatively, taking an ICC of 0.04 and 150 students per school, a trial involving 20 schools per arm will provide 90% power to identify an effect size of 0.25 SD with a 5% significance level. If two schools per arm (that is 10%) were to be lost to follow-up over the course of the trial, we would still have 80% power to detect an effect size of 0.25 SD.

The total student sample size will be approximately 6,000. As we will be surveying all young people in the relevant school year at each follow-up, this sample is likely to remain similar across the study.

Economic evaluation

The aim of the economic evaluation is to assess the costs, consequences and cost-effectiveness of the INCLUSIVE intervention compared with standard school-based practices for managing aggression.

The primary economic evaluation will take the form of a within-trial cost-consequence analysis, with a secondary analysis that will report relative cost-utility with health outcomes expressed in terms of Quality-Adjusted Life-Years (QALYs), as recommended by the National Institute for Health and Care Excellence (NICE)'s public health methods guidance.

This NICE guidance also 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. This statement is particularly pertinent to INCLUSIVE as the costs of implementing it are likely to fall on the educational sector, yet there are potential cost implications for sectors such as the NHS, the police and the judiciary through reduced anti-social behaviour.

The costs to the education sector include cost of the facilitator to deliver the intervention and the cost of staff time. The facilitator costs for the delivery of the intervention will be collected using log sheets. The impact on staff time for training and delivering bullying policy will be obtained as part of the process evaluation. It is possible that the intervention might offset some of the staff time related to dealing with pupil aggression or bullying behaviour and this will be captured as part of the teacher survey. It might also impact on teacher health and we will capture this by valuing the number of days off work, which will be captured as a secondary outcome measure. The implications for NHS resource use and policing will be identified with specific questions in the student survey and valued accordingly. The time horizon will capture costs and outcomes within the trial.

Changes in health-related QoL (as expressed using QALYs) will be measured from the study participants' (that is student's/teacher's) perspective.

The *Child Health Utility (CHU) 9D measure* (CHU-9D) [71] will be used to assess student's health-related QoL as part of the economic evaluation. The CHU-9D is a validated age-appropriate measure that was explicitly developed using children's input and has been suggested to be more appropriate and function better than other health utility measures for children and adolescents. For teachers, we will use the SF-12 for this purpose [67]. Student and teacher utility values will be collected (at baseline and at follow-up surveys at 24 and 36 months) using the CHU-9D and by converting the SF-12 questionnaires respectively.

Process evaluation

Data will be used to examine intervention implementation and receipt and examine possible causal pathways in order to facilitate interpretation of outcome data. In line with Medical Research Council (MRC) guidance on complex interventions, this component of the trial will also enable refinement of the intervention logic model. Informed by existing frameworks, the process evaluation will examine the following:

Trial context

We will assess the context within the intervention and control arms, including what other relevant services and practices operate, such as the nature of school discipline systems, staff training, social skills curricula and student participation in decision-making. This will draw on annual: interviews with intervention facilitators (n = 5); telephone interviews with action-team members (n = 2 per school) in intervention schools; interviews with the Senior Leadership Team (SLT) (n = 1 per school) and other staff (n = 2 per school) in intervention and control schools; and 2 focus group discussions (FGDs) with students and one FGD with staff in 8 randomly selected intervention and control schools (purposively sampled by students participation, gender and age and staff participation and role), which will also allow us to explore mechanisms of actions.

Trial arm fidelity

We will assess the fidelity with which INCLUSIVE is delivered in each school. In addition to the above sources, we will draw on: annual structured quantitative researcher observational data of a random selection from each school of one action team meeting (n = 20), staff training (n = 20) and one curriculum session (n = 20); structured diaries of action team meetings and staff training maintained by intervention facilitators in each school; qualitative data from action-team minutes (from 10 randomly selected schools in the full trial). We will assess fidelity and acceptability rates for each facilitator.

Participation, reach and dose

We will assess the extent to which students and staff are involved in or in receipt of intervention processes and outputs. This will draw on quantitative data from 24- and 36-month follow-up surveys of students, staff and action group members. The last of these will also assess the extent to which members felt empowered to participate in decision-making using a modified version of the Learner Empowerment Scale [72].

Reception and responsiveness

We will assess the *experiences* of participation in INCLUSIVE and in school environments shaped by this, to

assess *acceptability and any barriers or facilitators to this*. This will draw on the annual interviews with action-team members ($n = 2$ per school) in intervention schools; interviews with SLT ($n = 2$ per school) and other staff ($n = 2$ per school) in intervention and control schools; and FGDs with students in 8 randomly selected intervention schools described above.

Intermediate outcomes

To assess possible intervention causal pathways and examine whether these mediate intervention effects in order to assess and refine our logic model, we will use two measures that examine students' perception of the school environment and their connection to the school:

- a. Beyond Blue School Climate Questionnaire (BBSCQ) [73] which will be used to measure students' perceptions of the school climate. It consists of twenty-eight items which produce an overall score and also assesses four key domains of school climate (subscale): supportive teacher relationships, sense of belonging, participative school environment, and student commitment to academic values.
- b. Student reports of anti-school actions will be assessed using the ESYTC Self-Reported Delinquency (SRD) subscale. Involvement with anti-school peer groups will be assessed using a single item measure previously used in the YPDP evaluation measure.

Analyses

Outcome analyses

All primary analyses will be carried out according to the principle of intention-to-treat (ITT) and using multilevel modelling to take into account clustering at the school level. The primary analysis will be a repeat cross-sectional analysis that includes data from all students at both time points for two main reasons: (1) the intervention is a whole school intervention and, based on a school-level theory of change, is expected to impact on all pupils, not just on those pupils who were present at baseline; (2) the literature suggests that in cluster randomised trials, when migration into or out of the clusters is high over time, the baseline cohort may not remain representative of the cluster and therefore repeated cross-sectional analysis is preferred to minimise bias. Based on our pilot data and existing research on student mobility, we anticipate student turn-over of up to 25% in some schools over 36 months. Because of this we will use multilevel analyses that include all students at all time-points, which essentially provides a repeat cross-sectional analysis with a nested longitudinal cohort.

Data will be analysed by appropriate multivariate regression models, fitting pre-hypothesised potential confounders as covariates. Note that data on ethnicity and socio-economic status will be collected by self-report from students. Both primary outcomes will be fully analysed and reported separately, using separate multi-level models. A small number of secondary analyses based on explicit hypotheses, for example, subgroup effects/causal pathway analyses will be specified in advance. These secondary analyses will include a longitudinal analysis of pupils present at both baseline and follow-up, with further analyses using individual-level baseline data to explore the implications of missing individual-level outcome data.

Secondary analyses will include staff outcomes and will be carried out according to the principle of ITT using the same approach to modelling as described for the student outcomes. Secondary analyses will also examine moderators and mediators. We will examine whether intervention effects are moderated by individual-level gender and socio-economic status measured using the Health Behaviours in School-aged Children (HBSC) Family Affluence Scale [74] and sex, as well as by school-level stratifying factors (single sex versus mixed sex school; school-level deprivation; value added strata); and facilitator, though these analyses may be underpowered. We will examine whether intervention effects are mediated by process and intermediate outcome measures. Other such analyses will be informed by hypotheses derived from analysis of qualitative data.

Economic analyses

The primary economic evaluation will be a cost-consequence analysis. We will undertake a cost-utility analysis as a secondary analysis. These analyses will be linked and use of both is consistent with NICE methods guidance for evaluating public health interventions. We propose using a multi-level modelling approach with random intercepts to estimate the mean and standard errors for both cost and effects along with the covariance matrix. From these data mean incremental net benefit and confidence intervals will then be estimated. Missing data will be handled using multiple imputation.

Process evaluation analyses

Qualitative data will be entered into the data analysis package NVivo (QSR International (UK) Limited, Vanguard House, Keckwick Lane, Daresbury, Cheshire, WA4 4AB, United Kingdom, Telephone: +44 (0) 1925 357 960) which will be used to manage and code data. Qualitative data from the process evaluation will be subjected to a thematic content analysis. Codes will be applied to transcripts, which identify key themes and how these interrelate in order to develop an analytical framework. Each transcript will be coded to indicate the type of participant,

school and date, allowing analytical themes to be explored in relation to different groups' experiences and to compare processes across schools. Drawing on methods associated with 'grounded theory', we will make constant comparisons and examine deviant cases to refine our analysis. Analysis will explore implementation and receipt and contextual factors affecting these, as well as potential causal pathways in order to develop hypotheses to examine in secondary moderator and mediator analyses. Additionally, quantitative data from surveys and observations will be used in analyses of intervention fidelity and reach using simple descriptive statistics.

Ethical issues

The study has been approved by the Institute of Education Research Ethics Committee (18/11/13 ref. FCL 566) and the University College London Research Ethics Committee (30/1/14, Project ID: 5248/001).

Consent

Written consent will be obtained at school level (head-teacher) for random allocation and for intervention, and at the individual student, staff and intervention facilitator level for data collection. For students, written age-appropriate information sheets will be provided in class one to two weeks before the baseline survey, together with oral explanation by teachers. Written consent will be required from all participating young people, which will be collected immediately before conducting the baseline survey. Young people will also be asked to take home written information sheets for parents. Parents who do not wish their child to participate will be asked to notify this opt-out in writing using a prepared form.

Confidentiality

All information collected during the trial will be kept confidential and adhere to the 1998 Data Protection Act.

Risk, burdens and benefits

Benefits

If successful, the INCLUSIVE intervention will result in the following benefits:

1. Reduction of bullying and aggression which will be of benefit to all participants, the whole school, local communities and society in general.
2. Reduction in other health-risk outcomes (for example, substance use) and improvements in mental health, emotional well-being and QoL.
3. Reduction in costs to society related to bullying and aggression. These include reductions in NHS costs (related to violence and mental health problems), and in social costs including costs within the justice system.

4. Benefits to school staff through increased access to restorative training and an improved school environment, which may improve staff well-being and QoL.
5. Benefits to students who participate in the intervention, through opportunities for learning and improved self-efficacy.

Risks

There are no anticipated risks to participants or to schools. However, as in all interventions, there may be unanticipated risks. Harms will be assessed through examination of outcomes at 24 and 36 months. An independent Data Monitoring Committee (DMC) will examine any potential harms at 24 months. If any major harms are detected, the DMC will inform the Trial Steering Committee (TSC) who will decide what action should be taken.

It is possible that our approach may be ineffective, and its introduction in trial schools may prevent the use of more effective techniques to reduce aggression. Although some educational interventions to raise awareness of risk behaviours during adolescence have been shown to increase participation in these behaviours, we believe this is extremely unlikely in the case of this study because as our approach is based upon what is shown to be effective in systematic reviews. Because of the above, we believe that risks are minimal and that benefits justify the risks.

Study governance

Trial documentation

Relevant trial documentation will kept for a minimum of 15 years.

Trial registration and conduct

The trial is registered with www.controlled-trials.com (ISRCTN 10751359); note that the ISRCTN for the pilot study was 88527078. As the trial is not within clinical settings nor using clinical samples nor using a medicinal product, there is no requirement to comply with the 'The Medicines for Human Use (Clinical Trials) Regulations 2004'. We will follow the UK MRC Guidelines on Good Clinical Practice in Clinical Trials. Note that the chief Investigators (CI) and the majority of the other investigators have been trained in Good Clinical Practice for clinical trials.

Sponsor

The UCL Institute of Child Health, the employer of one of the CIs, will act as the sponsor of this trial.

TSC: the trial will be overseen by a TSC, including an independent chair (Professor Laurence Moore, University of Glasgow), at least two other independent members, Patient and Public Involvement representatives including

young people and teachers involved in our pilot study, and an investigator representative of each institution involved in the research. Observers from the PHR programme will be invited to all TSC meetings. The TSC will meet six-monthly throughout the trial.

Data Monitoring Committee (DMC)

A DMC will be established independent of the investigators and of the TSC, but reporting to the TSC and (via the TSC) to the sponsors and the HTA programme. This will consist of an independent chair, a senior statistician and at least one other senior academic independent of the investigators. This will meet approximately yearly during the study. The DMC will monitor data for quality and completeness. Data quality, follow-up and trial monitoring will be facilitated through the development of a trial-specific database, including validation, verification, monitoring and compliance reports and follow-up report functionalities. The DMC will examine the results of an interim analysis at 24 months to consider any potential harms.

Study management

Russell Viner (RV) will direct the study together with Chris Bonell (CB) as co-CI. The intervention and research teams will be functionally independent. The research team will be managed by RV, CB and Anne Mathiot (AM), the trial manager. CB will direct the process evaluation.

The trial manager will have day-to-day responsibility for the conduct of the trial and the operations of the research team. The trial manager will report to the CIs and to a trial management group made up of RV, CB, AM together with the lead study statistician Elizabeth Allen (EA) and the lead for the intervention team, Meg Wiggins (MW). The trial management group will meet monthly throughout the study, and report to the Scientific Steering Committee (SSC) made up of all named investigators. The SSC will meet four- to six-monthly throughout the trial. Responsibility for data integrity and analysis will be held by the Clinical Trials Unit (CTU) at the London School of Hygiene & Tropical Medicine (LSHTM) (Diana Elbourne and EA). Responsibility for economic evaluation will be held by Richard Grieve at the LSHTM.

The intervention team will be managed by MW at the Institute of Education, together with Miranda Perry (MP), the intervention educational consultant who will direct day-to-day operation of the intervention and coordinate the educational facilitators.

Discussion

The INCLUSIVE trial is part of a growing number of cluster randomised trials related to health but conducted within the education system in the UK. We have built

upon evidence from US and Australian studies, modified the intervention to include restorative justice elements and shown feasibility and acceptability in a pilot study. This full trial of the INCLUSIVE intervention is a pragmatic 'realist' trial, evaluating not only the facilitated intervention (for the primary outcome) but also a further year of the intervention when continued by schools without external facilitation.

A number of elements of the trial will aid generalisability and scalability if shown to be effective. We have included a very wide range of participating schools, including all but schools whose current functioning we judge to be too low to be able to implement or benefit from the intervention. The intervention is flexible and can be tailored to each school's needs and we have partnered with a number of school networks to facilitate future scalability.

Funding was obtained from both the health sector (through the National Institute of Health Research) and the education sector (the Education Endowment Fund). Each is funding separate teams to undertake the research (health sector) and the intervention (education sector).

The trial will be overseen by an independent TSC and DMC appointed by the main funders (NIHR)

Trial status

At time of submission (2 June 2014) the trial has recruited all schools and is currently recruiting and surveying students for the baseline survey. Schools will be randomised after all baseline data are collected.

Abbreviations

AAYP: Aban Aya Youth Project; BBSCQ: Beyond Blue School Climate Questionnaire; CHU: Child Health Utility; CHU-9D: Child Health Utility 9D; CI: chief investigators; CTU: Clinical Trials Unit; DMC: Data Monitoring Committee; EEF: Educational Endowment Fund; ESTYC: Edinburgh Study of Youth Transitions and Crime; FGD: focus group discussion; GBS: Gatehouse Bullying Scale; HBSC: Health Behaviours in School-aged Children; HTA: Health Technology Assessment; ICC: intra-correlation coefficients; ITT: intention-to-treat; LSHTM: London School of Hygiene & Tropical Medicine; MRC: Medical Research Council; NHS: National Health Service; NICE: National Institute for Health and Care Excellence; NIHR: National Institute of Health Research; PedsQL: Paediatric Quality of Life Inventory; PHR: NIHR Public Health Research programme; PSHE: Personal, Social and Health Education; QALYs: Quality-Adjusted Life-Years; QoL: quality of life; RCT: randomised controlled trial; SDQ: Strengths and Difficulties Questionnaire; SF: Short Form; SLT: Senior Leadership Team; SMT: Senior Management Team; SRD: Self-Reported Delinquency; SWEMWBS: Short Warwick-Edinburgh Mental Well-Being Scale; TSC: Trial Steering Committee; VA: value added; YPDP: Young People's Development Programme.

Competing interests

All authors declare that they have no competing interests.

Authors' contributions

RV and CB conceived the study, led funding applications, participated in the design of the study and led the drafting of the manuscript. CB designed the process evaluation. EA and DE contributed to the design of the study and wrote the analysis plan. AF contributed to the intervention development, design of the study and design of the process evaluation. RG and RLG designed the economic evaluation. AM obtained ethical permission and contributed to the writing of the manuscript. DC contributed to the design

of the study and writing of the manuscript. SS contributed to the design of the study and writing of the manuscript. MW contributed to intervention development and design of the study. All authors read and approved the final manuscript.

Acknowledgements

Funding for this trial has been provided by the UK National Institute for Health Research (NIHR) Public Health Research Programme and by the Educational Endowment Fund. The funding bodies will not be involved in the study design, in collection, analysis and interpretation of data or in the writing of the manuscript.

Author details

¹Institute of Education, 20 Bedford Way, London WC1H 0AL, UK. ²London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK. ³UCL Hospitals NHS Foundation Trust, 250 Euston Rd, London NW1 2PG, UK. ⁴Cardiff School of Social Sciences, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff CF10 3WT, UK. ⁵UCL Institute of Child Health, 30 Guilford St, London WC1N 1EH, UK. ⁶Institute of Psychiatry, Kings College London, 16 De Crespigny Park, London SE5 8AF, UK.

Received: 2 June 2014 Accepted: 9 September 2014
Published: 30 September 2014

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**(7306):191.
2. Krug EG, Mercy JA, Dahlberg LL, Zwi AB: **The world report on violence and health.** *Lancet* 2002, **360**(9339):1083–1088.
3. McKee M, Paine R: **Riots on the streets.** *BMJ* 2011, **343**:d5248.
4. Fletcher A, Gardner F, McKee M, Bonell C: **The British Government's Troubled Families Programme.** *BMJ* 2012, **344**:e3403.
5. **World report on violence and health.** 2002, Retrieved 14 February 2010 from [http://www.who.int/violence_injury_prevention/violence/world_report/en/full_en.pdf]
6. Craig WM, Henderson K, Murphy JG: **Prospective teachers' attitudes toward bullying and victimization.** *Sch Psychol Int* 2000, **21**(1):5–21.
7. Newman RS, Murray B, Lussier C: **Confrontation with aggressive peers at school: students' reluctance to seek help from the teacher.** *J Educ Psychol* 2001, **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 Pediatr Adolesc Med* 2004, **158**(8):730.
9. Radford L, Corral S, Bradley C, Fisher H, Bassett C, Howat N, Collishaw S: *Child Abuse and Neglect in the UK Today: National Society for the Prevention of Cruelty to Children.* 2011.
10. 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**(1):440.
11. Chamberlain T, Britain G: *Tellus4 National Report: Department for Children, Schools and Families.* UK: National Foundation for Educational Research; 2010. Ref: DCSF-RR218.
12. Office H: *The Offending, Crime and Justice Survey 2006.* London: HMSO; 2006.
13. Wolke D, Woods S, Stanford K, Schulz H: **Bullying and victimization of primary school children in England and Germany: prevalence and school factors.** *Br J Psychol* 2001, **92**(4):673–696.
14. Stansfeld SHM, Booy R, Taylor S, Viner RM, Head JJ, Bhui K, Hillier S, Isenwater W, Choudhry-Dormer S, Brentnall S, Klineberg E, Ahmed G: *Health of Young People in East London: The RELACHS Study 2001.* 2003.
15. Battistich VHA: **The relationship between students' sense of their school as a community and their involvement in problem behaviors.** *Am J Public Health* 1997, **87**(12):1997–2001.
16. National Centre for Educational Statistics: *Indicators of school crime and safety: 2007.* Executive summary; 2007.
17. 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**(1):130–138.
18. 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**(7206):344–348.
19. Kalliala-Heino R, Rimpelä M, Rantanen P, Rimpelä A: **Bullying at school - an indicator of adolescents at risk for mental disorders.** *J Adolesc* 2000, **23**(6):661–674.
20. Juvonen J, Graham S, Schuster MA: **Bullying among young adolescents: the strong, the weak, and the troubled.** *Pediatrics* 2003, **112**(6):1231–1237.
21. 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**(4):441–455.
22. Bond LCJ, Thomas L, Rubin K, Patton G: **Does bullying cause emotional problems? A retrospective study of young teenagers.** *BMJ* 2001, **323**:480–484.
23. Arseneault L, Bowes L, Shakoor S: **Bullying victimization in youths and mental health problems: 'much ado about nothing'.** *Psychol Med* 2010, **40**(5):717–729.
24. Van der Wal MF, De Wit CA, Hirasing RA: **Psychosocial health among young victims and offenders of direct and indirect bullying.** *Pediatrics* 2003, **111**(6):1312–1317.
25. Woods S, Wolke D: **Direct and relational bullying among primary school children and academic achievement.** *J Sch Psychol* 2004, **42**(2):135–155.
26. Glew GM, Fan M-Y, Katon W, Rivara FP, Kernic MA: **Bullying, psychosocial adjustment, and academic performance in elementary school.** *Arch Pediatr Adolesc Med* 2005, **159**(11):1026.
27. Crick NR, Ostrov JM, Werner NE: **A longitudinal study of relational aggression, physical aggression, and children's social-psychological adjustment.** *J Abnorm Child Psychol* 2006, **34**(2):127–138.
28. Price LH, Kao H-T, Burgers DE, Carpenter LL, Tyrka AR: **Telomeres and early-life stress: an overview.** *Biol Psychiatry* 2012, **73**(1):15–23.
29. 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.* Sarah Beinart ... [et al.], Communities That Care; 2002.
30. Olweus D: *Bullying at School: What we Know and What we can do.* Oxford UK and Cambridge USA: Blackwell Publishing; 1993.
31. 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**(2):99–106.
32. Brame BN, Nagin DS, Tremblay RE: **Developmental trajectories of physical aggression from school entry to late adolescence.** *J Child Psychol Psychol* 2001, **42**:503–512.
33. Boxer P, Edwards-Leeper L, 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.
34. Steer A: *Learning Behaviour: Lessons learned - A review of behaviour standards and practices in our schools.* London, UK: DCSF; 2009.
35. Sainsbury Centre for Mental Health: *Diversion: A Better way for Criminal Justice and Mental Health.* 2009.
36. Department for Education and Skills (DfES): *White Paper: The Future of Higher Education.* London: DfES; 2003.
37. Department for Children, Families and Schools (DfCFS): *Your child, your schools, our future: building a 21st century schools system.* London: TSO; 2009.
38. Department of Health (DH): *Healthy lives, brighter futures.* London: HMSO; 2009.
39. Vreeman RC, Carroll AE: **A systematic review of school-based interventions to prevent bullying.** *Arch Pediatr Adolesc Med* 2007, **161**(1):78.
40. Smith JD, Schneider BH, Smith PK, Ananiadou K: **The effectiveness of whole-school antibullying programs: a synthesis of evaluation research.** *Sch Psychol Rev* 2004, **33**(4):547–560.
41. Langford R, Bonell CP, Jones HE, Poulou T, Murphy SM, Waters E, Komro KA, Gibbs LF, Magnus D, Campbell R: **The WHO Health Promoting School framework for improving the health and well being of students and their academic achievement.** *Cochrane Database Syst Rev* 2014, **4**, CD008958.
42. Viner RM, Ozer EM, Denny S, Marmot M, Resnick M, Fatusi A, Currie C: **Adolescence and the social determinants of health.** *Lancet* 2012, **379**(9826):1641–1652.
43. Bonell CJF, Harden A, Wells H, Parry W, Fletcher A, Petticrew M, Thomas J, Whitehead M, Campbell R, Murphy S, Moore L: *Systematic Review of the*

- Effects of Schools and School Environment Interventions on Health: Evidence Mapping and Synthesis.* Southampton: National Institute of Health Research; 2013.
44. Dahlgren G, Whitehead M: *Policies and Strategies to Promote Social Equity in Health: Institute for Future Studies.* Stockholm, Sweden: 1991.
 45. Marmot M: **Creating healthier societies.** *Bull World Health Organ* 2004, **82**(5):320.
 46. 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**(2):98–99.
 47. Buck DFF: *Clustering of Unhealthy Behaviours Over Time. The King's Fund;* 2012.
 48. Flay BR, Graumlich S, Segawa E, Burns JL, Holliday MY: **Effects of 2 prevention programs on high-risk behaviors among African American youth: a randomized trial.** *Arch Pediatr Adolesc Med* 2004, **158**(4):377.
 49. Bond L, Patton G, Glover S, Carlin JB, Butler H, Thomas L, Bowes G: **The Gatehouse Project: can a multilevel school intervention affect emotional wellbeing and health risk behaviours?** *J Epidemiol Community Health* 2004, **58**(12):997–1003.
 50. Patton GC, Bond L, Carlin JB, Thomas L, Butler H, Glover S, Catalano R, Bowes G: **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**(9):1582.
 51. Hopkins B: *Just Schools: A Whole School Approach to Restorative Justice.* London and New York: Jessica Kingsley Publishers; 2004.
 52. Wales YJBEa: *National Evaluation of the Restorative Justice in Schools Programme.* London: YJD; 2004.
 53. Kane JGLG, McCluskey S, Riddell J, Stead J, Weedon E: *Restorative Practices in Scottish Schools.* Edinburgh: Scottish Executive; 2007.
 54. Skinnis LDRN, Hough M: *An Evaluation of Bristol RAIS.* London: ICPR, King's College London; 2009.
 55. Minnesota Department of Children FaL: *A Three-Year Evaluation of Alternative Approaches to Suspensions and Expulsions, Report to the Legislature.* 2002.
 56. Shaw G: **Restorative practices in Australian schools: changing relationships, changing culture.** *Conflict Resolution Quarterly* 2007, **25**(1):127–135.
 57. 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 Comp Criminol* 2011, **55**(6):846–862.
 58. Farrington DPTM: *School-Based Programs to Reduce Bullying and Victimization, Systematic Review for the Campbell Collaboration Crime and Justice Group.* 2010.
 59. Tobler AL, Komro KA, Dabroski A, Aveyard P, Markham WA: **Preventing the link between SES and high-risk behaviors: 'value-added' education, drug use and delinquency in high-risk, urban schools.** *Prev Sci* 2011, **12**(2):211–221.
 60. Hawe P, Shiell A, Riley T: **Complex interventions: how 'out of control' can a randomised controlled trial be?** *BMJ* 2004, **328**(7455):1561.
 61. Goodman R: **The strengths and difficulties questionnaire: a research note.** *J Child Psychol Psychiatry* 1997, **38**(5):581–586.
 62. Smith DJ: *School Experience and Delinquency at Ages 13 to 16.* Edinburgh, UK: The University of Edinburgh, Centre for Law and Society; 2006.
 63. Varni JW, Burwinkle TM, Seid M, Skarr D: **The PedsQL™ 4.0 as a pediatric population health measure: feasibility, reliability, and validity.** *Ambul Pediatr* 2003, **3**(6):329–341.
 64. Clarke A, Friede T, Putz R, Ashdown J, Martin S, Blake A, Adi Y, Parkinson J, Flynn P, Platt S, Stewart-Brown S: **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):1471–2458.
 65. Stephenson J, Strange V, Allen E, Copas A, Johnson A, Bonell C, Babiker A, Oakley A, Team RS: **The long-term effects of a peer-led sex education programme (RIPPLE): a cluster randomised trial in schools in England.** *PLoS Med* 2008, **5**(11):e224.
 66. Wiggins M, Bonell C, Sawtell M, Austerberry H, Burchett H, Allen E, Strange V: **Health outcomes of youth development programme in England: prospective matched comparison study.** *BMJ* 2009, **339**:b2534.
 67. Ware JEJ, Kosinski M, Keller SD: **A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity.** *Med Care* 1996, **34**(3):220–233.
 68. Maslach C, Jackson SE, Leiter MP: *Maslach Burnout Inventory;* 1986.
 69. Mytton J, DiGiuseppi C, Gough D, Taylor R, Logan S: **School-based secondary prevention programmes for preventing violence.** *Cochrane Database Syst Rev* 2006, **3**, CD004606.
 70. Hale DR, Patalay P, Fitzgerald-Yau N, Hargreaves DS, Bond L, Görzig A, Wolpert M, Stansfeld SA, Viner RM: **School-level variation in health outcomes in adolescence: analysis of three longitudinal studies in England.** *Prev Sci* 2014, **15**(4):600–610.
 71. 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**(3):157–169.
 72. Frymier AB, Shulman GM, Houser M: **The development of a learner empowerment measure 1.** *Commun Educ* 1996, **45**(3):181–199.
 73. Sawyer MG, Pleiffer S, Spence SH, Bond L, Graetz B, Kay D, Patton G, Sheffield J: **School based prevention of depression: a randomised controlled study of the beyondblue schools research initiative.** *J Child Psychol Psychiatry* 2010, **51**(2):199–209.
 74. Currie CE, Elton RA, Todd J, Platt S: **Indicators of socioeconomic status for adolescents: The WHO health behaviour in school-aged children survey.** *Health Edu Res* 1997, **12**(3):385–397.

doi:10.1186/1745-6215-15-381

Cite this article as: Bonell et al.: Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial. *Trials* 2014 15:381.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



UPDATE

Open Access



Initiating change locally in bullying and aggression through the school environment (INCLUSIVE) trial: update to cluster randomised controlled trial protocol

Chris Bonell¹, Anne Mathiot², Elizabeth Allen¹, Leonardo Bevilacqua², Deborah Christie³, Diana Elbourne¹, Adam Fletcher⁴, Richard Grieve¹, Rosa Legood¹, Stephen Scott⁵, Emily Warren¹, Meg Wiggins⁶ and Russell M. Viner^{2*}

Abstract

Background: Systematic reviews suggest that multi-component interventions are effective in reducing bullying victimisation and perpetration. We are undertaking a phase III randomised trial of the INCLUSIVE multi-component intervention. This trial aims to assess the effectiveness and cost-effectiveness of the INCLUSIVE intervention in reducing aggression and bullying victimisation in English secondary schools. This paper updates the original trial protocol published in 2014 (Trials 15:381, 2014) and presents the changes in the process evaluation protocol and the secondary outcome data collection.

Methods: The methods are summarised as follows.

Design: cluster randomised trial.

Participants: 40 state secondary schools. Outcomes assessed among the cohort of students at the end of year 7 ($n = 6667$) at baseline.

Intervention: INCLUSIVE is a multi-component school intervention including a social and emotional learning curriculum, changes to school environment (an action group comprising staff and students reviews local data on needs to review rules and policies and determine other local actions) and staff training in restorative practice. The intervention will be delivered by schools supported in the first two years by educational facilitators independent of the research team, with a third intervention year involving no external facilitation but all other elements.

Comparator: normal practice.

Outcomes:

Primary: Two primary outcomes at student level assessed at baseline and at 36 months:

1. Aggressive behaviours in school: Edinburgh Study of Youth Transitions and Crime school misbehaviour subscale (ESYTC)
2. Bullying and victimisation: Gatehouse Bullying Scale (GBS)

Secondary outcomes assessed at baseline, 24 and 36 months will include measures relating to the economic evaluation, psychosocial outcomes in students and staff and school-level truancy and exclusion rates.

(Continued on next page)

* Correspondence: r.viner@ucl.ac.uk

²UCL Great Ormond Street Institute of Child Health, 30 Guilford St, London WC1N 1EH, UK

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



© The Author(s). 2017 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

(Continued from previous page)

Sample size: 20 schools per arm will provide 90% power to identify an effect size of 0.25 SD with a 5% significance level.

Randomisation: eligible consenting schools were randomised stratified for single-sex versus mixed-sex schools, school-level deprivation and measures of school attainment.

Discussion: The trial involves independent research and intervention teams and is supervised by a Trial Steering Committee and a Data Monitoring Committee.

Trial registration: Current Controlled Trials, ISRCTN10751359. Registered on 11 March 2014.

Keywords: Bullying, Cluster randomised trial, School intervention, Violence prevention, Adolescent

Changes to the original protocol

Amendment 1

The team suggested changes to the process evaluation section of the original protocol [1]. These changes were endorsed by our Trial Steering Committee (TSC) and approved by UCL Research Ethics Committee (5/10/2015, ref 5248/001).

The deviations from the original protocol and rationales for these changes are provided in Table 1.

The main reason for changing the protocol is to limit the data collection's burden imposed on schools and re-direct the resources to in-depth data analysis and additional data collection collected from intervention schools.

Table 1 Changes to the original protocol approved in Amendment 1

	Change to the original protocol	Rationale behind the change
Staff telephone interviews	The protocol originally included conducting interviews with 1 member of the school senior leadership team (SLT) and 2 teaching staff annually (years 1–3) across 40 schools (intervention and control). These were completed as per the protocol for year 1. We do not intend to conduct staff telephone interviews in year 2. We will conduct interviews with 1 SLT member in each of the 40 schools (intervention and control) in year 3. Control schools will be interviewed in term 1, and intervention schools will be interviewed in term 3	Interviews in year 2 were considered unnecessary since we are already collecting other data (e.g. via interviews with action team members, curriculum surveys, focus groups) on how the intervention is progressing in intervention schools. Interviews in years 3 and 1 are sufficient to assess provision in control schools. Some control schools have also reported overburden following year 1 interviews, so we have reduced the number of interviews for year 3. Resources are being re-directed to in-depth case studies of intervention schools (and away from superficial data collection across all schools)
Researcher observations of curriculum delivery	We originally intended to observe $n = 1$ curriculum session in each school but are now using a curriculum survey circulated to the intervention curriculum co-ordinator in each school to assess what was delivered, how and when. Interviews with curriculum leads will also be conducted	The lead intervention facilitator advised us that observations would create an excessive administrative burden for schools, and our modified approach provides fuller data on implementation of this component
Action group meeting observations	This will be done in $n = 10$ schools per year rather than $n = 20$ schools	We are collecting substantial amounts of other data on action groups via facilitator diaries and collection of all action group documentation. The observations act as a check on the validity of diary data provided by facilitators and do not need to be done across all 20 schools each year. We will re-direct the researcher time that would have been spent on this to more in-depth data from case study schools
Case study schools	The protocol originally specified case studies in $n = 4$ control schools and $n = 4$ intervention schools. We now plan to conduct case studies in $n = 6$ intervention schools only	Control schools have complained about being overburdened with fieldwork requests, and we think that asking too much of them may threaten follow-up rates in the trial. The main purpose of the case studies is to capture data on intervention mechanisms. Case studies of control schools will not be informative about mechanisms, but will only inform us about what activities constitute the control condition in the trial, which we are already collecting across all control schools. We have re-directed resources so that we are doing more work in intervention schools ($n = 6$ schools as case study sites; conducting 1 focus group with staff, 2 focus groups with students and 2 interviews with students who were involved in restorative practices in each school)

Amendment 2

The study executive team thought it would be in the interest of the study to add a question on bullying perpetration. The change was supported by our TSC and has been approved by UCL Research Ethics Committee (23/03/2016, ref 5248/001). This added a new secondary outcome to the study and an additional question in the students' questionnaire delivered in the year 2 and year 3 follow-up surveys. The protocol has been amended accordingly in the secondary outcome section, and with a minor correction in the statistical section.

The question is taken from the Centers for Disease Control and Prevention (CDC) guidance document on bullying measures [2]. The only measure that it recommends that focuses on specific occasions of recent bullying perpetration is the Modified Aggression Scale Bullying subscale (Cronbach's alpha = 0.83) [3]. This is an existing, established measure with evidence of reliability.

Amendment 4

The current approved protocol (v1.5) had some details missing in the Process Evaluation section of the protocol. These details were in our Process Evaluation (PE) protocol, approved by our TSC, so the team thought it important to align the main protocol with the PE protocol by adding more details in the main section, new version 1.6. The amendment was approved on 10/10/2016.

The additional details added:

- Section *Trial arm fidelity*: “termly (from year 3 annual) restorative practice surveys ($n = 20$)” and “We will also draw on administrative documents (e.g. minutes, attendance sheets, training satisfaction feedback)”
- Section *Reception and responsiveness*: “We will also interview $n = 2$ students involved in restorative practice sessions per year in each case study school.”

Abbreviations

CDC: Centers for Disease Control and Prevention; DMC: Data Monitoring Committee; ESTYC: Edinburgh Study of Youth Transitions and Crime; GBS: Gatehouse Bullying Scale; TSC: Trial Steering Committee

Acknowledgements

Funding for this trial has been provided by the UK National Institute for Health Research (NIHR) Public Health Research Programme and by the Educational Endowment Fund. The funding bodies will not be involved in the study design, in collection, analysis and interpretation of data or in the writing of the manuscript.

Funding

The study has been funded by the UK National Institute for Health Research (NIHR) Public Health Research (12/153/60) and by the Educational Endowment Fund.

Authors' contributions

RV and CB conceived the study, led funding applications, participated in the design of the study and led the drafting of the manuscript. CB designed the process evaluation. EA and DE contributed to the design of

the study and wrote the analysis plan. AF contributed to the intervention development, design of the study and design of the process evaluation. RG and RLG designed the economic evaluation. AM and LB obtained ethical permission and contributed to the writing of the manuscript. DC, SS and EW contributed to the design of the study and writing of the manuscript. MW contributed to intervention development and design of the study. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The study has been approved by the Institute of Education Research Ethics Committee (18/11/13 ref. FCL 566) and the University College London Research Ethics Committee (30/1/14, Project ID: 5248/001). All participants to the surveys and the interviews are giving informed consent before each data collection activity. Further information on the consent pathways in the trial are provided in the full protocol paper [1].

Publisher's Note

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

Author details

¹London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK. ²UCL Great Ormond Street Institute of Child Health, 30 Guilford St, London WC1N 1EH, UK. ³UCL Hospitals NHS Foundation Trust, 250 Euston Rd, London NW1 2PG, UK. ⁴Cardiff School of Social Sciences, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff CF10 3WT, UK. ⁵Institute of Psychiatry, Kings College London, 16 De Crespigny Park, London SE5 8AF, UK. ⁶UCL Institute of Education, 20 Bedford Way, London WC1H 0AL, UK.

Received: 24 February 2017 Accepted: 11 May 2017

Published online: 25 May 2017

References

1. Bonell C, Allen E, Christie D, Elbourne D, Fletcher A, Grieve R, LeGood R, Mathiot A, Scott S, Wiggins M, Viner RM. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial. *Trials*. 2014;15:381.
2. Hamburger ME, Basilem KC, Vivolom AM. Measuring bullying victimization, perpetration, and bystander experiences: a compendium of assessment tools. Atlanta: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2011.
3. Bosworth K, Espelage DL, Simon TR. Factors associated with bullying behavior in middle school students. *J Early Adolesc*. 1999;19:341–62.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



Appendix 7: Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial

Articles



Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial



Chris Bonell, Elizabeth Allen, Emily Warren, Jennifer McGowan, Leonardo Bevilacqua, Farah Jamal, Rosa Legood, Meg Wiggins, Charles Opondo, Anne Mathiot, Jo Sturgess, Adam Fletcher, Zia Sadique, Diana Elbourne, Deborah Christie, Lyndal Bond, Stephen Scott, Russell M Viner

Summary

Lancet 2018; 392: 2452-64
Published Online
November 22, 2018
[http://dx.doi.org/10.1016/S0140-6736\(18\)31782-3](http://dx.doi.org/10.1016/S0140-6736(18)31782-3)
See [Comment](#) page 2416

London School of Hygiene & Tropical Medicine, London, UK (Prof C Bonell PhD, Prof E Allen PhD, E Warren MSc, R Legood DPhil, C Opondo PhD, J Sturgess MSc, Z Sadique PhD, Prof D Elbourne PhD); University College London Great Ormond St Institute of Child Health, London, UK (J McGowan PhD, L Bevilacqua MSc, A Mathiot MSc, Prof R M Viner PhD); University College London Institute of Education, London, UK (F Jamal PhD, M Wiggins MA); Cardiff University, Cardiff, UK (A Fletcher PhD); University College London Hospitals NHS Trust, London, UK (Prof D Christie PhD); Victoria University, Melbourne, VIC, Australia (Prof L Bond PhD); and Institute of Psychiatry Psychology & Neurology, London, UK (Prof S Scott FMedSci)

Correspondence to: Prof Chris Bonell, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK
chris.bonell@lshtm.ac.uk

or
Prof Russell M Viner, UCL Institute of Child Health, London WC1N 1EH, UK
r.viner@ud.ac.uk

Background Bullying, aggression, and violence among children and young people are some of the most consequential public mental health problems. We tested the Learning Together intervention, which involved students in efforts to modify their school environment using restorative practice and by developing social and emotional skills.

Methods We did a cluster randomised trial, with economic and process evaluations, of the Learning Together intervention compared with standard practice (controls) over 3 years in secondary schools in south-east England. Learning Together consisted of staff training in restorative practice; convening and facilitating a school action group; and a student social and emotional skills curriculum. Primary outcomes were self-reported experience of bullying victimisation (Gatehouse Bullying Scale; GBS) and perpetration of aggression (Edinburgh Study of Youth Transitions and Crime (ESYTC) school misbehaviour subscale) measured at 36 months. We analysed data using intention-to-treat longitudinal mixed-effects models. This trial was registered with the ISRCTN registry (10751359).

Findings We included 40 schools (20 in each group); no schools withdrew. 6667 (93.6%) of 7121 students participated at baseline and 5960 (83.3%) of 7154 at 36 months. Mean GBS bullying score at 36 months was 0.34 (SE 0.02) in the control group versus 0.29 (SE 0.02) in the intervention group, with a significant adjusted mean difference (-0.03, 95% CI -0.06 to -0.001; adjusted effect size -0.08). Mean ESYTC score at 36 months was 4.33 (SE 0.20) in the control group versus 4.04 (0.21) in the intervention group, with no evidence of a difference between groups (adjusted difference -0.13, 95% CI -0.43 to 0.18; adjusted effect size -0.03). Costs were an additional £58 per pupil in intervention schools than in control schools.

Interpretation Learning Together had small but significant effects on bullying, which could be important for public health, but no effect on aggression. Interventions to promote student health by modifying the whole-school environment are likely to be one of the most feasible and efficient ways of addressing closely related risk and health outcomes in children and young people.

Funding National Institute for Health Research, Educational Endowment Foundation.

Copyright © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Introduction

Bullying, aggression, and violence among children and young people are among the most consequential public mental health problems.^{1,2} WHO defines bullying as the intentional use of physical or psychological force against others,³ and violence as the intentional use of physical force against oneself or others.⁴ Aggression consists of hostile or destructive behaviour, and is a common part of bullying or violence. Bullying is more prevalent among British young people⁵ than in other western European countries,⁶ with cyber-bullying becoming one of the most common forms.⁷ Childhood exposure to bullying and violence results in multiple physical and mental health harms in childhood and in adult life,⁸⁻¹⁴ as well as lower educational attainment.¹⁵ Prevention of bullying and violence is therefore a major priority for public health and education systems internationally,^{2,16} with schools a

key focus of initiatives to improve young people's mental health and wellbeing.¹⁷ A challenge is to address these inter-related behaviours using single coherent interventions rather than overburdening busy schools with multiple interventions.

We developed and piloted a school-based intervention based on the three most promising approaches to reducing bullying and other health risks. The first are whole-school interventions aiming to modify overall school policies and systems rather than merely to deliver classroom-based lessons addressing bullying or other outcomes.¹⁸ A key element of many such interventions is to increase student engagement with school as a social determinant of health, particularly for the most socially disadvantaged students.^{19,20} Systematic reviews and trials suggest that such approaches reduce risk behaviours including violence and anti-social behaviour^{21,22} and

Research in context**Evidence before this study**

Reviews have shown the pervasive effect of bullying in adolescence on contemporary and later health, wellbeing, and social functioning. Systematic reviews indicate that whole-school interventions are among the most promising approaches to the promotion of young people's health, and that these are effective in reducing bullying victimisation. Restorative practice is increasingly used in schools to address bullying and antisocial behaviour. We undertook a systematic review in January, 2018, of PubMed using the search terms (((("Schools"[Mesh]) AND "Randomized Controlled Trial" [Publication Type]) AND ("Bullying"[Majr]) OR "Aggression"[Majr])) AND restorative justice. We identified no published randomised trials or systematic reviews of restorative practice interventions in schools.

Added value of this study

We present the first evidence from a randomised trial that a whole-school intervention including restorative practice and

social and emotional learning elements, has positive effects on bullying; mental health and wellbeing; quality of life; smoking, alcohol and drug use; and police contact.

The Learning Together intervention is very low cost compared with other educational interventions and offers a coherent means of addressing clustered risks and health outcomes in schools.

Implications of all the available evidence

Interventions aiming to promote student health by modifying the whole-school environment can have effects of public health importance across a broad range of important outcomes in young people. The inclusion of restorative practice within such interventions can reduce bullying among all young people and reduce aggressive behaviour in those with high baseline aggression.

bullying victimisation.¹⁸ The large SEHER trial²³ in Bihar, India, showed that such interventions can be effective in resource-poor settings.

The second promising approach is based on restorative practice, which aims to prevent or resolve conflicts between students or between staff and students to prevent further harms.²⁴ It enables victims to communicate to perpetrators the effects of the harm, and for perpetrators to acknowledge and amend their behaviour to avoid further harms. Restorative practice can involve primary prevention of incidents (such as so called circle-time, in which students are brought together with their teacher to discuss their feelings, identify problems, and maintain good relationships) or secondary prevention to resolve incidents (such as conferencing, bringing together parties to a conflict and, when necessary, external agencies, to reflect on more serious incidents and develop strategies to avoid future harms). Restorative practice is increasingly used within schools in the UK and internationally to address bullying and antisocial behaviour, with encouraging results from non-randomised evaluations.^{16,25} However, there have been no randomised trials of restorative practice in schools.²⁶

The third is social and emotional education. Lessons to teach young people the skills needed to manage their emotions and relationships can enhance social relationships, improve mental health, and reduce bullying.²⁷

In 2014, we developed the Learning Together intervention,²⁸ which aimed to modify the school environment by using all three of these approaches to reduce bullying and aggression, and promote student health and wellbeing across various domains. A pilot trial²⁸ in eight schools showed that the intervention was feasible

and acceptable to participants. We then did the INCLUSIVE trial, a cluster-randomised controlled trial of Learning Together. We hypothesised that secondary schools using the intervention would have lower rates of self-reported bullying and perpetration of aggression, and improved student and staff secondary outcomes compared with control schools, and that Learning Together would be cost-effective compared with standard school practice. Here, we report student health and behaviour outcomes. Data on student educational outcomes and staff outcomes will be published when routine administrative data become available in 2019.

Methods**Study design and participants**

We did a cluster-randomised controlled trial, along with process evaluation (an assessment of implementation, context, reach, and acceptability) and an economic assessment, in 40 secondary schools in southeast England between 2014 and 2017, with schools as the unit of allocation.²⁹ We included all students in the school at the end of year 7 (age 11–12 years) at baseline, with follow-up at 24 months and 36 months (end of year 10; age 14–15 years). There were no ineligibility criteria for students.

We enrolled mainstream secondary schools within the state education system that had a most recent school quality rating by the Ofsted (the national education inspectorate in the UK) of: "requires improvement", "satisfactory", "good", or "outstanding". We excluded schools with an "inadequate/poor" rating because such schools are subject to special measures, which were likely to impede delivery of the intervention. We identified and contacted all eligible schools in Greater London and surrounding counties between March, and June, 2014.

The protocol was amended during the trial to refine the methods. All amendments were approved by the independent study steering committee. The only change to trial outcomes was adding a measure of bullying perpetration as a secondary outcome. All refinements were completed before collection of the 36-month surveys and before trial analyses.

The trial was approved by the University College London ethics committee (ref 5248/001). Written, informed consent was obtained from head teachers for random allocation and intervention, and from individual students, staff, and intervention facilitators for data collection. Information sheets and consent forms for student surveys were identical in intervention and control schools and did not refer to the intervention. Parents were informed about the study and could withdraw their children from research activities.

See Online for appendix

Randomisation and masking

We randomly allocated schools (1:1) to the intervention group (Learning Together) or the control group (standard practice) immediately after baseline surveys. We stratified randomisation by key school-level determinants of violence,³⁰ with data obtained from the Department for Education:³¹ (1) single sex versus mixed sex school; (2) school-level deprivation, as measured by the percentage of students eligible for free school meals (low or moderate 0–23%; high >23%; 23% is the 75th centile for England); and (3) student attainment in General Certificate of Secondary Education examinations normally sat by students aged 16 years, with a total score based on the best eight grades achieved by each student accounting for previous attainment at age 11 years (above and below median score of 1000 across English schools), a school-level measure of students' attainment.

Sequence allocation was generated by the Clinical Trials Unit at The London School of Hygiene & Tropical Medicine using Stata's `ralloc` command, and was concealed from schools and the wider evaluation and intervention teams. Allocation was communicated to the research team who then communicated it to schools and the intervention team.

Schools, the intervention team, and process and economic evaluators could not be masked to allocation status. However, fieldwork staff were masked to allocation as was the outcome research team lead (RMV), and staff who entered and analysed data.

Procedures

Baseline surveys were done between March and July, 2014; 24-month follow-up surveys were done in April to June, 2016; and 36-month follow-up surveys were done in April to June, 2017. Student self-reported data were collected using paper questionnaires, which students completed in classrooms under examination conditions facilitated by trained researchers with teachers present but unable to read student responses. Questionnaires were

double-entered by trained personnel. Questionnaires with additional text, regardless of content, were scanned and password-protected scans were sent to the study team to assess serious adverse events and abuse requiring safeguarding interventions. Password-protected electronic data were securely transferred to the London School of Hygiene and Tropical Medicine and stored on secure servers.

Informed by prior theory,³² the intervention aimed to enable young people to choose healthier behaviours by promoting their autonomy, motivation, and reasoning ability. These were to be promoted by increasing engagement with school via improving relationships between and among students and teachers, and between academic education and broader student development, as well as by reorienting school practices and organisation to centre on student needs (appendix p 4).

In the first year, all school staff were trained in restorative practices with in-depth training for selected staff from accredited providers over 3 days. Schools were provided with a manual to guide action group meetings of at least six staff and six students, held twice per term, to revise relevant school policies and coordinate the intervention. These groups ensured that local implementation of the intervention was appropriate for students, with scope for some locally decided actions. For the first 2 years, the groups who attended action group meetings were encouraged to discuss and take action by an external facilitator with school management experience. Schools were sent a report on local needs, derived from the student surveys, to inform decisions. They were also provided with lesson plans and slides to guide teachers' delivery of 5–10 h per year of lessons on social and emotional skills for students in years 8–10 (age 12–15 years). School staff delivered primary restorative practices using respectful language to challenge or support behaviour and circle time to build relationships, and secondary restorative practices involved some staff implementing restorative conferences to address more serious behaviour problems.

Schools randomised to the control group continued with their normal practices and received no additional input. The sample of schools happened to be spread over a wide geographical area and there were no intervention and control schools in close proximity. Head teachers and a few staff were aware that the school was participating in the INCLUSIVE trial but were not informed of the name or detailed contents of the intervention.

In line with the UK Medical Research Council guidance on complex interventions,³³ we did a process evaluation assessing trial context and trial group fidelity in all schools. For trial context, we examined services and practices relating to bullying, discipline, and social and emotional skills education, and student participation in school policy in control schools to assess how these differed from the intervention. This assessment drew on interviews with one member of each control school's senior leadership team and two other members of staff

done in the first year of intervention, and interviews with a senior leadership team member in the third year.

We scored fidelity to the intervention in the first 2 years out of eight points for each school, assessing whether: at least five staff attended in-depth training; six action-group meetings occurred per year; policies and rules were reviewed; locally decided actions were implemented; members assessed that action groups had a good or very good range of members; members assessed that action groups were well or very well led; schools delivered at least 5 h or two modules each year; at least 85% of staff reported that if there was trouble at the school, staff responded by talking to those involved to help them get on better. We assessed fidelity in the third year using a narrower range of data because the research teams had less access to schools. Schools were scored out of four on whether: six action groups were convened; local decisions were implemented; schools delivered at least 5 h or at least two modules; and at least 85% of staff reported that if there was trouble at the school, staff responded by talking to those involved to help them get on better. The appendix (p 16–29) provides additional data on the process evaluation.

Process evaluation interviews were done annually in each intervention school with two action group members. They used purposive sampling to involve participants with diversity in terms of characteristics thought important for exploring implementation.³³ Six intervention schools were chosen (encompassing a range of percentages of students entitled to free school meals, types of state school, and facilitator and school responsiveness to intervention activities) as case studies for more in-depth process evaluation involving two focus groups with students and one with staff each year. The six were selected to encompass variation by percentage of students entitled to free school meals, type of state school, facilitator and school responsiveness to intervention activities rated by facilitators after three months.

For the economic evaluation, we used a cost-consequence analysis including all main outcomes and evaluated incremental effects at 24 months and 36 months since randomisation. Costs included use of education, police, and NHS resources (appendix p 31).³⁴ We collected data on the costs of delivering the intervention from the invoices for facilitators and trainers and data from the process evaluation on school staff time requirements. We determined the costs of staff time taken to deal with bullying through the staff survey questionnaire, and the costs of NHS and police resource data through the student survey questionnaires and valued them accordingly.

We defined serious adverse events as (1) any death, serious injury, or hospital admission in any student in a trial school that was reported to investigators; or (2) responses on study questionnaires that prompted significant concerns about mental health, sexual risk, or child safety, which were then communicated to the school.

Outcomes

The primary outcomes were self-reported experience of bullying victimisation and perpetration of aggression measured at 36 months. Outcome data were collected by a research team (led by RMV), independent of the intervention team (lead by CB).

We measured bullying victimisation using the Gatehouse Bullying Scale (GBS), a 12-item validated³⁵ self-reported measure of being subject to teasing, name-calling, rumours, being left out of things, and physical threats or actual violence from other students, including face-to-face and cyber-bullying, within the past 3 months. Students reported the frequency and upset related to each experience. Items are summed to make a total bullying score (higher represents more frequent, upsetting bullying).

We measured perpetration of aggressive behaviour using the Edinburgh Study of Youth Transitions and Crime (ESYTC) school misbehaviour subscale,³⁶ a 13-item scale measuring self-reported aggression towards students and teachers. Each item was coded from hardly ever or never; less than once a week; at least once a week; to most days. Items are summed to provide a total score; high scores indicate greater aggressive behaviour.

The secondary outcomes included GBS and ESYTC scores at 24 months. The other secondary outcomes assessed at 36 months were quality of life measured with the Paediatric Quality of Life Inventory,³⁷ version 4.0 (higher scores indicate better quality of life); wellbeing, measured with the validated Short Warwick-Edinburgh Mental Well-Being Scale³⁸ (higher scores indicate greater emotional wellbeing); psychological problems, measured with the Strengths and Difficulties Questionnaire³⁹ (a brief, validated instrument for detecting behavioural, emotional, and peer problems in children and adolescents; a higher score indicates greater problems); bullying perpetration, measured with the Modified Aggression Scale, Bullying Subscale⁴⁰ (used at follow-up only; higher scores indicated greater bullying); substance use, assessed using validated age-appropriate questions about cigarette smoking, alcohol use, and illicit drug use taken from national surveys;⁴¹ sexual risk behaviour⁴² (age of sexual debut and use of contraception at first sex; assessed only at follow-up); use of NHS health services (self-reported use of primary care, accident and emergency, or other service in the past 12 months); and contact with police (self-report of being stopped, reprimanded, or picked up by the police in the past 12 months).

Statistical analysis

We calculated that, using a conservative intraclass correlation coefficient⁴³ of 0.04 and an estimate of 150 students per school, a trial involving 20 schools per group would provide 90% power to identify an effect size of 0.25 SD with a 5% significance level. This difference is considered to represent a moderate size of effect and is

	Control group	Intervention group
School characteristics		
Number of schools	20	20
School sex mix		
Mixed	15 (75%)	15 (75%)
Girls only	3 (15%)	4 (20%)
Boys only	2 (10%)	1 (5%)
School type		
Voluntary	1 (5%)	3 (15%)
Community school	3 (15%)	2 (10%)
Academy (converter mainstream)	9 (45%)	10 (50%)
Academy (sponsor led)	3 (15%)	3 (15%)
Foundation school	4 (20%)	2 (10%)
Ofsted rating*		
Excellent	5 (25%)	6 (30%)
Good	13 (65%)	12 (60%)
Requires improvement	2 (10%)	2 (10%)
Mean value added score	1003 (24.8)	1004 (20.4)
Mean proportion of students on free school mean (%)	36 (18.0)	35 (22.0)
Mean school size	1122 (322.7)	1046 (323.3)
Mean IDAC1 score	0.26 (0.2)	0.24 (0.2)
Student characteristics		
Number of students	3347†	3320†
Mean age (years)	12 (0.4)	12 (0.4)
Sex		
Male	1639 (49.9%)	1464 (44.9%)
Female	1649 (50.2%)	1804 (55.2%)
Ethnicity		
White British	1391 (41.5%)	1221 (37.3%)
White other	291 (8.8%)	273 (8.3%)
Asian or Asian British	859 (25.9%)	786 (24.0%)
Black or Black British	384 (11.6%)	535 (16.4%)
Chinese or Chinese British	11 (0.3%)	35 (1.1%)
Mixed ethnicity	238 (7.2%)	224 (6.9%)
Other	140 (4.2%)	198 (6.1%)

(Table 1 continues in next column)

	Control group	Intervention group
(Continued from previous column)		
Religion		
None	983 (29.6%)	787 (24.0%)
Christian	1073 (32.3%)	1173 (35.8%)
Jewish	9 (0.3%)	13 (0.4%)
Muslim	878 (26.5%)	817 (24.9%)
Hindu	90 (2.7%)	176 (5.4%)
Sikh	71 (2.1%)	88 (2.7%)
Don't know	145 (4.4%)	126 (3.8%)
Other	73 (2.2%)	100 (3.1%)
Family structure		
Two parents	2393 (71.9%)	2369 (72.1%)
Single mothers	604 (18.2%)	626 (19.0%)
Single fathers	37 (1.1%)	56 (1.7%)
Reconstituted	246 (7.4%)	204 (6.2%)
Other	48 (1.4%)	33 (1.0%)
At least one parent in work		
No	298 (8.7%)	233 (7.2%)
Yes	2437 (74.0%)	2381 (73.4%)
Don't know	566 (11.2%)	632 (19.5%)
Housing tenure		
Renting from council or housing association	474 (14.4%)	559 (17.3%)
Renting from a landlord	391 (11.9%)	396 (12.2%)
Owned by family	1451 (44.1%)	1273 (39.3%)
Other	62 (1.9%)	59 (1.8%)
Don't know	912 (27.7%)	951 (29.4%)
Mean family affluence scale	6 (1.8)	6 (1.8)

Data are n (%) or mean (SD) unless otherwise stated. *One control school did not have an Ofsted rating. †The number of students who responded at this survey; actual number of responses to each question varies, but item non-response was similar in each group.

Table 1: Baseline characteristics

in line with effect sizes in previous studies.⁴⁴ We therefore planned to include roughly 6000 students.

The primary analysis of outcomes was by intention to treat, including all randomly assigned schools and students. We analysed each measure using a separate mixed model with the outcomes from each timepoint treated as a repeated measures. Fixed effects of group (intervention vs control), time (baseline, 24 months, 36 months), and the interaction between treatment and time were specified, and the estimated baseline measures were constrained to be identical in the two groups of the trial. This approach is equivalent to adjusting for baseline and permitting the relationship between baseline and follow-up scores to differ at each timepoint, but offers the additional advantage that the data from all participants contribute to the analysis, even when there were missing

data at follow-up. Details of missing data are shown in the appendix (p 35). We used random effects for school and participants to allow for correlations within schools and repeated measures within participants. Statistical significance for these analyses was taken at the 5% level ($p < 0.05$). We did analyses adjusted only for baseline measures of the outcomes and in the primary analysis adjusted for baseline measures of outcomes, sex, ethnicity, socioeconomic status, and school-level stratification factors.

For the two primary outcomes, we used mixed linear regression models with random effects at the participant and school levels to estimate the mean difference in GBS and ESYTC scores between the two arms of the trial. We restricted formal testing to the prespecified of secondary outcomes, and we used appropriate multilevel models to examine the effect of the intervention. For continuous outcomes, we calculated unadjusted and adjusted mean differences with 95% CIs and adjusted effect sizes (standardised mean difference). For binary and ordinal outcomes, we calculated unadjusted and adjusted odds

ratios. Additionally, we also calculated adjusted risk differences for binary outcomes. We assessed differential effects of the intervention on the primary and secondary outcomes by subgroup using likelihood ratio tests for the treatment by subgroup interaction terms. We estimated the effects in the different subgroups directly from the regression model with the interaction term included.

We did four subgroup analyses: (1) by sex; (2) by socioeconomic status, measured using the Health Behaviours in School-aged Children Family Affluence Scale (<5 for low vs ≥6 for high);⁴⁵ (3) baseline bullying experience based on GBS (high, defined as at least weekly experience of bullying or being upset by it vs medium or low, defined as less than weekly experience of bullying and not being upset by it); and (4) baseline behaviour problems based on the ESYTC (>0 for high vs 0 for low).

For the process evaluation of trial context and intervention fidelity, we used thematic content analysis for qualitative data and descriptive statistics for quantitative data. For the economic analysis we used general linear mixed regression models that allowed for clustering of students within schools, and including school as a random effect variable.

The trial is registered with ISRCTN (ISRCTN10751359).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

6667 (93.6%) of 7121 registered students in the 40 participating schools provided data at baseline (3320 [94.4%] of 3516 in the intervention group vs 3347 [92.8%] of 3605 in the control group). Table 1 shows the baseline characteristics of schools and students, the characteristics of students at 24 months and 36 months are shown in the appendix (pp 6–7). The 40 participating schools did not differ significantly from 450 non-recruited schools in terms of size, population, deprivation, or gross or value-added attainment, but participating schools were more likely to have an Ofsted rating of good or outstanding (appendix p 4). All schools participated in the follow-up surveys at 24 months and 36 months; the numbers of students who completed the questionnaires at baseline, 24 months, and 36 months were similar in each group (figure). Student and school characteristics and outcomes at baseline were well balanced across arms. Primary and secondary outcomes at baseline are shown in the appendix (pp 8–9).

Mean GBS bullying score at 36 months was 0.34 (SE 0.02) in the control group versus 0.29 (SE 0.02) in the intervention group, with a significant adjusted mean difference (−0.03, 95% CI −0.06 to −0.001; adjusted effect size −0.08; table 2). Mean ESYTC score at

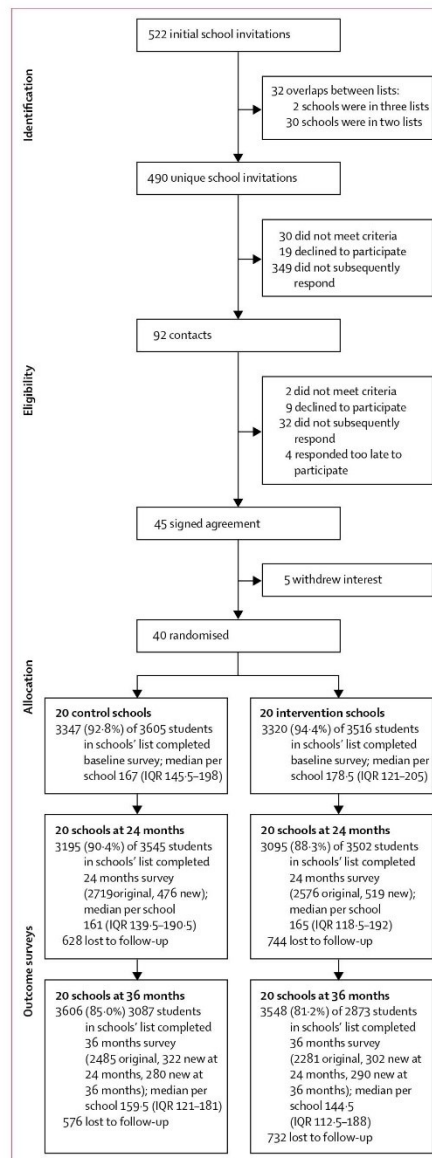


Figure: Trial profile

36 months was 4.33 (SE 0.20) in the control group versus 4.04 (0.21) in the intervention group, with no evidence of a difference between groups (adjusted difference -0.13, 95% CI -0.43 to 0.18; adjusted effect size -0.03).

With regards to the secondary outcomes, the GBS overall score and the ESYTC scores at 24 months were higher in the intervention groups than in the control groups, but we found no evidence of a significant difference (table 3). At 36 months, students in intervention schools had a higher

quality of life and psychological wellbeing and lower psychological difficulties than did students in control schools (table 3). There was also evidence that those in intervention schools had lower emotional, conduct, hyperactivity, and peer problems (table 3).

Students in intervention schools also had lower odds of having ever smoked regularly, lower odds of having ever drunk alcohol, and lower odds of having ever been offered or tried illicit drugs (table 4). Among students in

	Control group (3087 students)*	Intervention group (2281 students)*	Unadjusted difference (95% CI)	p value	Adjusted difference (95% CI)	p value	Adjusted effect size
GBS overall score	0.34 (0.02)	0.29 (0.02)	-0.03 (-0.06 to -0.002)	0.0395	-0.03 (-0.06 to -0.001)	0.0441	-0.08
Teasing	0.55 (0.03)	0.47 (0.03)	-0.04 (-0.09 to 0.01)	..	-0.05 (-0.10 to 0.000)	..	-0.07
Rumours	0.37 (0.02)	0.31 (0.02)	-0.06 (-0.10 to -0.02)	..	-0.07 (-0.11 to -0.02)	..	-0.10
Deliberate exclusion	0.24 (0.01)	0.22 (0.02)	-0.04 (-0.08 to -0.004)	..	-0.04 (-0.08 to 0.01)	..	-0.06
Threatened or hurt	0.21 (0.02)	0.18 (0.02)	0.01 (-0.02 to 0.05)	..	0.01 (-0.03 to 0.05)	..	0.02
ESYTC overall score	4.33 (0.20)	4.04 (0.21)	-0.07 (-0.38 to 0.25)	0.6820	-0.13 (-0.43 to 0.18)	0.4199	-0.03

Data are mean (SE) unless otherwise stated. GBS=Gatehouse Bullying Scale. ESYTC=Edinburgh Study of Youth Transitions and Crime. *Shows the number of students who responded at this survey; actual number of responses to each question varied, but non-response for each item was similar across arms.

Table 2: Primary outcomes at 36 months

	Control group (mean, SE)	Intervention group (mean, SE)	Unadjusted difference (95% CI)	p value	Adjusted difference (95% CI)	p value	Adjusted effect size
24 months	3195 students*	3095 students*
GBS overall score	0.42 (0.02)	0.37 (0.02)	-0.02 (-0.05 to 0.01)	0.2198	-0.02 (-0.05 to 0.01)	0.1581	-0.05
Teasing	0.66 (0.03)	0.59 (0.03)	-0.02 (-0.07 to 0.03)	..	-0.03 (-0.08 to 0.01)	..	-0.05
Rumours	0.44 (0.02)	0.41 (0.02)	-0.02 (-0.06 to 0.02)	..	-0.02 (-0.06 to 0.02)	..	-0.04
Deliberate exclusion	0.31 (0.02)	0.30 (0.02)	-0.03 (-0.07 to 0.01)	..	-0.03 (-0.07 to 0.01)	..	-0.05
Threatened or hurt	0.26 (0.02)	0.22 (0.02)	0.01 (-0.02 to 0.05)	..	0.01 (-0.03 to 0.04)	..	0.01
ESYTC overall score	4.24 (0.28)	3.96 (0.28)	-0.04 (-0.34 to 0.27)	0.8113	-0.06 (-0.35 to 0.24)	0.7206	-0.01
36 months	3087 students*	2281 students*
PedsQL overall score	78.82 (0.54)	80.65 (0.55)	1.16 (0.41 to 1.90)	0.0022	1.44 (0.70 to 2.17)	0.0001	0.14
Physical health	84.46 (0.61)	85.84 (0.63)	0.80 (-0.02 to 1.62)	..	1.05 (0.26 to 1.84)	..	0.09
Psychosocial health	75.75 (0.56)	77.87 (0.58)	1.46 (0.62 to 2.29)	..	1.71 (0.87 to 2.54)	..	0.15
Emotional functioning	71.41 (1.04)	73.38 (1.07)	1.52 (0.30 to 2.75)	..	2.08 (0.88 to 3.28)	..	0.12
Social functioning	87.16 (0.41)	88.57 (0.43)	0.98 (0.12 to 1.85)	..	1.04 (0.17 to 1.91)	..	0.08
School functioning	68.72 (0.70)	71.65 (0.72)	1.75 (0.72 to 2.79)	..	1.87 (0.83 to 2.91)	..	0.13
SDQ total difficulties score†	12.20 (0.18)	11.51 (0.19)	-0.51 (-0.80 to -0.22)	0.0005	-0.54 (-0.83 to -0.25)	0.0002	-0.14
Emotional problems	3.68 (0.12)	3.57 (0.12)	-0.13 (-0.25 to 0.001)	..	-0.14 (-0.26 to -0.02)	..	-0.08
Conduct problems	2.14 (0.05)	1.90 (0.05)	-0.14 (-0.24 to -0.05)	..	-0.17 (-0.26 to -0.07)	..	-0.13
Hyperactivity	4.48 (0.09)	4.27 (0.09)	-0.11 (-0.23 to 0.01)	..	-0.14 (-0.26 to -0.02)	..	-0.08
Peer problems	1.92 (0.04)	1.78 (0.05)	-0.11 (-0.21 to -0.02)	..	-0.10 (-0.20 to -0.02)	..	-0.08
Pro-social strengths	6.91 (0.10)	7.09 (0.10)	0.07 (-0.04 to 0.18)	..	0.08 (-0.02 to 0.19)	..	0.06
SWEMWBS total wellbeing index	22.88 (0.19)	23.32 (0.19)	0.27 (-0.06 to 0.60)	0.1150	0.33 (0.00 to 0.66)	0.0487	0.07
Age of sexual debut	13.11 (0.43)	12.54 (0.49)	-0.58 (-1.97 to 0.81)	0.4155	-0.35 (-1.48 to 0.78)	0.5409	-0.12
Modified aggression scale, bullying subscale	2.75 (0.21)	2.33 (0.21)	-0.28 (-0.84 to 0.29)	0.3333	-0.26 (-0.57 to 0.05)	0.0976	-0.12

GBS=Gatehouse Bullying Scale. ESYTC=Edinburgh Study of Youth Transitions and Crime. PedsQL=Paediatric Quality of Life Inventory. SDQ=Strengths and Difficulties Questionnaire. SWEMWBS=Short Warwick Edinburgh Mental Wellbeing Scale. *Shows the number of students who responded at this survey; actual number of responses to each question varied, but non-response for each item was similar across arms. †Does not include the pro-social score subscale.

Table 3: Continuous secondary outcomes

	Control group (3087 students)*	Intervention group (2281 students)*	Unadjusted odds ratio (95% CI)	p value	Adjusted odds ratio (95% CI)	p value	Adjusted risk difference (95% CI)	p value
Ever smoked regularly								
No	2293 (77.70%)	2318 (84.17%)	1.00	0.0011	1.00	0.0009	..	0.0006
Yes	658 (22.30%)	436 (15.83%)	0.59 (0.43 to 0.81)	..	0.58 (0.43 to 0.80)	..	-0.03 (-0.05 to -0.01)	..
If yes, how long since last smoked	1.46 (1.06 to 2.01)†	..	1.40 (1.02 to 1.93)†
<1 day	105 (16.20%)	49 (11.56%)
1-3 days	61 (9.41%)	26 (6.13%)
4-7 days	37 (5.71%)	24 (5.66%)
1 week-1 month	85 (13.12%)	57 (13.44%)
1-2 months	63 (9.72%)	50 (11.79%)
3-6 months	87 (13.43%)	69 (16.27%)
>6 months	210 (32.41%)	149 (35.14%)
Ever drunk alcohol?								
No	1677 (56.43%)	1735 (62.43%)	1.00	0.0290	1.00	0.0094	..	0.0082
Yes	1295 (43.57%)	1044 (37.57%)	0.75 (0.58 to 0.79)	..	0.72 (0.56 to 0.92)	..	-0.03 (-0.06 to -0.01)	..
If yes, had alcohol in the past week	1.00	..	1.00
No	949 (75.80%)	800 (80.00%)	1.00	..	1.00
Yes	303 (24.20%)	200 (20.00%)	0.71 (0.52 to 0.98)	..	0.67 (0.50 to 0.91)
Number of times really drunk	0.57 (0.33 to 0.98)†	0.0426	0.51 (0.33 to 0.80)†	0.0029
Never	788 (53.14%)	721 (61.21%)
Once	283 (19.08%)	178 (15.11%)
2-3 times	221 (14.90%)	144 (12.22%)
4-10 times	124 (8.36%)	67 (5.69%)
>10 times	67 (4.52%)	68 (5.77%)
Binge drinking (≥5 drinks in a row) in past 30 days	0.78 (0.53 to 1.14)†	0.2071	0.77 (0.59 to 1.00)†	0.0521
0	1162 (73.97%)	974 (76.57%)
1-2	276 (17.57%)	209 (16.43%)
3-5	69 (4.39%)	45 (3.54%)
6-9	30 (1.91%)	13 (1.02%)
≥10	34 (2.16%)	31 (2.44%)
Ever been offered illicit drugs	0.52 (0.34 to 0.79)†	0.0023	0.51 (0.36 to 0.73)†	0.0003
No	1913 (64.41%)	1997 (72.54%)
Yes, but did not try them	744 (25.05%)	567 (20.60%)
Yes, and tried them	313 (10.54%)	189 (6.87%)
Used any contraception at first sex	1.00	0.6583	1.00	0.8410	..	0.8395
No	64 (23.10%)	36 (21.95%)
Yes	213 (76.90%)	128 (78.05%)	1.18 (0.56 to 2.48)	..	1.08 (0.50 to 2.35)	..	0.01 (-0.08 to 0.10)	..
Use of NHS services in past 12 months	1.00	0.6392	1.00	0.5652	..	0.5647
No	1605 (53.22%)	1472 (52.59%)
Yes	1411 (46.78%)	1327 (47.41%)	0.96 (0.83 to 1.12)	..	0.96 (0.82 to 1.11)	..	-0.01 (-0.04 to 0.02)	..
Contact with police in past 12 months	1.00	0.0403	1.00	0.0269	..	0.0222
No	2626 (86.52%)	2485 (88.43%)
Yes	409 (13.48%)	325 (11.57%)	0.75 (0.57 to 0.99)	..	0.74 (0.56 to 0.97)	..	-0.02 (-0.04 to -0.003)	..

All assessed at 36 months. Data are n (%), unless stated otherwise. NHS=National Health Service. *The number of students who responded at this survey, actual number of responses to each question varies, but item non-response is similar across arms. †Proportional odds ratio.

Table 4: Categorical secondary outcomes

the intervention group who had ever smoked, there was evidence that the time since the last cigarette was longer than in those in the control group and that, among those who had ever drunk alcohol, there were lower odds of having drunk in the past week, and number of

times having been really drunk (table 4). Students in intervention schools had lower odds of having ever been in contact with police in the past 12 months than did those in control schools (table 4). We found no evidence of differences in age of sexual debut or use of

contraception at first sex, bullying perpetration, or use of NHS services. The appendix shows results for secondary outcomes at 24 months not selected for formal testing (pp 10–13).

Subgroup analyses suggest that the intervention had a greater effect in boys than in girls for many secondary outcomes (quality of life, psychological problems, wellbeing, having ever smoked regularly, having ever drunk alcohol, bullying perpetration, and contact with police; table 5). The intervention was also more effective in students with higher baseline bullying experience, with greater effects on bullying and psychological problems, quality of life, and wellbeing (table 5). The intervention was more effective in those with greater baseline aggression, with greater effects on both primary outcomes (bullying victimisation and aggressive behaviour), psychological secondary outcomes (quality of life, psychosocial problems, and wellbeing) and some risk behaviours (ever smoked regularly, ever drunk alcohol; table 5). There was no suggestion of any difference in the outcomes by socioeconomic status (appendix pp 14–15).

Other process evaluation findings will be reported elsewhere. Fidelity to the intervention varied between schools and over time, with a reduction in the fidelity of formal intervention activities in the third year. The median fidelity score for the first and second years was 6 out of 8 (IQR 5–7), whereas for the third year the median was 1 out of 4 (IQR 0–3). In the third year, 15 schools sustained restorative practice. Interviews with action group members and focus groups with staff in case-study schools suggested that in the third year, schools commonly incorporated what they regarded as the most useful action group functions into mainstream school structures and processes. Training, action groups, and restorative practices but not the curriculum were delivered with good fidelity (appendix pp 23–24). Increased fidelity of delivery in the first 2 years of the intervention was associated with lower bullying victimisation at 24 months but not with lower aggression (appendix p 17). The fidelity score in the third year was not associated with either primary outcome (appendix p 17).

Slightly over half of staff in intervention schools were aware that the school had been taking steps to reduce bullying and aggression, falling slightly between the second and third years (appendix pp 18–19). About a third of students reported being aware that the school had been taking steps to reduce bullying (appendix pp 20–21). About half reported that if there was trouble at school, staff responded by talking to those involved to help them get on better. About two-thirds of students reported that teachers and students got together to build better relationships or discuss their views and feelings. Other data on the process evaluation are shown in the appendix (pp 22–29).

Many schools in the control group implemented similar activities to those prescribed in the intervention but with

variable degrees and quality. Five control schools used restorative practice, social and emotional skills education, and consultation with students on policy. A per protocol analysis excluding these schools showed no discernible differences in the intervention effects compared with the intention-to-treat analyses (appendix p 30).

The main time-consuming activities for school staff were attending the training and curriculum delivery. We included staff training in the intervention costs but staff interviews suggested that training was not additional but part of existing training periods, suggesting that the intervention costs might be overestimated. Mean total costs to the education sector to address bullying were £116 per pupil (SD 47) in the control group compared to £163 per pupil (SD 69) in the intervention group over the first 2 years, and £63 (33) versus £74 (37) in the third year. For the intervention schools, the mean cost of facilitators and trainers was £11039 per school (SD 993). The mean cost to address bullying per school of all staff time combined was £232 670 (SD 113 634) for the intervention group and £202 405 (SD 103 090) for the control group. Costs for health-service use and police contacts were similar in both groups (appendix p 34). Overall, the intervention increased costs and reduced bullying, leading to incremental costs averted of £2352 at 36 months. Further details of resource use and costs are reported in the appendix (31–34).

The number of reported serious events was similar in each group although patterns differed (table 6). Two each of suicide and stabbing incidents were reported by intervention group schools, which could reflect increased reporting in intervention schools.

Discussion

We report results of the first randomised controlled trial of restorative approaches to reduce bullying and aggression and promote student health, using a whole-school approach, engaging students in school decision making, and providing social and emotional skills education. Learning Together reduced student reports of bullying victimisation compared with schools continuing their standard practice. We did not identify a reduction in overall student reports of aggression. Learning Together seemed to have larger benefits for many secondary outcomes, from improved psychological function, wellbeing, and quality of life, to reductions in police contact, smoking, and alcohol and drug use. The effects on bullying and other continuous outcomes by the third year approximated 0.1 SD, which could be important at the population level. We found intervention effects both in the whole sample and in schools with higher levels of bullying or aggression at baseline, implying that the intervention worked to curtail existing bullying and aggression (secondary prevention) as well as prevent new bullying (primary prevention). We also found that the Learning Together intervention had greater effects for boys than in girls for secondary psychological and

	Main adjusted effect		By sex		P _{interaction}	By baseline bullying		P _{interaction}	By baseline aggression		P _{interaction}
	Effect size*	p value	Boys (95% CI)	Girls (95% CI)		Low (95% CI)	High (95% CI)		Low (95% CI)	High (95% CI)	
Continuous outcomes											
GBS overall score	-0.03 (-0.06 to -0.001)	0.0441	-0.04 (-0.08 to 0.001)	-0.03 (-0.06 to 0.01)	0.6113	0.15 (0.12 to 0.18)	-0.41 (-0.45 to -0.36)	<0.0001	0.01 (-0.03 to 0.05)	-0.06 (-0.10 to -0.02)	0.0024
ESYTC overall score	-0.13 (-0.43 to 0.18)	0.4199	-0.33 (-0.73 to 0.06)	0.04 (-0.32 to 0.39)	0.0890	0.02 (-0.34 to 0.37)	-0.17 (-0.63 to 0.29)	0.4422	0.71 (0.33 to 1.10)	-0.65 (-1.03 to -0.27)	<0.0001
PedsQL overall score	1.44 (0.70 to 2.17)	0.0001	3.85 (2.89 to 4.80)	-0.41 (-1.28 to 0.46)	<0.0001	0.34 (-0.52 to 1.19)	3.93 (2.81 to 5.04)	<0.0001	-0.14 (-1.13 to 0.85)	2.70 (1.74 to 3.65)	<0.0001
SDQ total difficulties score	-0.54 (-0.83 to -0.25)	0.0002	-1.29 (-1.67 to -0.92)	0.04 (-0.30 to 0.39)	<0.0001	-0.08 (-0.42 to 0.26)	-1.61 (-2.05 to -1.17)	<0.0001	0.34 (-0.05 to 0.72)	-1.31 (-1.69 to -0.94)	<0.0001
SWEMWBS total well-being index	0.33 (0.00 to 0.66)	0.0487	1.32 (0.89 to 1.74)	-0.42 (-0.81 to -0.04)	<0.0001	0.14 (-0.24 to 0.53)	0.93 (0.43 to 1.43)	0.0034	-0.18 (-0.62 to 0.26)	0.78 (0.36 to 1.21)	0.0001
Age of sexual debut (years)	-0.35 (-1.48 to 0.78)	0.5409	-0.63 (-1.99 to 0.74)	0.01 (-1.51 to 1.53)	0.4879	-0.14 (-1.34 to 1.05)	-0.46 (-1.89 to 0.98)	0.7085	-0.63 (-2.40 to 1.15)	-0.42 (-1.58 to 0.73)	0.8308
Modified aggression scale score	-0.26 (-0.57 to 0.05)	0.0978	-0.53 (-0.89 to -0.18)	-0.03 (-0.37 to 0.31)	0.0029	-0.18 (-0.52 to 0.17)	-0.22 (-0.62 to 0.18)	0.8100	0.03 (-0.33 to 0.39)	-0.25 (-0.60 to 0.10)	0.0933
Categorical outcomes											
Ever smoked	0.58 (0.43 to 0.80)	0.0009	0.33 (0.22 to 0.50)	0.87 (0.60 to 1.25)	<0.0001	0.66 (0.46 to 0.96)	0.52 (0.33 to 0.80)	0.2887	0.93 (0.58 to 1.49)	0.46 (0.32 to 0.68)	0.0053
Ever drunk alcohol	0.72 (0.56 to 0.92)	0.0094	0.52 (0.38 to 0.70)	0.95 (0.71 to 1.26)	0.0002	0.81 (0.61 to 1.07)	0.57 (0.41 to 0.81)	0.0541	1.08 (0.78 to 1.49)	0.56 (0.41 to 0.75)	0.0009
Been offered illicit drugs	0.51† (0.36 to 0.73)	0.0003	0.44† (0.29 to 0.68)	0.57 (0.38 to 0.86)	0.2350	0.46† (0.31 to 0.68)	0.58† (0.37 to 0.93)	0.2895	0.62† (0.40 to 0.96)	0.50† (0.34 to 0.74)	0.3137
Used any contraception at first sex	1.08 (0.50 to 2.35)	0.8410	1.02 (0.39 to 2.69)	1.19 (0.36 to 3.90)	0.8409	0.55 (0.20 to 1.57)	2.63 (0.61 to 11.35)	0.0905	0.18 (0.02 to 1.45)	1.35 (0.45 to 4.13)	0.1003
Use of NHS in past 12 months	0.96 (0.82 to 1.11)	0.5652	0.96 (0.79 to 1.17)	0.95 (0.80 to 1.14)	0.9590	0.96 (0.81 to 1.16)	1.02 (0.80 to 1.29)	0.6717	1.12 (0.90 to 1.38)	0.93 (0.76 to 1.14)	0.1556
Contact with police in past 12 months	0.74 (0.56 to 0.97)	0.0269	0.62 (0.45 to 0.85)	0.93 (0.66 to 1.31)	0.0371	0.66 (0.48 to 0.92)	0.79 (0.53 to 1.17)	0.4280	0.93 (0.58 to 1.47)	0.67 (0.49 to 0.93)	0.1917

Low and high baseline bullying were defined on the basis of the GBS. High was defined as at least weekly experience of bullying or being upset by it, low was defined as less than weekly experience of bullying and not being upset by it. Low and high aggression were defined on the basis of the ESYTC school misbehaviour subscale, with high levels of behaviour problems defined as scores >0 and low levels defined as scores of 0. GBS=Gatehouse Bullying Scale. ESYTC=Edinburgh Study of Youth Transitions and Crime. PedsQL=Pediatric Quality of Life Inventory. SDQ=Strengths and Difficulties Questionnaire. SWEMWBS=Short Warwick-Edinburgh Mental Well-Being Scale40. NHS=National Health Service. *Effects are difference (95% CI) for continuous outcomes and odds ratios (95% CI) for categorical outcomes. †Proportional odds ratio.

Table 5: Subgroup analyses at 36 months

	Control group (n=20)	Intervention group (n=20)
Suicide	0	2
Responses showing potential for self-harm	0	4
Stabbing incidents	0	2
Possible non-consensual sex (including age <10 years)	6	0
Disability or long-term illness	1	0
Total	7	8

Reported at the school level (any relevant events in any student) or student level (from survey responses).

Table 6: Serious adverse events

behavioural outcomes, although not for primary outcomes. The intervention was cheap, falling into the very low cost category for UK school interventions.⁴⁶ The costs of trainers, facilitators, and school staff were an additional £47–58 per pupil in the intervention group compared with control schools over the 3 years.

We found an effect of the intervention on bullying at 36 months (as hypothesised for our primary outcome) but not at 24 months, and we found a similar strengthening of effects over time for most secondary outcomes. This probably reflects the time needed for components of the intervention to be translated into organisational change within schools, consistent with evidence from the Gatehouse Project, a previous trial of an intervention to modify the whole-school environment to reduce health risk behaviours among Australian adolescents.⁴⁷ Although many schools did not deliver formal intervention components so well in the third year as earlier, our process evaluation suggested that by the third year schools had integrated components of the intervention into mainstream school structures and processes.

We found no effect on perpetration of aggressive behaviours, contrary to a study by Flay and colleagues,²² although consistent with the Gatehouse study⁴⁷ and findings from reviews, which suggest that school-based studies of bullying prevention interventions consistently have stronger effects on victimisation than on perpetration.¹⁸ The Gatehouse Project showed no effect on bullying or psychological problems, by contrast with our findings,²¹ although it did show similar effects on risky behaviours including substance use.⁴⁸ As predicted by our theory of change,³⁰ intervention effects were concentrated on behaviours that could be markers of disengagement from school, such as bullying, smoking, and drunkenness. We found no effects on sexual health outcomes, perhaps because our intervention did not explicitly address sexual health, or because, unlike bullying and substance use, sexual behaviours occur off the school site and in private.

In terms of strengths, participating schools were representative of the approximately 500 schools initially approached and all schools were retained in the trial. Our follow-up was sufficiently lengthy to allow both time for intervention effects to develop and investigation of

persistence of intervention effects after the end of the facilitated intervention. Student participation was high. Our outcome research team and intervention team remained independent throughout the trial and masking of lead researchers, fieldworkers, and analysts was maintained. We assessed outcomes using age-appropriate validated instruments. Although self-reported outcomes can be open to recall bias, we collected baseline data before randomisation, used instruments with standardised recall periods, and actions at the school-level are unlikely to have biased reporting between intervention and control groups.

In terms of limitations, absence of students at baseline or at some follow-up points could have introduced bias. However, if non-responders are more likely to have experienced bullying or behaviour problems, this limitation is likely to have underestimated the intervention effect. The large number of secondary outcomes investigated necessitated multiple statistical testing. To mitigate the weakness of this, we only tested prespecified secondary and subgroup analyses. Had we applied an overly conservative Bonferroni correction, four of the secondary outcomes (paediatric quality of life score, strengths and difficulties score, ever smoked regularly, ever been offered illicit drugs) would have remained significant (data not shown). Some schools in the control group implemented activities that resembled some elements of Learning Together intervention. However, only five control schools implemented activities that resembled the three key elements of the intervention (restorative practice, social and emotional skills education, and student participation in decision making) and a per-protocol analysis excluding these control schools found similar intervention effects. A sensitivity analysis excluding the six schools selected for more intensive process evaluation showed no discernible differences in intervention effects compared with the intention-to-treat analyses (data not shown).

Our study adds to the evidence that whole-school approaches to prevent bullying and aggression, and promote student health are feasible and have positive effects on a range of outcomes in a broad range of high-income, middle-income, and low-income settings.^{18,21–23} Learning Together offers the potential for broad improvements in behaviour and health in secondary schools and the results of this trial provide strong support for further development of restorative approaches in secondary schools. The findings are important for public health policy in that a single, very low cost intervention affected a related set of outcomes of public health importance. The findings provide the first experimental evidence that multiple health outcomes can be promoted by transforming the school environment and increasing educational engagement.

We found positive effects of Learning Together despite variable fidelity to the intervention. For such organisational-change interventions, traditional fidelity of form (what intervention components were delivered)

might be less important than overall fidelity of function (whether overall the intervention triggered the mechanisms in the ways theorised, albeit in locally appropriate ways).⁴⁸ Our findings are particularly encouraging given that many of the control schools were delivering broadly similar activities, including restorative practice and student involvement in decision making, suggesting that Learning Together packaged and promoted these activities more effectively than most schools could do on their own.

The poor fidelity for the curriculum element suggests this aspect was less likely to have contributed significantly to the benefits of Learning Together. Given that participating schools were representative of those invited for participation and included a good range in terms of attainment, deprivation, and inspectorate ratings, Learning Together could have similar effects in other schools in England and beyond. The wider value of Learning Together should be examined in further trials in diverse settings.

At a time when the mental health of young people is a major public health concern internationally,^{17,49} countries such as the UK⁷ and Australia⁹ have identified schools as a key part of improving mental health. Interventions to promote student health by modifying the whole-school environment, such as Learning Together, are likely to be one of the most efficient ways of promoting mental health and wellbeing while also addressing other health harms in adolescence, because of their potential to modify population-level risk and their wide reach across health outcomes and likely sustainability.⁵⁰

Contributors

RMV and CB led development of the Learning Together intervention, had the idea for the study, designed the trial, led the evaluation. RMV, AM (trial manager), LBe, and JM evaluated outcomes. CB led the process evaluation and intervention teams. EA and DE designed the trial. EA and CO did the statistical analysis. JS managed the trial data and did the analysis. FJ and EW did the process evaluation. RL and ZS did the economic analysis. AF, LBo, and MW developed the intervention. SS and DC contributed to contributed towards the choice and interpretation of outcomes relating to psychological problems and mental wellbeing. All authors with the exception of FJ wrote the report.

Declaration of interests

CB, LG, JS, EA, EW, JM, LB, RL, MW, CO, AM, JS, AF, ZS, DE, LB, SS, and RMV declare no competing interests. RMV is President of the Royal College of Paediatrics & Child Health. DC has received personal fees from Medtronic, Pfizer, and ASCEND.

Acknowledgments

This study was funded by the National Institute for Health Research in England under its Public Health Research Board (12/153/60) and the Education Endowment Foundation. The views expressed in this publication are those of the authors and do not necessarily reflect those of the UK NHS, the National Institute for Health Research, or the Department of Health for England. Miranda Perry played a key role in developing the intervention and coordinated its implementation in the pilot and phase 3 trials. We are grateful to the staff and students of participating schools for their dedication to the intervention and completion of the outcome surveys and process evaluation surveys and interviews. We are very grateful for the advice and support of our Trial Steering Committee and Data Monitoring Committee. We acknowledge the work of and mourn the loss of Dr Farah Jamal during the trial, whose death at the age of 30 was a tragic loss for public health research.

References

- 1 WHO. Global status report on violence prevention 2014. Geneva: World Health Organization, 2014.
- 2 Public mental health priorities: investing in the evidence. Annual Report of the Chief Medical Officer, 2013. London: Department of Health, 2014.
- 3 Krug EG, Mercy JA, Dahlberg LL, Zwi AB. The world report on violence and health. *Lancet* 2002; **360**: 1083–88.
- 4 WHO Global Consultation on Violence and Health. Violence: a public health priority. Geneva: World Health Organization, 1996.
- 5 Brooks F, Magnusson J, Klemra E, Chester K, Spencer N, Smeeton N. HBSC England National Report 2014. Hatfield: University of Hertfordshire, 2015.
- 6 Inchley J, Currie D, Young T, et al. Growing up unequal: gender and socioeconomic differences in young people's health and well-being. Copenhagen: WHO Europe, 2016.
- 7 Bevilacqua L, Shackleton N, Hale D, et al. The role of family and school-level factors in bullying and cyberbullying: a cross-sectional study. *BMC Pediatr* 2017; **17**: 160.
- 8 Arseneault L. Annual Research Review: The persistent and pervasive impact of being bullied in childhood and adolescence: implications for policy and practice. *J Child Psychol Psychiatry* 2017; **59**: 405–21.
- 9 Hair EC, Park MJ, Ling TJ, Moore KA. Risky behaviors in late adolescence: co-occurrence, predictors, and consequences. *J Adolesc Health* 2009; **45**: 253–61.
- 10 Copeland WE, Wolke D, Angold A, Costello EJ. Adult psychiatric outcomes of bullying and being bullied by peers in childhood and adolescence. *JAMA Psychiatry* 2013; **70**: 419–26.
- 11 Lereya ST, Copeland WE, Costello EJ, Wolke D. Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries. *Lancet Psychiatry* 2015; **2**: 524–31.
- 12 Fisher HL, Moffitt TE, Houts RM, Belsky DW, Arseneault L, Caspi A. Bullying victimisation and risk of self-harm in early adolescence: longitudinal cohort study. *BMJ* 2012; **344**: e2683.
- 13 Winsper C, Lereya T, Zanarini M, Wolke D. Involvement in bullying and suicide-related behavior at 11 years: a prospective birth cohort study. *J Am Acad Child Adolesc Psychiatry* 2012; **51**: 271–82e3.
- 14 Fantaguzzi C, Allen E, Miners A, et al. Health-related quality of life associated with bullying and aggression: a cross-sectional study in English secondary schools. *Eur J Health Econ* 2018; **19**: 641–51.
- 15 Glew GM, Fan MY, Katon W, Rivara FP, Kernic MA. Bullying, psychosocial adjustment, and academic performance in elementary school. *Arch Pediatr Adolesc Med* 2005; **159**: 1026–31.
- 16 Steer A. Learning behaviour: lessons learned. A review of behaviour standards and practices in our schools. London: Department for Children, Families and Schools, 2009.
- 17 Transforming children and young people's mental health provision: a Green Paper. London: Department of Health and Department for Education, England, 2017.
- 18 Langford R, Bonell CP, Jones HE, et al. The WHO Health Promoting School framework for improving the health and well-being of students and their academic achievement. *Cochrane Database Syst Rev* 2014; **4**: CD008958.
- 19 Bonell C, Fletcher A, McCambridge J. Improving school ethos may reduce substance misuse and teenage pregnancy. *BMJ* 2007; **334**: 614–16.
- 20 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.
- 21 Bond L, Patton G, Glover S, 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.
- 22 Flay BR, Graumlich S, Segawa E, Burns JL, Holliday MY. Effects of 2 prevention programs on high-risk behaviors among African American youth: a randomized trial. *Arch Pediatr Adolesc Med* 2004; **158**: 377–84.
- 23 Shinde S, Weiss HA, Varghese B, et al. Promoting school climate and health outcomes with the SEHER multi-component secondary school intervention in Bihar, India: a cluster-randomised controlled trial. *Lancet* 2018; published online Nov 22. [http://dx.doi.org/10.1016/S0140-6736\(18\)31615-5](http://dx.doi.org/10.1016/S0140-6736(18)31615-5).

- 24 Morrison B. Restorative justice in schools. In: Elliot E, Gordon RM, eds. *New directions in restorative justice: issues, practices, evaluation*. Cullompton: Willan Publishing; 2005: 26–52.
- 25 National Evaluation of the Restorative Justice in Schools Programme. London: Youth Justice Board for England and Wales, 2004.
- 26 Farrington DP, Tofi MM. School-based programs to reduce bullying and victimization. *Campbell System Rev* 2009; 5.
- 27 Durlak JA, Weissberg RP, Dymnicki AB, Taylor RD, Schellinger KB. The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Dev* 2011; 82: 405–32.
- 28 Bonell C, Fletcher A, Fitzgerald-Yau N, et al. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): a pilot randomised controlled trial. *Health Technol Assess* 2015; 19: 1–110.
- 29 Bonell C, Allen E, Christie D, et al. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial. *Trials* 2014; 15: 381.
- 30 Tobler AL, Komro KA, Dabroski A, Aveyard P, Markham WA. Preventing the link between SES and high-risk behaviors: "value-added" education, drug use and delinquency in high-risk, urban schools. *Prev Sci* 2011; 12: 211–21.
- 31 UK School Performance Tables. <https://www.gov.uk/school-performance-tables> (accessed Sept 9, 2018).
- 32 Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ* 2015; 350: h1258.
- 33 Green J, Thorogood N. *Qualitative methods for health research*. London: Sage, 2009.
- 34 NICE. *Methods for the development of NICE public health guidance (3rd edition)*: National Institute for Health and Care Excellence, 2012.
- 35 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 Sch Health* 2007; 77: 75–79.
- 36 Smith DJ. *School experience and delinquency at ages 13 to 16*. Edinburgh: The University of Edinburgh, Centre for Law and Society; 2006.
- 37 Varni JW, Burwinkle TM, Seid M. The PedsQL 4.0 as a school population health measure: feasibility, reliability, and validity. *Qual Life Res* 2006; 15: 203–15.
- 38 Clarke A, Friede T, Putz R, 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.
- 39 Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatr* 2000; 177: 534–39.
- 40 A4: Modified Aggression Scale. In: Hamburger ME, Basile KC, Vriolo AM, eds. *Measuring bullying victimization, perpetration, and bystander experiences: a compendium of assessment tools*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, 2011: 12.
- 41 Bridges S, Gill V, Omole T, Sutton R, Wright V. Smoking, drinking and drug use among young people in England in 2010. London: NHS Information Centre for Health and Social Care, 2011.
- 42 Stephenson JM, Strange V, Forrest S, et al. Pupil-led sex education in England (RIPPLE study): cluster-randomised intervention trial. *Lancet* 2004; 364: 338–46.
- 43 Hale DR, Patalay P, Fitzgerald-Yau N, et al. School-level variation in health outcomes in adolescence: analysis of three longitudinal studies in England. *Prev Sci* 2014; 15: 600–10.
- 44 Mytton J, DiGiuseppe C, Gough D, Taylor R, Logan S. School-based secondary prevention programmes for preventing violence. *Cochrane Database Syst Rev* 2006; 3: CD004606.
- 45 Currie C, Molcho M, Boyce W, Holstein B, Torsheim T, Richter M. Researching health inequalities in adolescents: the development of the Health Behaviour in School-Aged Children (HBSC) family affluence scale. *Soc Sci Med* 2008; 66: 1429–36.
- 46 Higgins S, Matsipataki M, Kokotsaki D, Coe R, Major LE, Coleman RE. *The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit: Technical Appendices*. London: Educational Endowment Foundation, 2013.
- 47 Patton GC, Bond L, Carlin JB, 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: 1582–87.
- 48 Hawe P, Shiell A, Riley T. Complex interventions: how "out of control" can a randomised controlled trial be? *BMJ* 2004; 328: 1561–63.
- 49 Scott S, Kearney J. \$100m funding boost for school mental health programs and Headspace centres. ABC News Online. Jan 7, 2018. <http://www.abc.net.au/news/2018-01-08/more-dollars-for-kids-mental-health/9310172> (accessed Sept 9, 2018).
- 50 Bonell C, Wells H, Harden A, et al. The effects on student health of interventions modifying the school environment: systematic review. *J Epidemiol Community Health* 2013; 67: 677–81.

Appendix 8: Modifying the secondary school environment to reduce bullying and aggression: the INCLUSIVE cluster RCT

NIHR | National Institute
for Health Research

Journals Library



Public Health Research

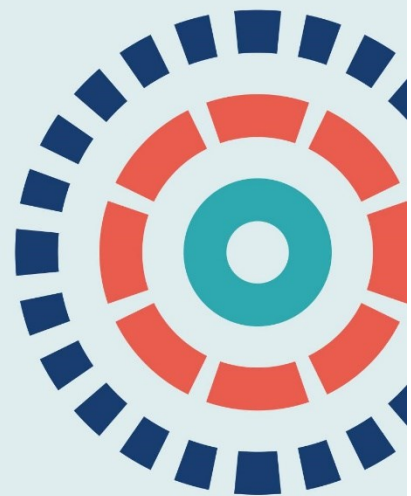
Volume 7 • Issue 18 • October 2019

ISSN 2050-4381

Modifying the secondary school environment to reduce bullying and aggression: the INCLUSIVE cluster RCT

Chris Bonell, Elizabeth Allen, Emily Warren, Jennifer McGowan, Leonardo Bevilacqua, Farah Jamal, Zia Sadique, Rosa Legood, Meg Wiggins, Charles Opondo, Anne Mathiot, Joanna Sturgess, Sara Papparini, Adam Fletcher, Miranda Perry, Grace West, Tara Tancred, Stephen Scott, Diana Elbourne, Deborah Christie, Lyndal Bond and Russell M Viner

DOI 10.3310/phr07180



Modifying the secondary school environment to reduce bullying and aggression: the INCLUSIVE cluster RCT

Chris Bonell,^{1*} Elizabeth Allen,² Emily Warren,¹
Jennifer McGowan,³ Leonardo Bevilacqua,³
Farah Jamal,^{4†} Zia Sadique,⁵ Rosa Legood,⁵
Meg Wiggins,⁴ Charles Opondo,² Anne Mathiot,³
Joanna Sturgess,² Sara Papparini,¹ Adam Fletcher,⁶
Miranda Perry,⁷ Grace West,³ Tara Tancred,¹
Stephen Scott,⁸ Diana Elbourne,² Deborah Christie,⁹
Lyndal Bond¹⁰ and Russell M Viner³

¹Department of Public Health, Environments and Society, London School of Hygiene & Tropical Medicine, London, UK

²Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK

³Institute of Child Health, University College London Great Ormond Street Institute of Child Health, London, UK

⁴Department of Social Science, University College London Institute of Education, London, UK

⁵Department of Health Services Research and Policy, London School of Hygiene & Tropical Medicine, London, UK

⁶School of Social Sciences, Cardiff University, Cardiff, UK

⁷World Class Schools Quality Mark, St Albans, UK

⁸Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

⁹University College London Hospitals NHS Foundation Trust, London, UK

¹⁰College of Health and Biomedicine, Victoria University, Melbourne, VIC, Australia

*Corresponding author

†In memoriam

Declared competing interests of authors: Russell M Viner is President of the Royal College of Paediatrics and Child Health. Stephen Scott reports grants from King's College London during the conduct of the study. Chris Bonell is a member of Public Health Research Research Funding Board. Deborah Christie reports personal fees from Medtronic plc (Dublin, Ireland), personal fees from Pfizer Inc. (New York, NY, USA) and personal fees from ASCEND (Watford, UK), all unrelated to the submitted work.

Published October 2019

DOI: 10.3310/phr07180

This report should be referenced as follows:

Bonell C, Allen E, Warren E, McGowan J, Bevilacqua L, Jamal F, *et al.* Modifying the secondary school environment to reduce bullying and aggression: the INCLUSIVE cluster RCT. *Public Health Res* 2019;**7**(18).

Public Health Research

ISSN 2050-4381 (Print)

ISSN 2050-439X (Online)

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

Editorial contact: journals.library@nihr.ac.uk

The full PHR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/phr. 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 *Public Health Research* journal

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

Reviews in *Public Health Research* 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.

PHR programme

The Public Health Research (PHR) programme, part of the National Institute for Health Research (NIHR), is the leading UK funder of public health research, evaluating public health interventions, providing new knowledge on the benefits, costs, acceptability and wider impacts of non-NHS interventions intended to improve the health of the public and reduce inequalities in health. The scope of the programme is multi-disciplinary and broad, covering a range of interventions that improve public health.

For more information about the PHR programme please visit the website: <https://www.nihr.ac.uk/explore-nihr/funding-programmes/public-health-research.htm>

This report

The research reported in this issue of the journal was funded by the PHR programme as project number 12/153/60. The contractual start date was in March 2014. The final report began editorial review in June 2018 and was accepted for publication in November 2018. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The PHR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report 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 PHR programme or the Department of Health and Social Care. 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 PHR programme or the Department of Health and Social Care.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Bonell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. 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).

NIHR Journals Library Editor-in-Chief

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Senior Clinical Researcher, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals) and Editor-in-Chief of HS&DR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Consultant Advisor, Wessex Institute, University of Southampton, UK

Ms Tara Lamont Director, NIHR Dissemination Centre, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

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

Professor Geoffrey Meads Professor of Wellbeing Research, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, 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 Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

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

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk

Abstract

Modifying the secondary school environment to reduce bullying and aggression: the INCLUSIVE cluster RCT

Chris Bonell,^{1*} Elizabeth Allen,² Emily Warren,¹ Jennifer McGowan,³ Leonardo Bevilacqua,³ Farah Jamal,^{4†} Zia Sadique,⁵ Rosa Legood,⁵ Meg Wiggins,⁴ Charles Opondo,² Anne Mathiot,³ Joanna Sturgess,² Sara Papparini,¹ Adam Fletcher,⁶ Miranda Perry,⁷ Grace West,³ Tara Tancred,¹ Stephen Scott,⁸ Diana Elbourne,² Deborah Christie,⁹ Lyndal Bond¹⁰ and Russell M Viner³

¹Department of Public Health, Environments and Society, London School of Hygiene & Tropical Medicine, London, UK

²Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK

³Institute of Child Health, University College London Great Ormond Street Institute of Child Health, London, UK

⁴Department of Social Science, University College London Institute of Education, London, UK

⁵Department of Health Services Research and Policy, London School of Hygiene & Tropical Medicine, London, UK

⁶School of Social Sciences, Cardiff University, Cardiff, UK

⁷World Class Schools Quality Mark, St Albans, UK

⁸Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

⁹University College London Hospitals NHS Foundation Trust, London, UK

¹⁰College of Health and Biomedicine, Victoria University, Melbourne, VIC, Australia

*Corresponding author Chris.Bonell@lshtm.ac.uk

†In memoriam

Background: Bullying, aggression and violence among children and young people are some of the most consequential public mental health problems.

Objectives: The INCLUSIVE (initiating change locally in bullying and aggression through the school environment) trial evaluated the Learning Together intervention, which involved students in efforts to modify their school environment using restorative approaches and to develop social and emotional skills. We hypothesised that in schools receiving Learning Together there would be lower rates of self-reported bullying and perpetration of aggression and improved student biopsychosocial health at follow-up than in control schools.

Design: INCLUSIVE was a cluster randomised trial with integral economic and process evaluations.

Setting: Forty secondary schools in south-east England took part. Schools were randomly assigned to implement the Learning Together intervention over 3 years or to continue standard practice (controls).

Participants: A total of 6667 (93.6%) students participated at baseline and 5960 (83.3%) students participated at final follow-up. No schools withdrew from the study.

ABSTRACT

Intervention: Schools were provided with (1) a social and emotional curriculum, (2) all-staff training in restorative approaches, (3) an external facilitator to help convene an action group to revise rules and policies and to oversee intervention delivery and (4) information on local needs to inform decisions.

Main outcome measures: Self-reported experience of bullying victimisation (Gatehouse Bullying Scale) and perpetration of aggression (Edinburgh Study of Youth Transitions and Crime school misbehaviour subscale) measured at 36 months. Intention-to-treat analysis using longitudinal mixed-effects models.

Results: Primary outcomes – Gatehouse Bullying Scale scores were significantly lower among intervention schools than among control schools at 36 months (adjusted mean difference -0.03 , 95% confidence interval -0.06 to 0.00). There was no evidence of a difference in Edinburgh Study of Youth Transitions and Crime scores. Secondary outcomes – students in intervention schools had higher quality of life (adjusted mean difference 1.44 , 95% confidence interval 0.07 to 2.17) and psychological well-being scores (adjusted mean difference 0.33 , 95% confidence interval 0.00 to 0.66), lower psychological total difficulties (Strengths and Difficulties Questionnaire) score (adjusted mean difference -0.54 , 95% confidence interval -0.83 to -0.25), and lower odds of having smoked (odds ratio 0.58 , 95% confidence interval 0.43 to 0.80), drunk alcohol (odds ratio 0.72 , 95% confidence interval 0.56 to 0.92), been offered or tried illicit drugs (odds ratio 0.51 , 95% confidence interval 0.36 to 0.73) and been in contact with police in the previous 12 months (odds ratio 0.74 , 95% confidence interval 0.56 to 0.97). The total numbers of reported serious adverse events were similar in each arm. There were no changes for staff outcomes. Process evaluation – fidelity was variable, with a reduction in year 3. Over half of the staff were aware that the school was taking steps to reduce bullying and aggression. Economic evaluation – mean (standard deviation) total education sector-related costs were £116 (£47) per pupil in the control arm compared with £163 (£69) in the intervention arm over the first two facilitated years, and £63 (£33) and £74 (£37) per pupil, respectively, in the final, unfacilitated, year. Overall, the intervention was associated with higher costs, but the mean gain in students' health-related quality of life was slightly higher in the intervention arm. The incremental cost per quality-adjusted life year was £13,284 (95% confidence interval $-\text{£}32,175$ to $\text{£}58,743$) and £1875 (95% confidence interval $-\text{£}12,945$ to $\text{£}16,695$) at 2 and 3 years, respectively.

Limitations: Our trial was carried out in urban and periurban settings in the counties around London. The large number of secondary outcomes investigated necessitated multiple statistical testing. Fidelity of implementation of Learning Together was variable.

Conclusions: Learning Together is effective across a very broad range of key public health targets for adolescents.

Future work: Further studies are required to assess refined versions of this intervention in other settings.

Trial registration: Current Controlled Trials ISRCTN10751359.

Funding: This project was funded by the National Institute for Health Research (NIHR) Public Health Research programme and will be published in full in *Public Health Research*; Vol. 7, No. 18. See the NIHR Journals Library website for further project information. Additional funding was provided by the Educational Endowment Foundation.

Contents

List of tables	xiii
List of figures	xv
List of supplementary material	xvii
List of abbreviations	xix
Plain English summary	xxi
Scientific summary	xxiii
Chapter 1 Introduction	1
Scientific background	1
Whole-school-based interventions	2
<i>Process evaluation of whole-school health interventions</i>	3
Restorative justice	4
Social and emotional education	5
Objectives and hypotheses	5
Chapter 2 Methods	7
Trial design	7
Refinements to the trial informed by the pilot	7
Refinements to the trial after commencement	7
Participants	8
Eligibility criteria for participants	8
Settings and locations	8
Intervention	9
<i>Learning Together</i>	9
Materials	9
Procedures	9
<i>Action groups</i>	10
<i>Social and emotional learning curriculum</i>	10
<i>Restorative practice</i>	10
Training	11
Modes of delivery	11
Location	11
Dose	11
Planned adaptations	11
Unplanned modifications	11
Comparator: control schools	12
Assessment and follow-up	12
<i>Assessment of effectiveness</i>	12
Data management	12
Ethics arrangements	13
<i>Informing participants</i>	13
<i>Consent</i>	13
<i>Duty of care and confidentiality</i>	13

CONTENTS

<i>Ethics review</i>	14
<i>Assessment of harms</i>	14
Outcomes	14
<i>Primary outcomes</i>	14
<i>Secondary outcomes</i>	14
Demographic information	15
<i>Sex and ethnicity</i>	15
<i>Socioeconomic status: the Family Affluence Scale</i>	15
<i>Data collected directly from schools</i>	15
<i>Individual staff-level outcomes</i>	16
<i>School-level outcomes</i>	16
Process evaluation	17
<i>Trial context</i>	17
<i>Trial arm fidelity</i>	17
<i>Participation, reach and dose</i>	17
<i>Reception and responsiveness</i>	17
<i>Data collection</i>	18
<i>Data collected in all schools</i>	18
<i>Data collected in case study schools</i>	20
<i>Ethics</i>	21
Economic evaluation	21
Changes to trial outcomes	21
<i>Amendments to and deviations from the protocol (process evaluation)</i>	21
Sample size	22
Randomisation	23
<i>Sequence generation</i>	23
Allocation concealment mechanisms	23
Implementation	23
Blinding	24
Statistical methods	24
<i>Primary and secondary outcome analysis</i>	24
<i>Process evaluation analyses</i>	25
Chapter 3 Results	27
Participant flow	27
<i>Recruitment</i>	27
Baseline data	27
Numbers analysed	27
Outcomes and estimation	27
<i>Primary and secondary analysis</i>	27
Staff outcome data	34
Ancillary analyses	34
Adverse events	34
Process evaluation	34
Response rates	36
Baseline context of intervention schools	39
<i>Motivation to participate in the trial</i>	39
Context of control schools	40
<i>Existing relevant services and practices in control schools during the trial period</i>	40
<i>Existing relevant policies and practices in intervention schools during the trial period</i>	43
Intervention implementation	45
<i>Intervention fidelity</i>	45
Participation and reach	49

Reception and responsiveness	50
<i>Needs assessment surveys</i>	53
<i>Training</i>	53
<i>Action group meetings</i>	54
<i>Restorative practice</i>	61
<i>Social and emotional education</i>	66
<i>Intermediate outcomes and mediation</i>	67
Chapter 4 Economic evaluation	71
Methodology	71
<i>Study question, selection of alternatives and form of evaluation</i>	71
<i>Effectiveness data and benefit measurement and evaluation</i>	71
<i>Resource use measurement and costing</i>	72
<i>Modelling and adjustments for timing of costs and benefits</i>	73
<i>Statistical analysis and uncertainty</i>	73
Results	73
<i>Resource use and costs</i>	73
<i>Health outcomes</i>	75
<i>Cost-effectiveness</i>	75
Chapter 5 Discussion	81
Primary and secondary results	81
Intervention cost and cost-effectiveness	82
Trial context	82
Fidelity and acceptability	82
Determinants of implementation	82
Reach and awareness	83
Intermediate outcomes	83
Strengths and limitations	83
Chapter 6 Conclusions	85
Implications for research and practice	85
Chapter 7 Other information	87
Patient and public involvement	87
Research governance	87
<i>Trial documentation</i>	87
<i>Trial registration and conduct</i>	87
<i>Sponsor</i>	87
<i>Trial Steering Committees</i>	88
<i>Data Monitoring and Ethics Committee</i>	88
<i>Study management</i>	88
Acknowledgements	89
References	93
Appendix 1 Tables and figures	103
Appendix 2 The young researchers group feedback example	127
Appendix 3 Example school report, year 3 (anonymised)	131

List of tables

TABLE 1 Primary and secondary outcomes at baseline by trial arm	29
TABLE 2 Primary outcomes and intervention effects at 36 months	30
TABLE 3 Secondary outcomes and intervention effects at 24 months	31
TABLE 4 Staff outcomes and intervention effects at 24 and 36 months	35
TABLE 5 Number of serious adverse events by arm across the trial	36
TABLE 6 Staff awareness of the intervention processes	37
TABLE 7 Student awareness of the intervention processes	38
TABLE 8 Adjusted intention-to-treat and per-protocol analysis for the main primary and secondary student outcomes at 36 months	39
TABLE 9 Overall fidelity per facilitator	46
TABLE 10 All-staff training implementation	46
TABLE 11 Needs assessment survey implementation	47
TABLE 12 Staff awareness of the intervention processes	51
TABLE 13 Student awareness of the intervention processes	52
TABLE 14 Participant satisfaction with the in-depth training	53
TABLE 15 Member satisfaction with AGs	55
TABLE 16 Learner Empowerment Scale for AG members	62
TABLE 17 Mediator measures response and reliability	68
TABLE 18 Intervention effects on mediators at 24 and 36 months	69
TABLE 19 School staff time associated with the delivery of the intervention and dealing with bullying and aggression per student	74
TABLE 20 Trainers, facilitators and staff costs associated with the delivery of the intervention and dealing with bullying and aggression per student	75
TABLE 21 Student health service and police resource use	76
TABLE 22 Cost–consequences results	77
TABLE 23 Cost–utility results	77

LIST OF TABLES

TABLE 24 Sensitivity analysis reporting costs, QALYs and ICERs at 36 months	79
TABLE 25 Effect of fidelity score on costs and QALYs at 36 months	79
TABLE 26 Comparison of Inclusive study schools with non-recruited schools and average for secondary schools in England	104
TABLE 27 Demographic characteristics of participants at baseline, 24 months and 36 months	105
TABLE 28 Staff outcomes at baseline	106
TABLE 29 Subgroup analyses by sex, baseline bullying and baseline aggression for key primary and secondary analyses at 36 months	107
TABLE 30 Effect of fidelity score on the primary outcomes at 24 months and 36 months	109
TABLE 31 Process evaluation response rates	109
TABLE 32 Overall fidelity in years 1–2 and 3	110
TABLE 33 All-staff training implementation	110
TABLE 34 Participant satisfaction with in-depth training	111
TABLE 35 Member satisfaction with action groups	112
TABLE 36 Response rates	113
TABLE 37 Overall fidelity in years 1–2	114
TABLE 38 Overall fidelity in year 3	115
TABLE 39 Staff and students participating in intervention activities	116
TABLE 40 Intervention effects on student primary and secondary outcomes at 36 months before and after adjusting for Beyond Blue School Climate Questionnaire and anti-school peer group mediators at 24 months	117
TABLE 41 Items and sources (tool) of resource use items and HRQoL	120
TABLE 42 Student HRQoL utility scores at baseline	121
TABLE 43 Student HRQoL utility scores at 24 months	122
TABLE 44 Student HRQoL utility scores at 36 months	123
TABLE 45 Proportion of missing values for all primary and secondary outcomes at baseline, 24 months and 36 months	124

List of figures

FIGURE 1 The CONSORT flow diagram	28
FIGURE 2 Uncertainty in the mean costs (£) and QALY differences and their distribution for intervention versus control at 24 months	71
FIGURE 3 Uncertainty in the mean costs (£) and QALY differences and their distribution for intervention versus control at 36 months	72
FIGURE 4 Cost-effectiveness acceptability curve, reporting the probability that the intervention is cost-effective (at 24 months) at alternative willingness to pay for a QALY gain	78
FIGURE 5 Cost-effectiveness acceptability curve, reporting the probability that the intervention is cost-effective (at 36 months) at alternative willingness to pay for a QALY gain	78
FIGURE 6 The INCLUSIVE logic model	125

List of supplementary material

Report Supplementary Material 1 Staff and student questionnaires, consent forms and information sheets

Supplementary material can be found on the NIHR Journals Library report project page (www.journalslibrary.nihr.ac.uk/programmes/phr/1215360/#/documentation).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

List of abbreviations

AAYP	Aban Aya Youth Project	MRC	Medical Research Council
AG	Action Group	NICE	National Institute for Health and Care Excellence
AGM	action group meeting	NIHR	National Institute for Health Research
CEA	cost-effectiveness analysis	Ofsted	Office for Standards in Education, Children's Services and Skills
CHU9D	Child Health Utility 9D	PedsQL	Paediatric Quality of Life Inventory
CI	confidence interval	PSHE	personal, social and health education
CTU	Clinical Trial Unit	QALY	quality-adjusted life-year
DMC	Data Monitoring Committee	QoL	quality of life
ESYTC	Edinburgh Study of Youth Transitions and Crime	RCT	randomised controlled trial
FAS	Family Affluence Scale	RP	restorative practice
FSM	free school meals	SD	standard deviation
GBS	Gatehouse Bullying Scale	SDQ	Strengths and Difficulties Questionnaire
GCSE	General Certificate of Secondary Education	SES	socioeconomic status
GP	general practitioner	SF-12	Short Form questionnaire-12 items
HRQoL	health-related quality of life	SLT	senior leadership team
HTA	Health Technology Assessment	SWEMWBS	Short Warwick–Edinburgh Mental Well-being Scale
ICER	incremental cost-effectiveness ratio	TSC	Trial Steering Committee
ID	identifier	UCL	University College London
IDACI	Income Deprivation Affecting Children Index	WHO	World Health Organization
INCLUSIVE	initiating change locally in bullying and aggression through the school environment	YPDP	Young People's Development Programme
LSHTM	London School of Hygiene & Tropical Medicine	YRG	young researchers group
LT	Learning Together		

Plain English summary

Bullying, aggression and violence among young people are important mental health problems. The trial reported here evaluated the Learning Together intervention, which involved school staff and students collaborating on an 'action group' to change school rules and policies and make other changes across the school to make it a healthier place. This included using restorative approaches (which focus on improving relationships) rather than merely punishment-based approaches to discipline and using a classroom curriculum aimed at fostering social and emotional skills. These aimed to reduce bullying and aggression and to promote student health and well-being.

We compared 20 schools in south-east England that were randomly allocated to deliver the intervention over 3 years with 20 schools continuing with existing practices. Schools were provided with an external facilitator to help convene the action group, with all-staff training in restorative approaches and with curriculum materials.

At the start, 6677 students (over 9 in every 10) completed questionnaires. No schools withdrew from the study. When questionnaires were repeated 3 years later, the numbers of students reporting experience of being bullied were significantly smaller among intervention schools than among comparison schools. There was no evidence of a difference in the numbers of students reporting acts of aggression. Students in intervention schools reported having higher quality of life and psychological well-being, lower psychological difficulties and lower chances of having smoked, drunk alcohol, been offered or tried illicit drugs and been in contact with the police in the previous 12 months. The intervention was acceptable to schools and provided strong value for money.

Learning Together is effective across a very broad range of key public health targets for adolescents.

Scientific summary

Background

Bullying, aggression and violence among children and young people are some of the most consequential public mental health problems. There is clear evidence of a range of physical and mental health harms associated with exposure to bullying and violence, including substance use, poorer long-term mental health, suicide and self-harm, and lower educational attainment. Childhood experiences of bullying and violence influence health and well-being both contemporaneously and well into adult life. Prevention of bullying and violence is, therefore, a major priority for public health and education systems internationally, with schools being a key focus of policy initiatives to improve young people's mental health and well-being.

The INCLUSIVE (initiating change locally in bullying and aggression through the school environment) trial evaluates the Learning Together intervention. In 2014 we developed this intervention based on the three most promising approaches to reduce bullying and other health risks. The first approach is 'whole-school' interventions, which aim to modify overall school policies and systems rather than merely deliver classroom-based lessons addressing bullying or other outcomes. A key element of many of these interventions appears to be increasing student engagement with school as a social determinant of health, particularly for the most socially disadvantaged students, who are at highest risk of poor health and educational outcomes. The second promising approach is restorative practice. This aims to prevent and/or resolve conflicts between students or between staff and students to prevent further harm. It enables victims to communicate the impact of the harm to perpetrators, and for perpetrators to acknowledge and take steps to remedy this, to avoid further harm. The third approach is social and emotional education. Evidence shows that classroom curricula that teaches young people the skills needed to manage emotions and relationships can enhance social relationships, improve mental health and reduce bullying.

Objectives

We hypothesised that in secondary schools randomly allocated to receive Learning Together there would be lower rates of self-reported bullying and perpetration of aggression, and improved student and staff secondary outcomes at follow-up compared with control schools, and that Learning Together would be cost-effective compared with standard school practice. In this paper, we report student health and behaviour outcomes. Data on student educational outcomes and staff outcomes will be published later because routine administrative data will not be available until later in 2019.

Methods

Design and participants

We undertook a two-arm repeat cross-sectional cluster randomised controlled trial of Learning Together with an integral economic and process evaluation in 40 secondary schools in south-east England, with schools as the unit of allocation. Our study population consisted of all students in the school at the end of year 7 (aged 11–12 years) at baseline, and at 24-month (end of year 9; aged 13–14 years) and 36-month (end of year 10; aged 14–15 years) follow-up, as well as school teaching and teaching assistant staff at each time point.

Intervention

School staff were offered training in restorative practices, with participants given written summaries of the material covered in training. Schools were provided with a manual to guide them in convening and running an action group. For the first 2 years of the intervention, schools were provided with an external

facilitator for the action group. Schools were sent a report on student needs, which detailed the findings from a survey of students aged 11–12 years about their attitudes to and experiences of school, and experiences of bullying, aggression and other risk behaviours, at the end of each year (see *Appendix 3*). Schools were provided with written lesson plans and slides to guide the delivery of a classroom-based social and emotional skills curriculum.

Guided by the manual and facilitator, schools instituted action groups comprising staff and students. In the first 2 years of the intervention, these action groups reviewed school rules and policies relating to discipline and behaviour management so that they supported the delivery of restorative practice, and co-ordinated intervention delivery across the school in all 3 years. The facilitator ensured that meetings were scheduled, and attended these to ensure that the meetings were participative and focused on deciding and implementing actions. Action groups reviewed the report of student needs to inform decisions. Schools delivered classroom-based social and emotional skills education in personal, social and health education lessons and/or integrated this into tutor time or various subject lessons (e.g. English) to students in the trial cohort as they moved through years 8–10 (aged 12–15 years). Schools selected modules for each year, such as 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.

Primary restorative practices delivered in schools in all three years involved staff using restorative language (the respectful use of language to challenge or support behaviour in a manner that preserves or enhances the relationship) and circle time (classes coming together to discuss their feelings and air any problems so that these may be addressed before they escalate), underpinned by supportive schools' rules and policies and the social and emotional skills curriculum. Secondary restorative practices involved some staff implementing restorative conferences (the parties to a conflict being invited to a facilitated face-to-face meeting to discuss the incident and its impact on the victim and for the perpetrator to take responsibility for their actions and avoid further harms).

Schools randomised to the control group continued with their normal practice and received no additional input.

Primary outcomes

The primary outcomes were self-reported experiences of bullying victimisation and perpetration of aggressive behaviour measured at 36 months. Bullying victimisation was assessed with the Gatehouse Bullying Scale. Perpetration of aggressive behaviour was measured using the Edinburgh Study of Youth Transitions and Crime school misbehaviour subscale.

Secondary outcomes

The Gatehouse Bullying Scale and the Edinburgh Study of Youth Transitions and Crime scale were assessed at 24 months as secondary outcomes. The following secondary outcomes were measured at 36 months: quality of life (Paediatric Quality of Life Inventory), well-being (Short Warwick–Edinburgh Mental Well-Being Scale), psychological problems (Strengths and Difficulties Questionnaire), bullying perpetration (Modified Aggression Scale), substance use (smoking, alcohol use and illicit drug use), sexual risk behaviour (age of sexual debut and use of contraception), use of NHS health services and contact with police.

Recruitment

We identified and contacted all potentially eligible schools in Greater London and surrounding counties (Surrey, Kent, Essex, Hertfordshire, Buckinghamshire and Berkshire) between March and June 2014. The 40 participating schools did not differ from the 450 non-recruited schools in school size, population, deprivation, student attainment or value-added education. However, participating schools were more likely to have an Office for Standards in Education, Children's Services and Skills (Ofsted) rating of good or outstanding.

Eligible schools:

- i. Were mainstream secondary schools within the state education system in south-east England.
- ii. Had a most recent school quality rating by Ofsted of 'requires improvement'/'satisfactory', 'good' or 'outstanding'. Schools with an 'inadequate/poor' rating were excluded, as these schools are subject to special measures that were likely to impede Learning Together delivery.

Data collection

Baseline surveys were completed March–July 2014, with 24-month follow-up in April–June 2016 and 36-month follow-up in April–June 2017. Student self-report data were collected using paper questionnaires; students completed these in lesson time in classrooms under exam conditions facilitated by trained researchers with teachers present but unable to read student responses. The field workers assisted students with questions that they did not understand and ensured that students completed as much of the questionnaire as possible. Students with mild learning difficulties or with limited command of written English were supported to complete the questionnaires by field workers.

Process evaluation

In line with Medical Research Council guidance on complex interventions and other frameworks, the process evaluation examined trial context, such as discipline systems, staff training, social and emotional learning curricula and student participation in decision-making, to assess how these differed from what was implemented in the intervention; trial fidelity; awareness (the extent to which students and staff were aware of the intervention); and reception and responsiveness.

Economic evaluation

The economic evaluation used a cost–consequences analysis with all main outcomes and evaluated incremental effects at 24 and 36 months since randomisation. Costs were identified from a public sector perspective, including education, police and NHS resources. The costs of delivering the interventions were collected from the invoices for facilitators and trainers and data from the process evaluation on school staff time requirements. The costs of staff time spent dealing with bullying were collected from the staff survey questionnaire, and the costs of NHS and police resource use data were collected from student survey questionnaire and valued accordingly.

Trial registration and amendments

The trial was prospectively registered as ISRCTN10751359 with the ISRCTN registry on 30 January 2014 and accepted for publication on 30 September 2014. The protocol was amended during the trial to refine the methods used. All amendments were approved by the independent study steering committee and the funder of the trial (National Institute for Health Research). The only change to trial outcomes was the addition of a measure of bullying perpetration (secondary outcome). All refinements were completed before the 36-month surveys were collected and before any trial analyses were conducted.

This paper presents independent research funded by the National Institute for Health Research in the UK under its Public Health Research programme (12/153/60) and the Education Endowment Foundation.

Analyses

The primary analysis of outcomes was intention to treat including all randomised schools and participants at each wave. Each measure was analysed using a separate mixed model with the outcomes from each time point treated as a repeated measures outcome. Fixed effects of treatment (Learning Together vs. control), time (baseline, 24 months and 36 months) and the interaction between treatment and time were specified, and the estimated baseline measures were constrained to be identical in the two arms of the trial.

As prespecified in the statistical analysis plan, we carried out analyses adjusted only for baseline measures of the outcomes and the analyses adjusted for baseline measures of outcomes, sex, ethnicity and

SCIENTIFIC SUMMARY

socioeconomic status, as well as for the school-level stratifying factors (single-sex vs. mixed-sex school, school-level deprivation, value-added strata), as the primary analysis.

For the co-primary outcomes (Gatehouse Bullying Scale and Edinburgh Study of Youth Transitions and Crime), mixed linear regression models with random effects at the participant and school levels were used to estimate a mean difference in Gatehouse Bullying Scale and Edinburgh Study of Youth Transitions and Crime scores between the two arms of the trial.

Economic analyses

The primary economic evaluation was a cost–consequences analysis. The economic analysis used general linear mixed regression models that allow for clustering of students within schools, and including school as a random effect variable.

Results

A total of 6667 students in the 40 participating schools provided data at baseline, with the participation rate being 93.6% of the students on the school roll (intervention arm, 92.9%; control arm, 94.3%).

Primary outcomes

Overall Gatehouse Bullying Scale bullying scores were lower among intervention than among control schools at 36 months (adjusted mean difference -0.03 , 95% confidence interval -0.06 to 0.00 ; adjusted effect size -0.08). There was no evidence of a difference in Edinburgh Study of Youth Transitions and Crime misbehaviour/delinquency scores (adjusted mean difference -0.13 , 95% confidence interval -0.43 to 0.18 ; adjusted effect size -0.03) between the arms; however, the direction of effect suggests a positive effect of the intervention.

Secondary outcomes

There was no evidence of difference in the Gatehouse Bullying Scale overall score or the Edinburgh Study of Youth Transitions and Crime misbehaviour/delinquency scores at 24 months. At 36 months, students in intervention schools had higher quality of life (Paediatric Quality of Life Inventory adjusted effect 1.44 , 95% confidence interval 0.07 to 2.17 ; adjusted effect size 0.14) and psychological well-being scores (Short Warwick–Edinburgh Mental Well-Being Scale 0.33 , 95% confidence interval 0.00 to 0.66 ; adjusted effect size 0.07) and lower psychological total difficulties (Strengths and Difficulties Questionnaire total score -0.54 , 95% confidence interval -0.83 to -0.25 ; adjusted effect size -0.14) than students in control schools. There was evidence that those in intervention schools also had lower emotional, conduct, hyperactivity and peer problems (Strengths and Difficulties Questionnaire subscales).

Students in intervention schools had lower odds of having ever smoked regularly (odds ratio 0.58 , 95% confidence interval 0.43 to 0.80 ; adjusted risk difference -0.03 , 95% confidence interval -0.05 to -0.01), lower odds of having ever drunk alcohol (odds ratio 0.72 , 95% confidence interval 0.56 to 0.92 ; adjusted risk difference -0.03 , 95% confidence interval -0.06 to -0.01) and lower odds of having ever been offered or tried illicit drugs (odds ratio 0.51 , 95% confidence interval 0.36 to 0.73). Among students in the intervention arm who had ever smoked, there was evidence that the time since the last cigarette was longer than among those in the control arm and, similarly, that, among those who had ever drunk alcohol, there were lower odds of having drunk in the past week (odds ratio 0.67 , 95% confidence interval 0.50 to 0.91), a lower number of times being really drunk (odds ratio 0.57 , 0.33 to 0.98) and lower odds of binge drinking (odds ratio 0.77 , 95% confidence interval 0.59 to 1.00). Similarly, students in intervention schools had lower odds of having ever been in contact with police in the past 12 months than those in control schools (odds ratio 0.74 , 95% confidence interval 0.56 to 0.97 ; adjusted risk difference -0.02 , 95% confidence interval -0.04 to -0.00). We found no evidence of differences in age of sexual debut or use of contraception at first sex, bullying perpetration or use of NHS services.

Exploratory analyses suggest that the intervention may be most effective for students with higher baseline levels of bullying or aggressive behaviours. The intervention also had greater effects for boys in terms of secondary psychological and behavioural outcomes, although not in terms of primary outcomes.

Process evaluation findings

Fidelity was variable, with a reduction in the fidelity of formal intervention activities in year 3. The median fidelity score for years 1–2 (maximum possible score of 8) was 6 (interquartile range 5–7), whereas for year 3 (maximum score of 4) the median was 1 (interquartile range 0–3). In year 3, 15 schools sustained restorative practice. Interviews with action group members and focus groups with staff in case study schools suggested that, in year 3, schools had commonly incorporated what they regarded as the most useful action group functions into mainstream school structures and processes. The fidelity score for year 3 was not associated with either primary outcome. The intervention was delivered more completely when it was led by a member of staff with sufficient authority and support to make decisions and drive delivery. In many, but not all, cases, this was required to be a staff member on the school's senior leadership team.

Slightly over half of staff in intervention schools were aware that the school had been taking steps to reduce bullying and aggression, with this falling slightly between years 2 and 3.

Economic evaluation

The main time components for school staff were attending the training and curriculum delivery. We included staff restorative practice training in intervention costs; however, staff interviews suggested that training was not additional but part of existing training periods, suggesting that our intervention costs may be overestimated. The mean (standard deviation) costs per school of all staff time combined were £232,670 (£113,634) for the intervention arm and £202,405 (£103,090) for the control arm. Costs for health service use and police contacts were similar in both arms. Overall, the intervention increased costs and reduced bullying, leading to incremental costs per Gatehouse Bullying Scale score averted of £2352 at 36 months.

Limitations

The large number of secondary outcomes investigated necessitated multiple statistical testing. The Gatehouse Bullying Scale is a well-established tool to measure the occurrence of bullying victimisation; it aligns with the World Health Organization's definition of bullying but aligns less well with some other definitions, such as that of Olweus, which focuses on repeat victimisation and power imbalances between the perpetrator(s) and the victim. Some aspects of the process evaluation had low response rates.

Conclusions

We present here what is, to our knowledge, the first randomised trial of restorative approaches to reduce bullying and aggression and promote student health in schools, within a multicomponent whole-school intervention engaging students in school decision-making, and providing restorative practice and social and emotional skills education. Learning Together resulted in a very broad range of benefits for behaviour and health outcomes. Learning Together reduced student reports of bullying victimisation compared with schools continuing their standard practice. We did not identify a reduction in student reports of aggression across the whole sample. Additionally, Learning Together appeared to have larger beneficial impacts on a wide range of important secondary outcomes among students, ranging from improved psychological function, well-being and quality of life, to reductions in police contact, smoking, alcohol and drug use. We found intervention effects both in the whole sample and in those with higher levels of bullying or aggression at baseline, implying that the intervention worked to curtail existing bullying and aggression (secondary prevention) as well as prevent new bullying (primary prevention). The intervention may be most effective for students with higher baseline levels of bullying or aggressive behaviours. The intervention also had greater effects for boys in terms of secondary psychological and behavioural outcomes, although not in terms of primary outcomes. The intervention was low cost, falling into the 'very low cost' category for

school interventions according to the Educational Endowment Foundation guidance. The intervention was feasible and acceptable to deliver, with delivery promoted by the involvement of senior staff.

Implications for research and practice

Our study adds to the evidence that whole-school approaches to preventing bullying and aggression and promoting student health are feasible to implement and have positive effects on a range of outcomes in a broad range of high-, middle- and low-income settings. Learning Together offers the potential for broad improvements in behaviour and health in secondary schools and, as the first randomised controlled trial of school-based restorative practice to our knowledge, provides strong support for further development of restorative approaches in secondary schools. The results are important for public health policy, in that a single, very low-cost intervention had an impact on a clustered set of outcomes of public health importance, including bullying, mental health, well-being and quality of life, as well as the use of tobacco, alcohol and drugs.

Trial registration

This trial is registered as ISRCTN10751359.

Funding

Funding for this study was provided by the Public Health Research programme of the National Institute for Health Research. Additional funding was provided by the Educational Endowment Foundation.

Chapter 1 Introduction

Parts of this chapter have been reproduced with permission from Bonell *et al.*¹ This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Scientific background

The study protocol for this trial is available in full in *Trials*.¹ It is presented again here. Bullying, aggression and violence among children and young people are some of the most consequential public mental health problems apparent today.^{2,3} The prevalence and harms of aggressive behaviours among young people makes addressing these a public health priority.⁴⁻⁷ The World Health Organization (WHO) considers bullying to be a major adolescent health problem, defining this to include the intentional use of physical or psychological force against others.^{5,8} This includes verbal and relational aggression that aims to harm the victim or their social relations, such as by spreading rumours or purposely excluding them.^{9,10} Some definitions of bullying¹¹ emphasise that bullying refers to abuse that is committed repeatedly over time and that involves a power imbalance between the perpetrator(s) and the victim. The prevalence of bullying among British young people (at around 33%)¹² is above the European average,¹³ with approximately 25% of young people reporting that they have been subjected to serious peer bullying.¹⁴ Cyberbullying, in particular, is rapidly becoming one of the most common forms of bullying.¹⁵ There are marked social gradients: family deprivation and school-level deprivation increase the risk of experiencing bullying.¹⁶ Bullying most commonly occurs in schools^{17,18} and prevalence varies significantly.¹⁹⁻²²

Being a victim of peer bullying has been associated with an increased risk of physical health problems;^{23,24} substance use;²⁵⁻²⁸ long-term emotional, behavioural and mental health problems;²⁹⁻³³ self-harm and suicide;³⁴⁻³⁶ and poorer educational attainment.^{37,38} Students who experience physical, verbal or relational bullying regularly tend to experience the most adverse health outcomes.³⁹ There has also been evidence that childhood exposure to bullying and aggression may influence lifelong health through biological mechanisms.^{23,29,40}

The perpetrators of peer bullying are also at risk of a range of adverse emotional and mental health outcomes, including depression and anxiety.^{13,19} Therefore, the prevention of bullying, aggression and violence is a major priority for public health and education systems internationally,^{3,41} with schools a key focus of policy initiatives to improve young people's mental health and well-being.⁴² In England,⁴³ schools have a legal obligation to prevent bullying.

Bullying is often a precursor to more serious violent behaviours. One UK study⁴⁴ 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 that they had attacked another with the intention of hurting them seriously. By the age of 15–16 years, 24% of students reported that they had carried a weapon and 19% had reported attacking someone with the intention of hurting them.⁴⁴ Interpersonal violence can cause physical injury and disability, and has been also associated with long-term emotional and mental health problems. Aggression refers to behaviour that is intended to harm, either directly or indirectly, another individual who does not wish to be harmed.⁴⁵ There are also links between aggression and antisocial behaviours in young people and violent crime in adulthood.^{46,47} This is thought to result from low-level provocation and aggressive behaviours in secondary schools being educationally disruptive and emotionally harmful, and reducing educational attainments and later life chances, and therefore leading to more overt physical aggression over time.⁴⁸⁻⁵⁰ The economic costs to society as a

© Queen's Printer and Controller of HMSO 2019. This work was produced by Bonell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. 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.

whole from bullying and violence are extremely high: the total cost of crime attributable to conduct problems in childhood has been estimated to be about £60B per year in England and Wales.⁵¹

Reducing aggression, bullying and violence in British schools has been a consistent priority in public health and education policies.⁵²⁻⁵⁴ The 2009 Steer Review⁵⁰ concluded that schools' approaches to discipline, behaviour management and bullying prevention vary widely and are rarely evidence based, and that further resources and research are urgently needed. There is, therefore, a pressing need to determine which interventions are effective in addressing bullying and aggression in schools, and to scale up such interventions to local and national school networks.

Whole-school-based interventions

A number of systematic reviews have assessed school-based interventions to address bullying and aggression. Interventions that promote change across school systems and address different levels of school organisation (i.e. 'whole-school' or 'school environment' interventions) are particularly effective in reducing victimisation and bullying in comparison with curriculum interventions.⁵⁵⁻⁵⁷ 'Whole-school' or 'school environment' interventions are interventions that modify the systemic operations of schools, and they have been shown to have an impact on a range of health outcomes and risk behaviours.⁵⁷ A key element of these interventions appears to be increasing student engagement with school, particularly the most socially disadvantaged students, who are at the highest risk of poor health and educational outcomes.^{58,59} Two trials have found that such approaches are associated with reductions in risk behaviours, including violence and antisocial behaviour.^{60,61}

The effectiveness of such interventions may be attributable to the way that they address bullying as a systemic problem meriting an 'environmental solution', rather than an individual student issue.⁵⁵ Whole-school interventions are also inherently universal in their reach and are likely to provide a cost-effective and non-stigmatising approach to preventing bullying.⁵⁶ 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 that can actively support healthy behaviours and outcomes.^{62,63}

Thus, these school environment interventions take a 'socio-ecological'⁶⁴ or 'structural'⁶⁵ approach to promoting health, whereby behaviours are understood to be influenced not only by characteristics of individuals, but also by the wider social context. A recent National Institute for Health Research (NIHR)-funded systematic review on 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' (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.⁶³ Other outcomes that are improved by school environment interventions can include mental health and physical activity and reduced substance use, including alcohol, tobacco and drugs.⁶³

School environment interventions, therefore, are likely to be one of the most efficient ways of addressing multiple health harms in adolescence owing to their potential for modifying population-level risk as well as their reach and sustainability.⁶³ Multiple risk behaviours in adolescence are subject to socioeconomic stratification, and are strongly associated with poor health outcomes, social exclusion, educational failure, and poor mental health in adult life.⁶⁶ A recent King's Fund report, *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.⁶⁷

The INCLUSIVE (initiating change locally in bullying and aggression through the school environment) trial aims to evaluate the Learning Together (LT) intervention. This has been particularly informed by two international evidence-based school environment programmes. The first is the Aban Aya Youth Project (AAYP),⁶¹ a multicomponent intervention enabling schools to modify their social environment as well as delivering a social skills curriculum. This approach was 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 (AG) was established (comprising both staff and students) to review policies and prioritise the actions needed to foster a more inclusive school climate. Among boys, the intervention was associated with significant reductions in violence and aggressive behaviour.⁶¹ The intervention also brought benefits in terms of reduced sexual risk behaviours and drug use, as well as provoking behaviour and school delinquency. Second, the Gatehouse Project^{60,68} in Australia also aimed to reduce health problems by changing the school climate and promoting security, positive regard and communication among students and school staff. As with the AAYP, an AG 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 randomised controlled trial (RCT) found consistent reductions in a composite measure of health risk behaviours, which included violence and antisocial behaviour.

Process evaluation of whole-school health interventions

Most evaluations of interventions taking a whole-school approach to preventing violence and promoting health in schools examine outcomes rather than processes.^{63,69} The process evaluation of the Gatehouse Project, which greatly informed LT, found that school staff perceived the various components (needs survey, action team, external facilitator) to function synergistically. Although specific actions varied between schools, these were completed with good fidelity. Implementation was facilitated by supportive school management and the broad participation of staff and students.^{70,71} However, this evaluation did not attempt to assess systematically how the completeness of the implementation might have been influenced by schools' baseline social climate.

The Healthy School Ethos intervention was also informed by the Gatehouse Project and included several elements similar to LT, but without a curriculum or any restorative practice elements.^{72,73} Using a structured process modelled closely on the Gatehouse Project, it aimed to enable schools involved in pilots in south-east England to carry out locally determined actions to increase students' sense of security, positive self-regard, and communication with staff and students. The intervention provided an external facilitator, survey data on student needs and training, and enabled schools to convene action teams to determine priorities and ensure delivery. Students and staff co-revised rules for appropriate conduct and revised policies on bullying and student feedback. Staff were trained to improve classroom management. Process evaluation reported that the intervention was delivered with good fidelity. Locally determined actions (e.g. peer-mediators) were generally more popular than mandatory actions. Implementation was more feasible when it built on aspects of schools' baseline ethos and when someone on the senior leadership team (SLT) led actions. Student awareness of the intervention was high. Student accounts suggested that benefits might arise as much from participation in intervention processes, such as rewriting rules, as from the effects of subsequent actions. Some components reached a large proportion of students.^{72,73}

Before the current Phase III trial, the LT intervention had been piloted in four schools.⁷⁴ Overall, school staff members were consistently supportive. Although some schools were already deploying some restorative approaches, the intervention was nonetheless attractive because it enabled restorative practices to be delivered more coherently and consistently across the school. 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. Staff valued the 'external push' that was provided by the external facilitator. The intervention was highly acceptable to school staff because of its fit with national policies and school metrics focused on attendance and exclusions. Some staff reported that it took time for them to understand how the various intervention components joined up, and this could have been better explained from the outset. Staff were positive about sustainability, with some reporting that activities would continue after the pilot was completed.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Bonell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. 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.

INTRODUCTION

Regarding particular components, staff reported that the needs assessment report allowed them to see the 'big picture' and identify priorities, but some suggested that the needs assessment could also feel too 'negative' at times, especially among established staff, who sometimes viewed this as a reflection on their work at the school. Negative aspects of the needs report could also present problems for schools, because if they were inspected by the Office for Standards in Education, Children's Services and Skills (Ofsted) they would be expected to share the results with the inspectors. As with the HSE (Healthy School Ethos) evaluation, AGs were viewed positively, and it was suggested that student participation may be an active ingredient in improving relationships and engagement across the school, particularly when these involved students who might be less committed to school and might be involved in antischool peer groups. Again, the presence of a SLT member on the group was seen as critical to driving actions forward. The training, however, was more critically received, with many staff suggesting that this was too didactic and contained too few examples from secondary schools. All schools successfully implemented the curriculum, welcoming its flexibility whereby modules could be implemented using the newly provided or existing materials. The pilot lacked a large enough sample to examine how implementation and processes might vary across a range of different school contexts, and focused only on the first year of implementation, so it could not examine the processes by which the intervention might become normalised within schools' institutional policies and practices and be sustained once external facilitation is withdrawn.

Restorative justice

The INCLUSIVE trial extends the AAYP and Gatehouse interventions by including a 'restorative justice' approach. The Steer Review⁶⁰ in 2009 called for English schools to consider adopting more restorative approaches to preventing bullying and other aggressive behaviour to help minimise the harms associated with such problems. The central tenet of such approaches is repairing the harms caused to relationships and communities by criminal behaviour, rather than merely assigning blame and enacting 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 the way in which staff communicate with students in order to improve relationships, reduce conflict and repair harm.

Restorative practice aims to prevent and/or resolve conflicts between students or between staff and students to prevent further harms.⁷⁵ It enables victims to communicate the impact of the harm to perpetrators, and for perpetrators to acknowledge and take steps to remedy this in order to avoid further harms. Restorative practice can involve methods to prevent incidents (e.g. 'circle-time', which brings students together with their teacher during registration periods or other lessons to maintain good relationships) and/or to resolve incidents (e.g. 'conferencing', bringing together relevant staff, students, and, where necessary, parents and external agencies such as police or social services). Restorative practice aims to prevent the occurrence or continuation not only of bullying but also of other forms of aggression and classroom disruption. Restorative practice can be delivered instead of, or alongside, more traditional punitive discipline.⁷⁶

The theoretical basis for restorative approaches has much in common with the theory of human functioning and school organisation.⁵⁹ It is theorised that the process of students coming together, discussing the harm, and working towards a reparative plan develops perpetrators' competency through accepting responsibility for the actions and contributing to a reparative solution, and develops offenders' understanding of the realities of others. Victims are also empowered in this process as they become an active participant in the decision-making process, and the acknowledgement of the offenders' ability to offer some healing to the victim (e.g. via an apology or carrying out a sanction) gives dignity to both parties. This resonates with the ideas of improving relationships as well as promoting practical reasoning and sense of connection to school. By eliciting accountability for the harm caused to the victim and the school community and negotiating a plan for restitution, the young person is encouraged to reclaim an identity as a participant of the school community, not a peripheral outsider.⁷⁶ Through this process, the young people involved develop relational competency, and enhance their relationships with staff and other students by improving their ability to empathise and communicate effectively. Restorative approaches

might indeed be particularly suitable for 'alienated' student offenders as they are given the opportunity to develop the necessary competencies to participate as a responsible member of the school community from which previously they might have felt excluded.⁷⁷ It may also be particularly helpful for female young people, as gender theory suggests that female adolescent identity is often based within a framework of relationship and connection. Thus, application of the principles of restorative approaches becomes a natural adjunct to the therapeutic process of self-identity and growth.⁷⁷

However, to date, restorative approaches in schools have been evaluated using only non-randomised designs, and systematic reviews have called for more rigorous evaluations of restorative practice in schools.⁷⁸ Those studies that have been carried out do suggest that the restorative approach is promising both in the UK^{79–81} and internationally,^{82–84} particularly when implemented at the whole-school level. 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 in students' attitudes to bullying and reductions in offending and victimisation in schools that adopted a whole-school approach to restorative practice. Restorative approaches thus appear to have the potential to complement school-environment interventions such as Aban Aya and the Gatehouse Project. They offer a highly promising way forward for reducing aggressive behaviours among British young people. A 2009 Cochrane review⁸⁵ found no RCTs of interventions employing restorative approaches to reduce bullying in schools and recommended that this be a priority for future research.

Process evaluations report positive results in terms of feasibility and acceptability of restorative practices in schools. In New Zealand,⁸⁶ case studies of five secondary schools and colleges found that all teachers valued restorative practice and felt that it was a good strategy for managing misbehaviour. In a pilot in London schools,⁸⁷ students in schools that had implemented restorative practice reported that their school was doing a good job of stopping bullying. Teachers reported that most restorative meetings were effective at addressing bullying, gossiping, and disagreements between students and teachers.

Common challenges reported in process evaluations of restorative practice in schools include consistency in the way restorative practice is implemented.^{82,88,89} Evaluations also suggest that restorative practice is implemented most successfully when it is delivered as part of a whole-school approach, when a positive ethos has been established, and when one-to-one problem-solving skills (e.g. listening and responsibility) have been introduced into the curriculum.^{90,91}

Social and emotional education

Social and emotional education aims to educate students not merely academically but using non-cognitive social skills and emotional self-management skills to enable young people to function at school and in other areas of life and to develop resilience or 'grit'. There was evidence that classroom curricula that teaches young people the skills needed to manage their emotions and relationships can enhance social relationships, improve mental health and reduce bullying.⁹² Many whole-school interventions, including the Aban Aya⁹³ and Gatehouse⁹⁰ interventions, also include social and emotional education elements, the intention being that these act synergistically with other components, helping students to take part in intervention activities within and beyond the classroom.

Objectives and hypotheses

In 2014, we developed the LT intervention, which aimed to modify the school environment by involving students, building better relationships, using restorative approaches, and developing social and emotional skills to reduce a range of risk behaviours including bullying and aggression.⁷⁴ The intervention combined AGs of staff and students to modify the whole-school environment; training for staff to implement restorative approaches; and a social and emotional skills curriculum. These components were intended to

© Queen's Printer and Controller of HMSO 2019. This work was produced by Bonell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. 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.

INTRODUCTION

address school- and individual-level determinants of bullying and aggression and to be synergistic with one another: the AG ensuring that schools took a whole-school approach to restorative practice and the social and emotional skills curriculum enabling students to participate actively in AGs and restorative practices.

A pilot cluster RCT in eight schools met all of the prespecified feasibility and acceptability criteria.⁷⁴ We report here the findings of a full-scale cluster RCT of LT (the INCLUSIVE trial).

We hypothesised that in secondary schools randomly allocated to receive LT there would be lower rates of self-reported bullying and perpetration of aggression among students aged 14–15 years at the 36-month follow-up.

We hypothesised that at the 36-month follow-up student and staff secondary outcomes would be improved in intervention compared with control schools. More specifically, we expected improvements in students' quality of life (QoL), well-being, psychological function and attainments; reductions in school exclusion and truancy, substance use, sexual risk, NHS use and police contacts among students; and improvements in staff QoL and attendance and reductions in staff burn-out.

We hypothesised that individual-level student socioeconomic status (SES), sex, and school-level stratifying factors (single-sex vs. mixed-sex school, school-level deprivation, and value-added strata) would moderate the effectiveness of the intervention for student outcomes.

We also hypothesised that LT would be cost-effective compared with standard school practice in terms of student quality-adjusted life-years (QALYs) and costs.

Chapter 2 Methods

Parts of this chapter have been reproduced with permission from Bonell *et al.*¹ This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Trial design

We undertook a two-arm repeat cross-sectional cluster 1 : 1 RCT of the LT trial with an integral economic and process evaluation in 40 secondary schools in south-east England, with schools as the unit of allocation.¹ Our study population consisted of all students in the school at the end of year 7 (aged 11–12 years) at baseline, and at 24-month (end of year 9; aged 13–14 years) and 36-month (end of year 10; aged 14–15 years) follow-up, as well as school teaching and teaching assistant staff at each time point.

Refinements to the trial informed by the pilot

The changes made to the trial informed by the pilot included:

- identifying the Gatehouse Bullying Scale (GBS) and Edinburgh Study of Youth Transitions and Crime (ESYTC) school misbehaviour subscale as primary outcomes of bullying victimisation and perpetration of aggressive behaviours, respectively
- including validated measures of drug use, sexual behaviour and educational attendance and attainment as additional secondary outcomes
- including all 'state' schools in the recruitment pool of schools to reflect the overall population profile of schools in south-east England
- using existing school networks to facilitate timely recruitment
- using revised timetabling – project initiation in February of the preceding school year, surveys of staff and students to be conducted in the summer term each year, timetabling of intervention and staff training to be undertaken prior to September school-year start
- enhancing quantitative data on intervention fidelity, including structured independent assessments of intervention delivery
- undertaking an economic evaluation to use the Child Health Utility 9D (CHU9D) scale and to be supplemented with a cost–consequences analysis.

Refinements to the trial after commencement

During the development of the trial, the following changes to the protocol were made:

- A measure of bullying perpetration (the Modified Aggression Scale Bullying subscale) was included after the baseline survey as we elected to add a measure of bullying perpetration as well as one of victimisation.
- We included administrative documents (e.g. minutes, attendance sheets, training satisfaction feedback) in our assessment of trial arm fidelity to provide us with a wider range of quantifiable data.

METHODS

The protocol was amended during the trial to refine the methods used. All amendments were approved by the independent study steering committee and the funder of the trial (NIHR). The only change to trial outcomes was the addition of a measure of bullying perpetration (secondary outcome). All refinements were completed before the 36-month surveys were collected and before any trial analyses were conducted.

Participants

The INCLUSIVE trial was a universal intervention aimed at all 11- to 16-year-olds in participating secondary schools in England. Although the intervention was expected to have effects on the whole school, our study population consisted of students at the end of year 7 (aged 11–12 years) at baseline and at the end of year 10 at the 36-month follow-up (aged 14–15 years), as well as all school teaching and teaching assistant staff. This year group should have experienced the intervention for 3 years, including the classroom curriculum for years 8–10. All students in the school in that year and all teaching staff were surveyed at each time point, not only those who participated at baseline.

Eligibility criteria for participants

Participant eligibility was assessed at the school level. Eligible schools were:

- Secondary schools within the state education system (including community, academy or free schools, and mixed- or single-sex schools) in south-east England. We took the widest definition of a 'state school' and excluded only private schools, schools exclusively for those with learning disabilities and pupil referral units. The last two were excluded as the INCLUSIVE study was unlikely to be appropriate for their populations.
- Schools whose most recent school quality rating by Ofsted was 'requires improvement'/'satisfactory' or better. Schools with an 'inadequate'/'poor' rating were excluded, as these schools would be subject to special measures that were likely to impede LT delivery.

Eligible schools were approached initially by letter and e-mail, with a telephone follow-up, complying with good practice and research governance for undertaking studies within the education system.

As the intervention was delivered at the whole-school level, there were no specific eligibility criteria for students, although parents who did not want their child to participate in the surveys were able to opt out on behalf of their child.

Settings and locations

Schools were recruited between March and June 2014 from secondary schools in Greater London and the surrounding counties (Surrey, Kent, Essex, Hertfordshire, Buckinghamshire and Berkshire) that had a maximum travel time of 1 hour from the study centres in London. To aid recruitment, we partnered with existing school networks, such as the University College London (UCL) Partners Schools Network and the Institute of Education Teaching Schools, and schools that are part of our collaborating schools network, Challenge Partners. We approached approximately 500 eligible schools, initially by letter and e-mail and with a telephone follow-up.

The 40 participating schools did not differ from 450 non-recruited schools in terms of school size, population, deprivation, student attainment or value-added education. However, participating schools were more likely to have an Ofsted rating of good or outstanding (see *Appendix 1, Table 26*).

Intervention

The INCLUSIVE trial involved 2 years of externally facilitated intervention and a final year without external facilitation. The LT intervention was intended principally to augment, rather than to replace, existing activities (e.g. training and curricula) in intervention schools. However, it was intended to replace existing non-restorative disciplinary school policies and practices when the AG deemed restorative approaches more appropriate. The intervention logic model is shown in *Appendix 1* (see *Figure 6*).

Below we describe the intervention informed by the TIDieR (Template for Intervention Description and Replication) checklist for better intervention reporting.⁹⁴ Fidelity assessment is described under process evaluation, the product of which is presented in our results.

Learning Together

Theory of change

Informed by Markham and Aveyard's⁵⁹ theory of human functioning and school organisation, the intervention's theory of change suggests that for young people to choose healthy behaviours over risky behaviours, such as bullying, aggression or substance use, they must possess the autonomy, motivation and reasoning ability to make informed decisions. These capacities and goals are theorised as facilitated by increased engagement with education (the school's 'instructional order') and connection to the school community (the school's 'regulatory order'). It is theorised that schools can increase such engagement by improving relationships between students and teachers, between different groups of student and between academic education and broader student development, as well as by reorienting learning and teaching, discipline, social support, and school management and organisation so that these centre on student needs and view conflict as an opportunity for learning. The intervention aims to strengthen relationships between and among staff and students through the use of primary (preventing conflict) and secondary (preventing the escalation of conflict) forms of restorative practice, and by enabling staff and students to work together on an AG co-ordinating intervention delivery in each school (see *Appendix 1*, *Figure 6*). AGs also aim to enable student participation in decision-making. Restorative practice aims to increase students' active participation in discipline systems. A social and emotional skills curriculum delivered in classrooms aims to promote student autonomy and reasoning ability, and to facilitate student participation both in AGs and in restorative practice.

Materials

Schools allocated to receive the intervention were provided with various resources. School staff were offered training in restorative practices, with participants given written summaries of the material covered in training. Schools were provided with a manual to guide the convening and running of an AG. For the first 2 years of the intervention, schools were provided with an external facilitator for the AG. Schools were sent a report on student needs detailing findings from a survey of students aged 11–12 years about their attitudes to and experiences of school, and their experiences of bullying, aggression and other risk behaviours (see *Appendix 3*). Schools were provided with written lesson plans and slides to guide delivery of a classroom-based social and emotional skills curriculum.

Procedures

Training was given to all staff, and in-depth training was given to selected staff, including training in formal 'conferencing' to deal with more serious incidents.

Action groups

Action groups were required to include, at a minimum, six students from the intervention classes (year 7 at the start of the intervention) and six members of staff, including at least one from the senior management team and one member of each of the teaching, pastoral and support staff. Having a member from specialist health staff, such as the school nurse and/or local child and adolescent mental health services staff, was desirable but optional. The AG was required to meet at least six times per school year (approximately once every half-term), and was tasked with developing action plans to co-ordinate the delivery of the intervention outputs:

- reviewing and revising *school rules and policies* relating to discipline, behaviour management and staff-student communication to incorporate restorative principles
- implementing *restorative practices* throughout the school to prevent and respond to bullying and aggression
- additional *tailored actions* to address local priorities
- delivering the *social and emotional skills curriculum* for years 8–10.

The facilitator ensured that meetings were scheduled, and attended these to ensure that the meetings were participative and focused on deciding and implementing actions. These actions were informed by findings about their students' experiences of bullying, aggression and the school environment from our baseline survey (conducted before randomisation) and from a 12-month survey of students at the end of year 8 (aged 12–13 years) in intervention schools only, as well as from the 24-month trial survey. In year 3, facilitation was to be internally led by the AG's chairperson, usually a SLT member or another experienced staff member. External facilitation in the first 2 years was theorised to be important to enable schools to initiate changes and particularly to empower students to participate in decisions. In year 3, schools were expected to facilitate implementation internally so that the trial could assess whether or not the intervention could be sustained by schools in the absence of an external facilitator.

Social and emotional learning curriculum

Schools delivered classroom-based social and emotional skills education in 'stand-alone' lessons, for example 'personal, social and health education' (PSHE) lessons, and/or integrated it into tutor time or various subject lessons (e.g. English) to students in the trial cohort as they moved through years 8–10 (aged 12–15 years). They received 5–10 hours of teaching and learning per year on restorative practices, relationships, and social and emotional skills based on the Gatehouse Project curriculum.⁶⁰

Schools selected modules for each year from 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.

Restorative practice

Primary restorative practices delivered in schools in all three years involved staff using restorative language (the respectful use of language to challenge or support behaviour in a manner that preserves or enhances the relationship) and circle time (classes coming together to discuss their feelings and air any problems so that these may be addressed before they escalate) underpinned by supportive school rules and policies and the social and emotional skills curriculum. Circle time takes place in an informal setting, and is overseen by a member of staff; it provides an opportunity for a class to discuss their relationships in the open. It could be undertaken during registration periods or other lessons and aims to maintain good relationships, or deal with specific problems, as well as making the whole class aware of the issues and responses ongoing.

Secondary restorative practices involve some staff implementing restorative conferences (the parties to a conflict being invited to a facilitated face-to-face meeting to discuss the incident and its impact on the victim and for the perpetrator to take responsibility for their actions and avoid further harms). Conferencing was suggested for use in more serious incidents; this is a more 'one-on-one' practice of restorative justice, which

brings together relevant staff, students, parents and, where necessary, external agencies to discuss ongoing issues between students.

Training

Staff training was implemented to ensure that teachers understood the necessary skills to engage in restorative practice. Training was provided by trainers accredited by the UK's Restorative Justice Council. Each school had its own named facilitator, who was a freelance consultant with experience of school leadership or organisational change, co-ordinated by a lead facilitator who trained them in the intervention theory and methods. AGs comprised at least six staff members (including one member of the school's SLT and one member of the school's teaching, student support and administrative staff) and at least six students from each school, led by a member of the school's SLT with support from the external facilitator in the first two years of the intervention but not in the third year. All of these staff attended the all-staff training and some attended the in-depth training. Staff who received basic training in restorative practice implemented this in the form of the use of restorative language and circle time. In addition, 5–10 staff members at each school who received in-depth training in restorative practice implemented this in the form of restorative conferences. The curriculum was delivered by teachers who specialised either in PSHE or in other subjects guided by lesson plans.

Modes of delivery

All intervention components were delivered face to face.

Location

All components were delivered on school premises.

Dose

Training occurred in the first year of intervention, comprising half a day for all staff plus in-depth 3-day training for 5–10 staff members at each school. AGs met six times per year in all 3 years. Restorative practices were delivered as frequently as required in each school. In their curriculum, students received 5–10 hours of teaching per year.

Planned adaptations

The intervention enabled local tailoring, informed by the needs survey and other local data sources. AGs ensured that implementation in their school was appropriate to local needs as identified by members and the survey of student needs. This included ensuring that revisions to policies and rules built on existing work, deciding which curriculum modules to deliver in each year, and implementing locally decided actions aiming to improve relationships and student participation (e.g. cascading restorative practice training to staff who had not attended or to student peer mentors).

Unplanned modifications

There were no unplanned modifications.

Comparator: control schools

Schools randomised to the control group continued with their normal practice and received no additional input. The sample of schools was spread over a wide geographical area and there were no cases in which intervention and control schools were near one another. Head teachers and a small number of staff in control schools were aware that the school was participating in a trial that was described as the INCLUSIVE trial. These individuals were not informed that the name of the intervention was LT and were not informed about the detailed contents of the intervention during recruitment. It is therefore unlikely that schools in the intervention and control arms would have shared information about the intervention.

Assessment and follow-up

Assessment of effectiveness

Student primary and secondary outcomes were assessed at 36 months, at the end of year 10 (when the students were aged 14–15 years), with a baseline survey having been undertaken at the end of year 7 (when the students were aged 11–12 years). Staff secondary outcomes were also assessed at 36 months. Additional student and staff surveys were conducted at 24 months to assess intervention process and intermediate outcomes to be used in the mediation analysis. Student surveys were conducted in exam conditions in schools, maximising privacy. The questionnaires used to collect these data can be found in full in the supplementary material (see *Report Supplementary Material 1*).

All students in the school in that year and all teachers and teaching assistant staff were surveyed at each time point, not only those who participated at baseline. Paper-based questionnaires were completed confidentially in a 45-minute class session devoted to that purpose. Field workers supervised the class as they completed the questionnaire, with the teacher present (for disciplinary purposes) but unable to see the questionnaires. The field workers assisted students with questions that they did not understand and ensured that the students completed as much of the questionnaire as possible. Students with mild learning difficulties or with limited command of written English were supported by field workers to complete the questionnaires.

We assessed the potential for measurement error and bias by asking the students completing surveys if their responses to questionnaires were completely truthful. We asked students in intervention schools involved in qualitative interviews whether or not their reporting (as opposed to their experience) of bullying and aggression might have been affected by the intervention.

Staff data were often collected on the same days as student data. However, owing to the busy nature of their work, staff questionnaires were often left at the school to be done in private time, and then mailed back to the study team. Staff were allowed to fill in their questionnaires in the staff breakroom, or to take them home to fill in.

Data management

The study centre received class lists for each school in advance of each survey. Participants were allocated a unique identifier (ID) prior to each survey and this ID was recorded on the questionnaire. All questionnaires were anonymous. Questionnaires were completed in classrooms and completed questionnaires were collected in schools on the day of the survey. If a participant was not in school on the day of the survey, a questionnaire was left at the school for them to complete later and was returned to the study centre by post. Completed questionnaires were transported from the school by study personnel to the study centre, where they were stored in a locked room.

Questionnaires were then securely transported for data entry by a third party, where they were double-entered into a database by trained personnel. Each questionnaire was checked at the time of data entry for any handwritten comments. Questionnaires with any additional text, regardless of content, were scanned, and password-protected scans were sent to the study team for safety reporting assessment. Electronic data generated from data entry were transferred via password-protected secure FTP and stored on secure servers at the London School of Hygiene & Tropical Medicine (LSHTM).

Following data entry, questionnaires were securely transported to the LSHTM Clinical Trial Unit (CTU) for archiving. An inventory of all questionnaires was maintained by LSHTM CTU.

Electronic data generated from data entry were transferred via password-protected secure FTP and stored on secure servers at LSHTM. Relevant trial documentation will be kept for a minimum of 15 years after study completion.

Ethics arrangements

The trial was approved by the UCL Ethics Committee (reference 5248/001). Ethics arrangements were informed by recent guidance on ethical issues in cluster RCTs.

Informing participants

Details of the research, including the possible benefits and risks, were provided to schools through written information and personal meetings and were provided to student participants through age-appropriate written information.

Consent

Written, informed consent was obtained at school level (chairperson of governors; head teacher) for random allocation and for the intervention, and at the individual student, staff and intervention facilitator level for data collection. For students, written age-appropriate information sheets were provided in class 2–4 weeks before the baseline survey, together with oral explanation by teachers. Written consent was required from all participating young people, which was collected immediately before conducting the baseline survey. Young people were also asked to take home written information sheets for their parents. Parents who did not want their child to participate were asked to notify this opt-out in writing using a prepared form. This 'opt-out' consent is standard practice in trials in secondary schools and was used in our pilot study, proving acceptable to schools, young people and families. Only < 1% of parents exercised an opt-out.

Information sheets and consent forms for student surveys were identical in intervention and control schools and did not refer to the intervention. Parents were informed about the study and could withdraw their children from research activities.

Duty of care and confidentiality

The researchers were experienced with the specific ethical issues involved in undertaking research with young people and other vulnerable participants. All work was carried out in accordance with the requirements of the Data Protection Act 1998.⁹⁵ Data storage and IT (information technology) systems were secure. All information remained confidential within the research team, except when child protection issues were raised. We consulted with a child protection social worker to define the issues that would prompt an exemption. The chief investigator, Russell Viner, as a paediatrician with training in child safeguarding, oversaw actions when safeguarding concerns were raised, and sought further advice, when necessary, from appropriate authorities. We followed Economic and Social Research Council ethics guidance and sought research ethics approval from the appropriate bodies. We also sought policy approval from local authorities related to each participating school.

Ethics review

Approval for the study was sought and obtained from the research ethics committees of the two lead universities, LSHTM and UCL. Our trial complied with the Economic and Social Research Council research ethics framework.

Assessment of harms

There were no anticipated risks to participants or to schools. However, as with all interventions, there may have been unanticipated risks. Harms were assessed by examining outcomes at 24 and 36 months. An independent Data Monitoring Committee (DMC) examined any potential harms at 24 months. If any major harms were detected, the DMC was to inform the Trial Steering Committee (TSC), which would decide what action should be taken.

Outcomes

Primary outcomes

Outcomes were collected by a research team (led by RV) independent of the intervention team. The primary outcomes were self-reported experience of bullying victimisation and perpetration of aggression measured at 36 months. The questionnaires used to collect these data can be found in full in *Report Supplementary Material 1*.

- Bullying victimisation was assessed with the GBS, a 12-item validated⁹⁶ self-report measure of being the subject of teasing, name-calling or rumours, being left out of things, and receiving physical threats or actual violence from other students within the previous 3 months. The questions and responses were worded to ensure that these assessed bullying occurring either face to face or online. Students reported the frequency of and upset related to each experience. Items were summed to make a total bullying score (higher scores represented more frequent upsetting bullying, with a maximum score of 3).
- Perpetration of aggressive behaviour was measured using the ESYTC school misbehaviour subscale, a 13-item scale measuring self-reported aggression towards students and teachers. Each item was coded as occurring hardly ever or never, less than once per week, at least once per week, or most days. Items were summed to provide a total score, with higher scores indicating greater aggressive behaviour (maximum score of 39).⁹⁷

Secondary outcomes

The GBS and ESYTC outcomes were measured at 24 months as secondary outcomes. In addition, we measured the following at 24 and 36 months.

Student-level self-report outcomes

These were measured through student survey self-reports.

Quality of life

The Paediatric Quality of Life Inventory (PedsQL) version 4.0 was used to assess overall QoL. The 30-item PedsQL has been shown to be a reliable and valid measure of QoL in normative adolescent populations.⁹⁸ It consists of 30 items representing five functional domains – physical, emotional, social, school and well-being – and yields a total QoL score, two summary scores for ‘physical health’ and ‘psychosocial health’, and three subscale scores for ‘emotional’, ‘social’ and ‘school’ functioning.

Health-related quality of life

The CHU9D, a validated, age-appropriate measure of students’ health-related quality of life (HRQoL),⁹⁹ was used to inform the economic evaluation.

Psychological function and well-being

The Strengths and Difficulties Questionnaire (SDQ)¹⁰⁰ is a brief screening instrument for detecting behavioural, emotional and peer problems and prosocial strengths in children and adolescents. It is brief, quick to complete and validated in national UK samples.

The Short Warwick–Edinburgh Mental Well-Being Scale (SWEMWBS)¹⁰¹ is a seven-item scale designed to capture a broad concept of positive emotional well-being, including psychological functioning, cognitive-evaluative dimensions, and affective-emotional aspects, with a total 'Well-Being Index' generated.

Risk behaviours

Substance use

Smoking, alcohol use and illicit drug use were assessed. Validated age-appropriate questions were taken from national surveys¹⁰² and/or previous trials were used in order to assess smoking (smoking in previous week; ever smoked regularly), alcohol use (use in previous week; number of times really drunk; binge drinking) and illicit-drug use (last month; lifetime use).

Sexual risk behaviours

Age of sexual debut and use of contraception at first sexual encounter were examined with measures used in the RIPPLE trial.¹⁰³ We consulted with schools about the acceptability of asking these questions at follow-up (year 10).

The Modified Aggression Scale Bullying subscale (Cronbach's alpha = 0.83)

This measure came from the Centers for Disease Control and Prevention's guidance document on bullying measures.¹⁰⁴ It includes a five-item scale assessing the level of bullying perpetration (last 3 months).

Use of NHS services

Self-report use of primary care, accident and emergency, or other service in the previous 12 months.

Contact with police

This was assessed using self-report of whether the young person had been stopped, told off or picked up by the police in the previous 12 months.

Demographic information

Sex and ethnicity

Student self-report.

Socioeconomic status: the Family Affluence Scale

The Family Affluence Scale (FAS) was developed specifically to describe the SES of young people.¹⁰⁵ A composite FAS score was calculated for each student based on his or her responses to four items relating to family car ownership, children having their own bedroom, the number of computers at home, and the number of holidays taken in the past 12 months. For our analyses, scores were collapsed into tertiles of low (score of 0, 1 or 2), medium (score of 3, 4 or 5) and high (score of 6, 7, 8 or 9) family affluence.

Data collected directly from schools

We planned to collect some data directly from schools for each year of the study using data routinely collected by schools:

- school attendance data, expressed as number of half-days absent over the previous year
- school rates of temporary and permanent exclusions
- staff attendance, expressed as number of half-days absent, for which staff members' informed consent to access was sought.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Bonell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. 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.

METHODS

Individual staff-level outcomes

The following secondary outcomes were measured using survey self-reports from teachers and teaching assistants (the questionnaires used to collect these data can be found in full in *Report Supplementary Material 1*):

- *staff QoL*, measured using the Short Form questionnaire-12 items (SF-12),¹⁰⁶ a brief, well-validated measure of adult health-related QoL
- *staff stress and burnout*, measured using Maslach Burnout Inventory,¹⁰⁷ an established scale that uses a three-dimensional description of exhaustion, cynicism and inefficacy.

School-level outcomes

Value-added education score

School median value-added scores were obtained from UK official statistics¹⁰⁸ relating to the progress students make between education Key Stage 2 or 3 (aged 7–14 years) and Key Stage 4 (aged 14–16 years). The value-added score for each student was calculated using the difference between their own output point score (end of Key Stage 4) and the median output point score achieved by others with the same or similar starting point (Key Stage 2 or 3), or input point score.¹⁰⁸ Schools that neither added nor subtracted educational value were given a score of 1000, with positive value added (> 1000) indicating a school where students on average made educational progress and negative value added (< 1000) indicating the reverse.

School size

The total number of students in the school, as identified from the school and college performance tables.¹⁰⁸

School-level deprivation

This was assessed using two variables:

1. *Proportions of students eligible for free school meals (FSM)* – the percentage of students eligible for FSM at each school at any time during the past 6 years is an accepted summary measure of school deprivation. Data were publicly accessible from the Department for Education.¹⁰⁸
2. *The Income Deprivation Affecting Children Index (IDACI) score* – a small-area indicator of deprivation specifically affecting children (< 16 years of age), which represents the proportion of children in a postcode who live in low-income households.¹⁰⁹ The value for each school is derived from the school's postcode and thus represents the deprivation level of the school's local area, rather than the school itself.

Office for Standards in Education, Children's Services and Skills rating

Ofsted ratings¹⁰⁸ are government inspectorate ratings of the quality of teaching, leadership and management, achievement of students, and behaviour and safety of students of a school. Schools are classified as 1 = 'outstanding', 2 = 'good', 3 = 'requires improvement' or 4 = 'inadequate'. Owing to eligibility criteria for the INCLUSIVE study, only schools rated from 1 to 3 were included in the sample.

School sex mix

Mixed- or single-sex schools were identified from the school and college performance tables.¹⁰⁸

School type

Our sample comprised five types of school, categorised by the source of school funding. These were (1) converter academy mainstream ($n = 18$), funded directly from central government; (2) sponsor-led academy ($n = 6$), which has an independent business or charitable sponsor but is funded directly from central government; (3) foundation ($n = 6$), where the school owns its premises but is funded by the local authority; (4) community ($n = 5$), where premises and funding are provided by local authorities; and (5) voluntary-aided ($n = 4$), where the premises are owned by a charity (e.g. a religious foundation) but funding is at least partly from the local authority.¹⁰⁸

Process evaluation

The process evaluation examined intervention implementation and receipt, and possible causal pathways, in order to facilitate interpretation of the outcome data and enable refinement of the intervention logic model. Informed by existing frameworks,^{110,111} data were collected to examine the following.

Trial context

We assessed the context of schools in the intervention and control arms, such as discipline systems, staff training, social and emotional skills curricula, and student participation in decision-making. This drew on interviews with intervention facilitators and trainers, members of AGs in intervention schools, staff on school SLTs, and other staff in the intervention and control arms; and focus group discussions with students and staff in schools selected as case studies.

Trial arm fidelity

We assessed the fidelity of intervention delivery by school and facilitator. In addition to the above sources, we drew on follow-up surveys with staff and students; structured researcher observation of AG meetings and staff training; surveys of adults leading curriculum implementation and implementing restorative practice; interviews with adults delivering the curriculum; structured diaries kept by facilitators of AG meetings and by trainers of all-staff training; and administrative documents such as minutes and attendance sheets.

Overall fidelity in the externally facilitated first 2 years was scored out of a possible eight points for each school, as assessed by researchers, based on whether or not (1) at least five staff attended in-depth training (indicated in training registers); (2) each year all six AGs were convened (indicated in minutes); (3) policies and rules were reviewed (indicated in minutes); (4) locally decided actions were implemented (indicated in minutes); (5) AGs were perceived to have had a good or very good range of members (indicated in survey of AG members); (6) AGs were perceived to have been well or very well led (indicated in survey of AG members); (7) schools delivered at least 5 hours and/or at least two modules each year (indicated in survey of lesson deliverers); and (8) at least 85% of staff reported that if there was trouble at the school, staff responded by talking to those involved to help them get on better (indicated in staff survey). Overall intervention fidelity in the internally facilitated third year of the intervention was assessed using a narrower range of data, as the research team's access to schools was expected to be reduced. Schools were scored out of a possible 4 points, on the basis of interviews with AG members to assess whether or not (1) all six AGs were convened and (2) locally decided actions were implemented; surveys and interviews with curriculum deliverers to assess whether or not (3) schools delivered at least 5 hours and/or at least two modules; and staff survey to assess whether or not (4) at least 85% of staff reported that, if there was trouble at the school, staff responded by talking to those involved to help them get on better.

Participation, reach and dose

We assessed the extent to which students and staff were aware of or involved in intervention delivery. This drew on surveys of AG members as well as follow-up surveys of students and staff and focus group discussions and interviews with students and staff.

Reception and responsiveness

We assessed the experiences of participation in INCLUSIVE, and in school environments shaped by this, to assess acceptability and barriers to facilitators. This drew on satisfaction surveys completed by staff attending in-depth training and of AG members; interviews with AG members, school staff and SLT, and students participating in restorative practice; and focus group discussions with students.

Intermediate outcomes

To assess possible intervention causal pathways, to examine whether or not these might mediate intervention effects, and to assess and refine our logic model, we used the Beyond Blue School Climate Questionnaire 28-item scale to measure students' perceptions of the school climate, including supportive

METHODS

teacher relationships, sense of belonging, participative school environment, and student commitment to academic values,¹¹² and the Young People's Development Programme (YPDP) single-item measure of involvement with antischool peer groups.¹¹³

Data collection

The data collection for the process evaluation was prospective and used mixed methods. The process evaluation was designed to explore the same intervention processes using different forms of data, and to compare findings between these. Purposive sampling was used for interviews to involve participants with diversity in terms of characteristics thought important for exploring implementation, and to explore diverse accounts and identify different themes.¹¹⁴ Samples were large enough to generate diversity, but small enough to keep the analysis manageable. Informed by Medical Research Council (MRC) guidance on process evaluation,¹¹⁵ we sampled participants on the basis of characteristics likely to be associated with diverse perspectives on the intervention and its implementation within schools. We balanced sampling of some participants in all schools to develop an overview of delivery with larger samples in a small number of case study schools to explore processes in more depth. When discrepancies or gaps in the data emerged, these were explored in the next applicable data collection round.

Data collected in all schools

Diaries completed by trainers

Individuals providing the all-staff training in restorative practice were asked to complete a diary for each session delivered, informed by the tool used in the pilot study. They were asked to rate the extent to which they covered topics/materials as intended, and the materials and activities [e.g. Microsoft PowerPoint® (Microsoft Corporation, Redmond, WA, USA) slides, small-group or paired activities] used. Trainers then sent the diaries to the research team.

Observations of training

Researchers aimed to conduct structured non-participant observations of training so that all schools could be observed at least once. Observation guides included what topics were covered and what activities were used, and were informed by the same tools as used in the pilot study.

Satisfaction survey for in-depth training

An anonymous satisfaction survey was given out to the 5–10 staff members from each school who attended the in-depth training on restorative practice. Informed by the same tool as used in the pilot study, questions assessed whether or not participants felt that the training was useful; whether or not they felt confident about putting into practice the skills learnt; whether or not they would recommend the training; and overall how they rated the training provided. Participants placed the questionnaires in an envelope, which the trainers collected and sent to the researchers.

Interviews with trainers

Semistructured interviews with the two trainers were conducted by telephone in year 1 and lasted between 30 and 45 minutes. These aimed to explore the trainers' views on participant responsiveness, any adaptations and deviations made, and barriers to and facilitators of delivery.

Diaries kept by facilitators of action groups

Facilitators were asked to complete a diary for each AG meeting they attended. Informed by the same tool as used in the pilot study, these explored general meeting information such as duration, date, number of attendees, chairperson and minute-taker names; members' roles, year group and gender; how and what data were used to inform setting up school actions; priorities set by the school and actions stemming from these; actions concerning the revision of school rules and school policies; identification of which modules of the curriculum were to be implemented and how this was decided; and the participation of AG members. This information was then sent to the research team.

Minutes of action groups

Facilitators were asked to collect minutes from each AG meeting and send these to the researchers. These were used to triangulate the validity of facilitator diary forms.

Observations of action groups

Researchers aimed to conduct structured non-participant observations of AGs in 10 randomly selected schools for each year of intervention. Observation guides focused on the same areas as diaries and were informed by the same tools used in the pilot study.

Survey of action group members

An anonymous survey was handed out to all members of AGs by facilitators at the end of each year of the intervention. Informed by the same tool as used in the pilot study, this explored its acceptability, functioning and composition. It asked questions, for example, on the diversity of staff and students on the AG, how well led they considered the group to be and how empowered members felt to make decisions, using an existing scale.¹¹⁶ Participants placed the questionnaires in an envelope, which the facilitators collected and sent to the researchers.

Interviews with facilitators of action groups

Semistructured telephone interviews with facilitators ($n = 6$) were conducted in years 1 and 2 and lasted between 45 and 90 minutes. These aimed to explore views on school culture, responsiveness and priorities; any adaptations and deviations made; and barriers to and facilitators of delivery.

Interviews with members of action groups

We aimed to interview two members of each school's AG per year. A member of the evaluation team contacted the member of staff tasked with co-ordinating the intervention at each school and asked them to identify two AG members (staff or student) to be interviewed. Identified staff participants were then contacted by e-mail and/or telephone to schedule an interview, which occurred either in person (if possible) in a private room on school premises or over the telephone and lasted between 30 and 60 minutes. Interviews with students were arranged via staff and were always conducted in the school. Interviews were semistructured and explored views on the acceptability of facilitators and the intervention; barriers to and facilitators of AG meetings and how they might be improved; the extent to which actions arising from meetings were implemented in the school; and their perceived impact on the school environment.

Survey of school staff leading implementation of the curriculum

This survey was sent annually to be completed by the teacher in each intervention school who was acting as the LT social and emotional skills co-ordinator. The research team sent the surveys by e-mail termly in the first and second years of the study, and in the final term of the third year. Staff were asked to complete the survey and return it to the research team by e-mail. Informed by the same tool as used in the pilot study, the survey covered what units and lessons were delivered, when, in which subjects, for how many hours, and what intervention materials (e.g. PowerPoint slides, lesson plans), if any, were used to deliver the content.

Interviews with school staff delivering the curriculum

The research team aimed to arrange semistructured interviews in each year of the intervention with the staff member responsible for delivery of the social and emotional curriculum. The curriculum co-ordinator at each school was contacted by e-mail and/or telephone and asked to identify a member of staff delivering the curriculum to participate. Interviews were carried out over the telephone or face to face in a private office on school premises. The interviews gathered views on the fidelity, reach and acceptability of the curriculum; which materials were used; delivery methods; student responsiveness; and contextual barriers to or facilitators of delivery.

Survey of staff implementing restorative practice

This survey aimed to examine the extent to which staff who attended in-depth training in restorative practices were implementing such practices in school. The survey was initially to be completed termly, but for the third year this was changed so that staff were asked to complete it only in the summer term. Staff who had attended the in-depth training were sent an e-mail inviting them to complete the survey. The survey assessed delivery of the use of affective language, circle time, mediation, restorative conferencing, family group conferencing and community conferencing.

Interviews with other school staff

Semistructured interviews were sought across intervention and control schools with one staff member from schools' SLT ($n = 40$) and two teaching staff ($n = 80$) at the beginning of year 1. Each school's member of staff liaising with the research team was contacted by e-mail and/or telephone and asked to identify three staff members to participate. A member of the research team then contacted these staff by e-mail to schedule a telephone interview. Interviews explored the context of schools, including their policies and practices relating to behaviour management, social and emotional skills education, staff training, and student participation in decisions. In year 3, such interviews were sought with one SLT member in all intervention and control schools. Individuals in control schools were interviewed in the autumn term and in intervention schools in the summer term.

Data collected in case study schools

Six schools in the intervention arm were selected as case studies in order to gather in-depth qualitative data on intervention processes and school context. In these schools, we aimed to conduct focus groups with staff and students, as well as interviews with students participating in restorative practices. To encompass diverse schools, schools were purposively sampled in terms of diversity in relation to the percentage of FSM (above and below national average in 2012 for secondary schools, 16.3%), type of school, the facilitator assigned to the school, and the extent to which the school was responsive (highly responsive, somewhat responsive, poorly responsive) to intervention activities, as rated by the intervention facilitators 3 months into the intervention start date.

Focus groups with staff

In each year of the intervention, we aimed to conduct one focus group with staff in each case study school, each involving four to six members of staff. Staff were purposively selected and invited to participate by the staff member liaising with the research team to include diversity according to degree of participation in the intervention and role within the school (including senior leaders, pastoral staff and classroom teachers). Focus group discussions aimed to explore school culture and ethos, views about the delivery and impacts of the intervention, how restorative practices were applied, and barriers to and facilitators of their use. Focus groups were conducted in private offices on school premises facilitated by one researcher.

Focus groups with students

Each year, we conducted two focus groups with students in each case study school, comprising 4–12 students each: one with students directly involved in intervention activities (e.g. AGs) and one with students not involved in such activities. Students were purposively selected and invited to participate by the staff member in each school liaising with the research team, such that they reflected the diversity of the school in terms of boys and girls, different ethnic groups, and varying degrees of educational engagement. Focus groups were conducted in private offices on school premises facilitated by one researcher.

Interviews with students involved in restorative practices

We aimed to conduct semistructured interviews with two students at each case study school that had been involved in a restorative practice. The staff member at each school liaising with the research team was asked to invite students to participate, recruiting either one boy and one girl or a perpetrator and victim in the same instance of bullying, where possible. Interviews aimed to understand the processes of restorative practice and assess the acceptability of the approach. Restorative interviews were not limited to

cases of bullying or aggression but also included classroom misbehaviour or friendship challenges. These interviews were conducted in private offices on school premises facilitated by one researcher.

Interviews and focus groups were audio-recorded and transcribed.

Ethics

All data were collected with research participants' informed consent. Student and staff participants received written information beforehand and were given the opportunity to ask questions to a member of the research team. They were then invited to give signed consent and reminded that they could skip any questions and/or end the interview at any time. For telephone interviews with staff, an e-mail from participants indicating consent was generally used instead of a written signature. Before the interview began, they were read a statement relating their rights, how the data would be used, and information about anonymity and confidentiality. They were then asked to give verbal consent, which was recorded. All data collected were stored on password-protected drives within separate password-protected folders. However, had any research participants reported that they had been involved in or were at risk of sexual or physical abuse that the school was not already aware of, the research team would have liaised with the safeguarding lead for the school in question, breaking anonymity. Participants were made aware of this policy as part of the consent procedures. No such reports were made.

Economic evaluation

Economic evaluation of the intervention also took place. In accordance with NIHR guidelines, the methodology and results for the economic evaluation are reported separately in *Chapter 4*.

Changes to trial outcomes

The following deviations from this plan occurred during data collection.

We were unable to collect school-level data on individual student and staff attendance and school rates of temporary and permanent exclusions, despite multiple attempts to contact schools and obtain these data after the intervention was completed. Small numbers of data on school-level exclusions were provided but these were not sufficient for analysis. In response to requests for data, schools either did not respond or notified us that this was a burden they were not prepared to undertake now that the trial had finished. Data were not available similarly across intervention and control schools. A further offer of money towards school staff funds (May 2018; see above) did not motivate schools to provide these data.

In discussion with NIHR, we came up with the following mitigation plan.

- School attendance data: these data are currently planned to be collected by Manchester University in a planned follow-up to this study in late 2019. These are therefore not reported here.
- School rates of temporary and permanent exclusions: these data are published at school level by the Department for Education, albeit with a significant delay. We plan to examine these in the future when available, although the data are not reported here.
- Staff attendance: individual staff attendance levels are not available through other means and these analyses cannot be undertaken.

Amendments to and deviations from the protocol (process evaluation)

All planned amendments were approved by the study steering committee (5 October 2015). We clarified that trainer diaries would be used to examine all-staff training and that a satisfaction survey would instead be used to assess in-depth training. A survey of all-staff training was judged impossible because of the short time available for the training.

METHODS

We reduced the number of observations of AG meetings from 20 to 10 per year. The rationale was that this was sufficient to explore AGs alongside all the other forms of data on these, and that any greater data collection would be very onerous for schools.

We dropped plans to observe one curriculum session per school per year, replacing this with surveys and interviews examining curriculum implementation, because schools advised us that scheduling observations would be challenging.

We introduced surveys examining implementation of restorative practice to ensure that we examined this important area of implementation. We clarified that we would undertake telephone interviews with one SLT member and two other staff per school in year 1 only and then undertake telephone interviews with one SLT member per school in year 3. The rationale was that this was sufficient to understand the context of trial schools and that any greater data collection would be very time-intensive and onerous for schools.

We reduced the number of case study schools from eight to six, with these being drawn from intervention schools only, rather than from intervention and control schools. The rationale was that six schools was more appropriate for in-depth research and that in-depth case studies in control schools would be informative. We purposively rather than randomly sampled case study schools as this would be more likely to be effective in ensuring the contextual diversity needed for qualitative research.

There were also several deviations from the protocol. We said that we would observe some training in all 20 intervention schools, but a research team error meant that we actually observed a random sample of 10 schools. We said that we would survey staff co-ordinating curriculum delivery annually but in years 1 and 2 we did this termly before reverting to an annual survey in year 3. Our protocol said that we would include as one of our intermediate outcomes a measure of student anti-school actions using the ESYTC self-reported delinquency subscale but this was a drafting error as this measure instead was part of one of our primary outcome measures.

Sample size

We proposed to recruit sufficient participants to detect a difference between groups of 0.25 standard deviation (SD) with 90% power and a 5% level of significance. This is considered to represent a moderate size of effect and is in line with the effect sizes seen in the literature.¹¹⁷

The average English school has approximately 150 students per year group of students, although this varies across schools. Using a conservative intraclass correlation coefficient (ICC) of 0.04¹¹⁸ and an estimate of 150 students per school, a trial involving 20 schools per arm would have provided 90% power to identify an effect size of 0.25 SD with a 5% significance level. If two schools per arm (i.e. 10%) were to be lost to follow-up over the course of the trial, we would still have 80% power to detect an effect size of 0.23. The total student sample size was estimated to be between 4000 (assuming 100 students per school) and 6000 (assuming 150 students per school). As we were surveying all young people in the relevant school year at each follow-up, this sample was likely to remain similar across the study. The total student sample was therefore planned to be approximately 6000.

No stopping guidelines were put in place for this study; any and all schools and students who agreed to take part were included.

Randomisation

Sequence generation

Schools were randomly allocated in a 1 : 1 ratio to the intervention or control arm immediately after their students and staff had completed the baseline data collection. To promote baseline similarity, randomisation was stratified by key school-level determinants of violence¹¹⁹ data obtained from the Department for Education (www.gov.uk/school-performance-tables, accessed 17 September 2018):

- single-sex versus mixed-sex school
- school-level deprivation, as measured by the percentage of students eligible for FSM (low/moderate 0–23% and high > 23%, with 23% being the median for England)
- school 'best eight value added' in General Certificate of Secondary Education (GCSE) exams (above and below the median for England of 1000), a school-level measure of students' attainment in public exams accounting for their attainment on entry to the school.

Data were obtained from the Department for Education (www.gov.uk/school-performance-tables, accessed 17 September 2018).

Schools were allocated randomly within each of these eight strata.

Sequence allocation was generated by the clinical trials unit (CTU) at LSHTM using Stata's® 'ralloc' command (version 15; StataCorp, College Station, TX, USA) and was concealed from schools and the wider evaluation and intervention teams. Allocation was communicated from the CTU to the research team, who communicated this to schools and the intervention team.

Protecting against selection bias

All schools were recruited before the intervention commenced. The randomisation schedule was drawn up once the schools had consented and after the baseline survey, thus guarding against selection biases at entry of clusters to the trial. Retention of control schools was maximised by ensuring regular senior liaison and provision of participation incentives (£500 per school).

We had very high student participation in our pilot study: 96% of students eligible at baseline, and 93% at follow-up. To minimise bias, we used in-school, mail, and telephone contacts to try to include all enrolled students absent at either baseline or follow-up questionnaires. Students absent on the day of questionnaire surveys were provided with the questionnaire to complete and an envelope in which to seal the completed questionnaire, which schools then collated and posted to the research team. We did not attempt to follow up students who had left the school.

Allocation concealment mechanisms

As with most social intervention trial schools, their students, teachers and other staff could not be 'blinded' to allocation status. However, field work staff were blinded to allocation, as were data-input staff. Analysis of follow-up quantitative data was undertaken blind to allocation.

Implementation

Randomisation and stratification was undertaken remotely by the CTU at LSHTM.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Bonell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. 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.

Blinding

After randomisation, schools, the intervention team and the process evaluators could not be 'blinded' to allocation status. However, other field work staff were blinded to allocation, as were the outcome evaluation research team lead (RV) and data input and data analysis staff. Process and economic evaluation researchers could not be blinded.

It was not possible to blind the costing conducted alongside the process evaluation. Researchers remained blinded to the arm and school name while preparing statistical analysis of cost data from questionnaires and the CHU9D until the analysis was completed.

Statistical methods

Primary and secondary outcome analysis

The primary analysis of outcomes was intention to treat, including all randomised schools and participants at each wave. Each measure was analysed using a separate mixed model, with the outcomes from each time point treated as a repeated measures outcome. Fixed effects of treatment (LT vs. control) and time (baseline, 24 months and 36 months) and the interaction between treatment and time were specified, and the estimated baseline measures were constrained to be identical in the two arms of the trial. This is equivalent to adjusting for baseline and permitting the relationship between baseline and follow-up scores to differ at each time point, but offers the additional advantage that the data from all participants contribute to the analysis, even when there were missing data at follow-up. Random effects for school and for participants were specified to allow for correlations within schools, and repeated measures within participants. Statistical significance for these analyses was taken at the 5% level ($p < 0.05$). As prespecified in the statistical analysis plan, we carried out analyses adjusted only for baseline measures of the outcomes, and the analyses adjusted for baseline measures of outcomes, sex, ethnicity and SES as well as for the school-level stratifying factors (single-sex vs. mixed-sex school; school-level deprivation; value-added strata) as the primary analysis.

For the joint primary outcomes (GBS and ESYTC), mixed linear regression models with random effects at the participant and school levels were used to estimate a mean difference in GBS and ESYTC scores between the two arms of the trial. Formal testing was restricted to a prespecified number of secondary outcomes, and appropriate multilevel models were used to examine the effect of the intervention. For continuous outcomes, we report unadjusted and adjusted mean differences with 95% confidence intervals (CIs) and adjusted effect sizes (standardised mean difference). For binary and ordinal outcomes, we report unadjusted and adjusted odds ratios. Additionally, we report adjusted risk differences for binary outcomes although these cannot be calculated for ordinal outcomes. Evidence for any differential effects of the intervention on the primary and secondary outcomes, by subgroup, was assessed using likelihood ratio tests for the treatment by subgroup interaction terms. The effects in the different subgroups were estimated directly from the regression model with the interaction term included.

The following four subgroups were prespecified: (1) sex; (2) SES, measured using the Health Behaviour in School-aged Children FAS,¹²⁰ described as low SES for FAS scores of 0–5 and high SES for FAS scores of 6–9; (3) baseline bullying experience (high, defined as frequent – at least weekly – experience of bullying or being upset by it, vs. medium/low, defined as rare – less than weekly – experience of bullying and not being upset by it) based on the GBS; and (4) baseline behaviour problems based on the ESYTC, with high levels of behaviour problems defined as ESYTC scores of > 0 and low levels defined as a score of 0.

When there was evidence of non-normality in the continuous outcome measures, non-parametric bootstrapping, with 1000 samples, was used to estimate the effect of the intervention, and bias-corrected CIs are reported. When this was done, p -values were estimated using permutation tests.

Process evaluation analyses

Quantitative data were entered into CSPro (version 7.2.1; United States Census Bureau, Washington, DC, USA) or Microsoft Excel® (2016; Microsoft Corporation, Redmond, WA, USA) and then transferred and analysed in Stata. Quantitative analysis used descriptive statistics to assess intervention implementation, awareness, reach, satisfaction and empowerment. Quantitative data also examined the association between trial arm and our potential mediators using the same approaches as were used to assess the primary and secondary outcomes. Qualitative data were organised in NVivo (version 11; QSR International, Warrington, UK) and analysed to explore views about intervention processes and contexts. Thematic content analysis of qualitative data was undertaken by Emily Warren and Chris Bonell using the following stages of analysis.¹²¹ First, the researcher created a coding framework based on the research questions and theory guiding the intervention. These starting codes were constructed to reflect the concepts informing our questions and theory, such as the intervention components, the aspects of the school setting included in our inclusion and stratification criteria, and the aspects of implementation such as fidelity, feasibility and acceptability. Other codes were developed inductively to reflect other concepts that featured in interview transcripts, or to subdivide and better describe our starting codes. Transcripts were analysed using *in vivo* codes to identify key themes that were immediately apparent in participant accounts, as well as axial codes to explore inter-relationships between these *in vivo* codes (e.g. to explore how implementation varied across different school contexts). Analyses used the method of constant comparisons and examination of deviant cases to refine the emerging analysis.

Data Monitoring Committee

A DMC was established independent of the investigators and of the TSC, but it reported to the TSC and (via the TSC) to the sponsors and the NIHR programme. This consisted of an independent chairperson, a senior statistician and at least one other senior academic independent of the investigators. It met approximately yearly during the study. The DMC monitored data for quality and completeness. The DMC examined the results of an interim analysis at 24 months to consider any potential harms.

Chapter 3 Results

Participant flow

A total of 6667 students in the 40 participating schools provided data at baseline, with the participation rate being 93.6% of the students on the school roll (intervention arm, 92.9%; control arm, 94.3%). Characteristics of the schools and students at baseline and at 24 and 36 months are shown by trial arm in *Appendix 1* (see *Tables 26 and 27*). All schools participated in the follow-up surveys at 24 and 36 months; the numbers of students who completed the questionnaires at baseline, 24 months and 36 months were similar between the arms (*Figure 1*). Student and school characteristics and outcomes at baseline were well balanced between the arms.

Recruitment

Baseline surveys took place in March–July 2014, with the 24-month follow-up in April–June 2016 and the 36-month follow-up in April–June 2017. The trial ended in summer 2017, as the designated intervention time limit had been reached.

Baseline data

Summaries of demographic characteristics of respondents are presented in *Appendix 1* (see *Table 27*). The arms were well balanced in terms of demographic characteristics and the primary and secondary outcome measures.

The data presented come primarily from mixed-sex schools (80%) with a 'good' Ofsted rating (60%). There was a good mix of male and female participants, ethnicities and religions. The majority of student participants came from a two-parent family (72%) in which at least one parent was employed (73%).

In terms of primary and secondary outcomes measures, reports of bullying and aggressive behaviour, emotional difficulties, QoL, risk behaviours, truancy, NHS service use and contact with police were similar between the control and intervention arms at baseline (see *Table 1*).

Staff outcomes in relation to stress and burnout and QoL were also similar across control and intervention conditions at baseline (see *Appendix 1, Table 28*).

Numbers analysed

A total of 3087 participants were included in analysis from the control condition and 2873 were included from the intervention condition. Analysis was undertaken by original assigned group. Details on missing data are shown in *Appendix 1* (see *Table 43*).

Outcomes and estimation

Primary and secondary analysis

Overall GBS bullying scores were lower among intervention schools than among control schools at 36 months (adjusted mean difference -0.03 , 95% CI -0.06 to 0.00 ; adjusted effect size -0.08). There was no evidence of a difference in misbehaviour/delinquency ESYTC scores (adjusted mean difference -0.13 ,

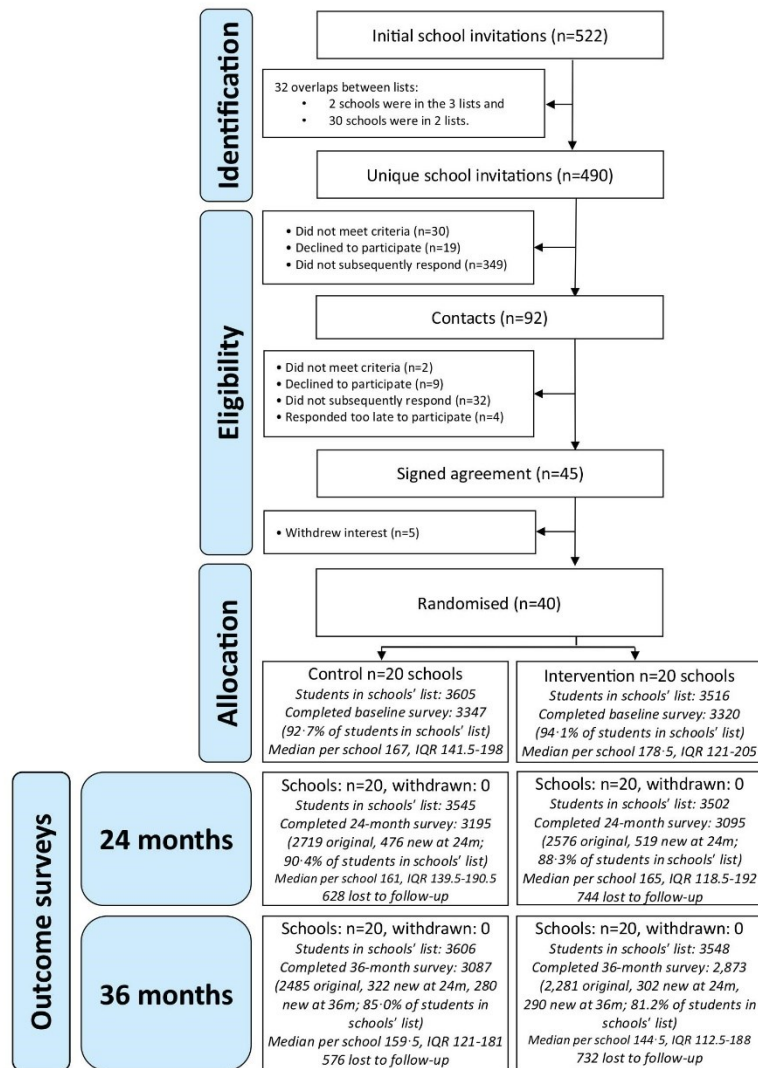


FIGURE 1 The CONSORT flow diagram. Reproduced from Bonell et al.¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

95% CI -0.43 to 0.18 ; adjusted effect size -0.03) between arms; however, the direction of effect suggests a positive effect of the intervention. These results are shown in *Table 2*.

Secondary outcomes and adjusted and unadjusted intervention effects for the GBS and ESYTC measured at 24 months and other secondary outcomes at 36 months are shown in *Table 3*. Although the direction of effect was, again, positive, there was no evidence of difference in the GBS overall score or the ESYTC misbehaviour/delinquency scores at 24 months. At 36 months, students in intervention schools had

TABLE 1 Primary and secondary outcomes at baseline by trial arm

	Control (3347 students) [†]	Intervention (3320 students) [†]	Overall (6667 students) [†]
Primary outcomes			
GBS overall score, mean (SD)	0.51 (0.63)	0.48 (0.60)	0.49 (0.61)
Teasing	0.79 (0.99)	0.74 (0.95)	0.76 (0.97)
Rumours	0.52 (0.85)	0.50 (0.81)	0.51 (0.83)
Deliberate exclusion	0.40 (0.79)	0.41 (0.80)	0.40 (0.79)
Threatened or hurt	0.36 (0.74)	0.30 (0.68)	0.33 (0.71)
ESYTC overall score, mean (SD)	2.92 (4.84)	2.72 (4.77)	2.82 (4.81)
Secondary outcomes			
SDQ total difficulties score, mean (SD)	11.00 (5.99)	10.70 (5.76)	10.85 (5.88)
Emotional problems	3.15 (2.39)	3.17 (2.35)	3.16 (2.37)
Conduct problems	2.00 (1.86)	1.90 (1.83)	1.95 (1.85)
Hyperactivity	4.00 (2.48)	3.84 (2.41)	3.92 (2.45)
Peer problems	1.86 (1.78)	1.81 (1.75)	1.83 (1.76)
Pro-social strengths	7.56 (1.91)	7.65 (1.93)	7.60 (1.92)
SWEMWBS overall score	24.11 (5.91)	24.33 (5.91)	24.21 (5.91)
PedsQL overall score	80.39 (14.31)	80.98 (14.08)	80.68 (14.20)
Physical	85.42 (14.88)	85.75 (14.81)	85.58 (14.84)
Emotional	75.90 (22.36)	76.40 (21.80)	76.15 (22.08)
Social	84.31 (17.88)	84.88 (17.70)	84.59 (17.79)
School	72.75 (18.76)	73.96 (18.55)	73.34 (18.67)
Psychosocial	77.70 (16.29)	78.41 (15.84)	78.04 (16.07)
Ever smoked, n (%)			
No	3083 (94.22)	3051 (95.28)	6134 (94.75)
Yes	189 (5.78)	151 (4.72)	340 (5.25)
If yes, how often in past month, n (%)			
None	132 (77.65)	100 (76.34)	232 (77.08)
Once or twice	30 (17.65)	20 (15.27)	50 (16.61)
About once a week	3 (1.76)	6 (4.58)	9 (2.99)
Daily or almost daily	5 (2.94)	5 (3.82)	10 (3.32)
Ever consumed alcohol, n (%)			
No	2751 (84.96)	2783 (87.63)	5534 (86.28)
Yes	487 (15.04)	393 (12.37)	880 (13.72)
If yes, how often in past month, n (%)			
None	349 (74.89)	268 (72.43)	617 (73.80)
Once or twice	88 (18.88)	88 (23.78)	176 (21.05)
Once a week or more	29 (6.22)	14 (3.78)	43 (5.14)

continued

RESULTS

TABLE 1 Primary and secondary outcomes at baseline by trial arm (*continued*)

	Control (3347 students) [†]	Intervention (3320 students) [†]	Overall (6667 students) [†]
Ever been offered illicit drugs, n (%)			
No	2924 (87.36)	2967 (89.37)	5891 (88.36)
Yes, but did not try them	258 (7.71)	175 (5.27)	433 (6.49)
Yes, and tried them	30 (0.90)	18 (0.54)	48 (0.72)
Truancy, n (%)			
No	3019 (94.11)	2989 (94.26)	6008 (94.18)
Yes	189 (5.89)	182 (5.74)	371 (5.82)
Exclusion from school, n (%)			
No	3154 (95.52)	3143 (96.92)	6297 (96.21)
Yes	148 (4.48)	100 (3.08)	248 (3.79)
Use of NHS services in past year, n (%)			
No	1836 (55.86)	1720 (53.33)	3556 (54.61)
Yes	1451 (44.14)	1505 (46.67)	2956 (45.39)
Contact with police			
No	3050 (92.06)	2982 (91.61)	6032 (91.84)
Yes	263 (7.94)	273 (8.39)	536 (8.16)

[†] The number of students who responded at this survey; actual number of responses to each question varies, but item non-response is similar across arms.
 Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

TABLE 2 Primary outcomes and intervention effects at 36 months

	Control group (3087 students) ^a	Intervention group (2281 students) ^a	Unadjusted difference (95% CI)	<i>p</i> -value	Adjusted difference (95% CI)	<i>p</i> -value	Adjusted effect size
GBS overall score	0.34 (0.02)	0.29 (0.02)	-0.03 (-0.06 to -0.002)	0.0395	-0.03 (-0.06 to -0.001)	0.0441	-0.08
Teasing	0.55 (0.03)	0.47 (0.03)	-0.04 (-0.09 to 0.01)		-0.05 (-0.10 to 0.00)		-0.07
Rumours	0.37 (0.02)	0.31 (0.02)	-0.06 (-0.10 to -0.02)		-0.07 (-0.11 to -0.02)		-0.10
Deliberate exclusion	0.24 (0.01)	0.22 (0.02)	-0.04 (-0.08 to 0.004)		-0.04 (-0.08 to 0.01)		-0.06
Threatened or hurt	0.21 (0.02)	0.18 (0.02)	0.01 (-0.02 to 0.05)		0.01 (-0.03 to 0.05)		0.02
ESYTC overall score	4.33 (0.20)	4.04 (0.21)	-0.07 (-0.38 to 0.25)	0.6820	-0.13 (-0.43 to 0.18)	0.4199	-0.03

^a Shows the number of students who responded at this survey; actual number of responses to each question varied, but non-response for each item was similar across arms.
 Data are mean (SE) unless otherwise stated.
 The GBS score decreased and the ESYTC increased in both arms over time. Similarly we observed an increase in risk taking behaviours over time in both arms of the trial.¹²³⁻¹²⁵
 Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

TABLE 3 Secondary outcomes and intervention effects at 24 months

Outcome	Control (3195 students), ^a mean (SE)	Intervention (3095 students), ^a mean (SE)	Unadjusted difference (95% CI); p-value	Adjusted difference (95% CI); p-value	Adjusted effect size (95% CI)
GBS overall score	0.42 (0.02)	0.37 (0.02)	-0.02 (-0.05 to 0.01); 0.2198	-0.02 (-0.05 to 0.01); 0.1581	-0.05 (-0.15 to 0.05)
Teasing	0.66 (0.03)	0.59 (0.03)	-0.02 (-0.07 to 0.03)	-0.03 (-0.08 to 0.01)	-0.05 (-0.14 to 0.04)
Rumours	0.44 (0.02)	0.41 (0.02)	-0.02 (-0.06 to 0.02)	-0.02 (-0.06 to 0.02)	-0.04 (-0.11 to 0.04)
Deliberate exclusion	0.31 (0.02)	0.30 (0.02)	-0.03 (-0.07 to 0.01)	-0.03 (-0.07 to 0.01)	-0.05 (-0.13 to 0.03)
Threatened or hurt	0.26 (0.02)	0.22 (0.02)	0.01 (-0.02 to 0.05)	0.01 (-0.03 to 0.04)	0.01 (-0.07 to 0.09)
ESYTC overall score	4.24 (0.28)	3.96 (0.28)	-0.04 (-0.34 to 0.27); 0.8113	-0.06 (-0.35 to 0.24); 0.7206	-0.01 (-0.12 to 0.09)
PedsQL overall score	79.75 (0.50)	80.97 (0.51)	0.77 (0.05 to 1.48)	0.88 (0.17 to 1.60)	
Physical health	85.39 (0.52)	86.01 (0.53)	0.30 (-0.49 to 1.09)	0.49 (-0.27 to 1.26)	
Psychosocial health	76.71 (0.55)	78.37 (0.56)	1.08 (0.28 to 1.89)	1.15 (0.35 to 1.96)	
Emotional functioning	73.90 (1.07)	74.65 (1.10)	0.58 (-0.60 to 1.77)	0.74 (-0.41 to 1.90)	
Social functioning	85.98 (0.42)	87.74 (0.43)	1.15 (0.32 to 1.99)	1.04 (0.20 to 1.88)	
School functioning	70.19 (0.60)	72.68 (0.61)	1.50 (0.50 to 2.50)	1.66 (0.65 to 2.66)	
SDQ total difficulties score ^b	11.83 (0.16)	11.23 (0.17)	-0.40 (-0.67 to -0.12)	-0.41 (-0.69 to -0.13)	
Emotional problems	3.46 (0.11)	3.40 (0.12)	-0.10 (-0.22 to 0.03)	-0.11 (-0.23 to 0.01)	
Conduct problems	2.22 (0.07)	2.01 (0.07)	-0.12 (-0.21 to -0.03)	-0.13 (-0.22 to -0.04)	
Hyperactivity	4.33 (0.07)	4.15 (0.07)	-0.05 (-0.17 to 0.07)	-0.05 (-0.17 to 0.07)	
Peer problems	1.84 (0.05)	1.68 (0.05)	-0.14 (-0.22 to -0.05)	-0.15 (-0.24 to -0.06)	
Pro-social strengths	6.96 (0.10)	7.18 (0.10)	0.12 (0.02 to 0.23)	0.15 (0.04 to 0.25)	
SWEMWBS total well-being index	23.54 (0.20)	23.79 (0.21)	0.09 (-0.24 to 0.41)	0.15 (-0.17 to 0.47)	
Age at sexual debut	12.30 (1.27)	12.74 (1.29)	0.06 (-1.53 to 1.66)	-0.22 (-1.60 to 1.16)	
Modified aggression scale, bullying subscale	2.74 (0.23)	2.61 (0.23)	0.00 (-0.56 to 0.57)	0.01 (-0.29 to 0.32)	

continued

RESULTS

TABLE 3 Secondary outcomes and intervention effects at 24 months (continued)

Outcome	Control (3195 students), ^a mean (SE)	Intervention (3095 students), ^a mean (SE)	Unadjusted difference (95% CI); p-value	Adjusted difference (95% CI); p-value	Adjusted effect size (95% CI)
Ever smoked regularly					
No	2568 (83.40)	2595 (87.97)	1.00	1.00	
Yes	511 (16.60)	355 (12.03)	0.67 (0.48 to 0.92)	0.63 (0.46 to 0.88)	
If yes, how long since last smoked			1.12 (0.80 to 1.56) ^c	1.14 (0.81 to 1.59) ^c	
< 1 day	65 (12.87)	51 (14.66)			
1–3 days	29 (5.74)	20 (5.75)			
4–7 days	27 (5.35)	20 (5.75)			
> 1 week, < 1 a month	69 (13.66)	40 (11.49)			
1–2 months	64 (12.67)	41 (11.78)			
3–6 months	90 (17.82)	49 (14.08)			
> 6 months	161 (31.88)	127 (36.49)			
Ever drunk alcohol					
No	2028 (64.88)	2127 (70.95)	1.00	1.00	
Yes	1098 (35.12)	871 (29.05)	0.79 (0.62 to 1.02)	1.26 (0.91 to 1.75)	
If yes, had alcohol in the past week					
No	885 (83.65)	675 (81.33)	1.00	1.00	
Yes	173 (16.35)	155 (18.67)	1.16 (0.82 to 1.63)	1.26 (0.91 to 1.75)	
Number of times really drunk			0.65 (0.37 to 1.14) ^c	0.65 (0.41 to 1.02) ^c	
Never	911 (66.35)	771 (70.80)			
Once	262 (19.08)	166 (15.24)			
2–3 times	120 (8.74)	97 (8.91)			
4–10 times	45 (3.28)	35 (3.21)			
> 10 times	35 (2.55)	20 (1.84)			
Binge drinking (> 5 drinks in a row) in past 30 days			0.89 (0.60 to 1.33) ^c	0.5832	
0	1204 (81.30)	978 (81.98)			
1 to 2	181 (12.22)	143 (11.99)			
3 to 5	20 (3.38)	46 (3.86)			
6 to 9	17 (1.15)	6 (0.50)			
≥ 10	29 (1.96)	20 (1.68)			

TABLE 3 Secondary outcomes and intervention effects at 24 months (*continued*)

Outcome	Control (3195 students), ^a mean (SE)	Intervention (3095 students), ^a mean (SE)	Unadjusted difference (95% CI); p-value	Adjusted difference (95% CI); p-value	Adjusted effect size (95% CI)
Ever been offered illicit drugs			0.63 (0.41 to 0.96) ^c	0.63 (0.44 to 0.92) ^c	
No	2346 (75.46)	2421 (81.19)			
Yes, but did not try them	618 (19.88)	469 (15.78)			
Yes, and tried them	145 (4.66)	92 (3.09)			
Truancy (skipped school in the past 3 months)					
No	2761 (88.95)	2685 (89.89)	1.00	1.00	
Yes	343 (11.05)	302 (10.11)	0.93 (0.72 to 1.20)	0.91 (0.70 to 1.18)	
Exclusion from school (temporarily or permanently) ^d					
No	2846 (90.18)	2791 (91.63)	1.00	1.00	
Yes	310 (9.82)	255 (8.37)	0.83 (0.54 to 1.29)	0.81 (0.55 to 1.21)	
Used any contraception at first sex					
No	36 (30.51)	32 (33.68)	1.00	1.00	
Yes	82 (69.49)	63 (66.32)	0.73 (0.28 to 1.88)	0.71 (0.25 to 1.99)	
Use of NHS services in past 12 months					
No	1633 (51.91)	1617 (53.33)	1.00	1.00	
Yes	1513 (48.09)	1415 (46.67)	0.87 (0.75 to 1.00)	0.87 (0.75 to 1.01)	
Contact with police in past 12 months					
No	2732 (86.26)	2673 (87.61)	1.00	1.00	
Yes	435 (13.74)	378 (12.36)	0.83 (0.64 to 1.09)	0.82 (0.63 to 1.07)	

a Shows the number of students who responded at this survey; actual number of responses to each question varied, but non-response for each item was similar across arms.
b SDQ total difficulties score does not include the pro-social score subscale.
c Proportional odds ratio.
d Based on repeat cross-sectional model because longitudinal model did not converge.

higher QoL (PedsQL adjusted effect 1.44, 95% CI 0.07 to 2.17; adjusted effect size 0.14) and psychological well-being scores (SWEMWBS 0.33, 95% CI 0.00 to 0.66; adjusted effect size 0.07) and lower psychological total difficulties (SDQ total score -0.54 , 95% CI -0.83 to -0.25 ; adjusted effect size -0.14) than students in control schools. There was evidence that those in intervention schools also had lower scores for emotional, conduct, hyperactivity and peer problems (SDQ subscales; see *Table 3*).

Students in intervention schools also had lower odds of having ever smoked regularly (OR 0.58, 95% CI 0.43 to 0.80; adjusted risk difference -0.03 , 95% CI -0.05 to -0.01), lower odds of having ever drunk alcohol (OR 0.72, 95% CI 0.56, 0.92; adjusted risk difference -0.03 , 95% CI -0.06 to -0.01) and lower odds of having ever been offered or tried illicit drugs (OR 0.51, 95% CI 0.36 to 0.73). There was evidence

RESULTS

that among students in the intervention arm who had ever smoked the time since their last cigarette was longer than in the control arm and, similarly, that, among those who had ever drunk alcohol, there were lower odds of having drunk in the past week (OR 0.67, 95% CI 0.50 to 0.91), number of times been really drunk (OR 0.57, 95% CI 0.33 to 0.98) and binge drinking (OR 0.77, 95% CI 0.59 to 1.00). Similarly, students in intervention schools had lower odds of having ever been in contact with the police in the past 12 months than those in control schools (OR 0.74, 95% CI 0.56 to 0.97; adjusted risk difference -0.02, 95% CI -0.04 to -0.00).

We found no evidence of differences in the CHU9D HRQoL score, sexual risk behaviour, use of contraception at most recent sexual contact, bullying perpetration or use of NHS services.

Staff outcome data

No differences were found in relation to staff stress or burnout of QoL between the control and intervention arms of the study at 24 or 36 months (*Table 4*).

Ancillary analyses

Analysis of the primary and secondary outcomes by prespecified subgroup (see *Appendix 1, Table 29*) suggests that the intervention had a greater effect in boys than in girls for many secondary outcomes [PedsQL overall score; SDQ total difficulties score; SWEMWBS total well-being index; bullying perpetration (modified aggression scale); CHU9D overall score; ever smoked; ever drunk alcohol; contact with police]. The intervention was also more effective in students with higher baseline bullying experience, with greater effects on bullying (GBS score) and psychological secondary outcomes (PedsQL overall score; SDQ total difficulties score; SWEMWBS total well-being index; CHU9D overall score). The intervention was more effective in those with greater baseline aggression, with greater effects on both primary outcomes (GBS score; ESYTC overall score), psychological secondary outcomes (PedsQL overall score; SDQ total difficulties score; SWEMWBS total well-being index; CHU9D overall score) and risk behaviours (ever smoked; ever drunk alcohol). There was no suggestion of any difference in the outcomes by level of SES).

Adverse events

A serious adverse event was defined as (1) any death or event requiring hospitalisation in a young person reported to investigators or (2) a response to study questionnaires that prompted significant concerns about mental health, sexual risk or child safety, which were then communicated to the school. There were 15 reported SAEs to date (*Table 5*). The total number of reported serious adverse events was similar in each arm.

Process evaluation

Fidelity was variable, with a reduction in the fidelity of formal intervention activities in year 3. The median fidelity score for years 1 and 2 (maximum possible score 8) was 6 (interquartile range 5–7), whereas for year 3 (maximum score 4) the median was 1 (interquartile range 0–3). In year 3, 15 schools sustained restorative practice. Interviews with AG members and focus groups with staff in case study schools suggested that, in year 3, schools commonly incorporated what they regarded as the most useful AG functions into mainstream school structures and processes. For example, involving students in decision-making and further review of policies to ensure that these supported restorative practice was integrated into existing school committees.

TABLE 4 Staff outcomes and intervention effects at 24 and 36 months

Measure	24 months						36 months					
	Arm, mean (SD)		Unadjusted effect		Adjusted effect		Arm, mean (SD)		Unadjusted effect		Adjusted effect	
	Control	Intervention	Difference (95% CI)	p-value	Difference (95% CI)	p-value	Control	Intervention	Difference (95% CI)	p-value	Difference (95% CI)	p-value
Staff stress and burnout – Maslach Burnout Inventory												
Emotional exhaustion score	23.03 (0.73)	24.23 (0.75)	0.81 (-0.24 to 1.86)	0.1322	0.80 (-0.34 to 1.93)	0.1680	23.97 (0.63)	23.81 (0.63)	-0.46 (-1.52 to 0.61)	0.3998	-0.55 (-1.74 to 0.64)	0.3625
Depersonalisation score	5.71 (0.32)	5.91 (0.33)	0.25 (-0.27 to 0.78)	0.3434	0.26 (-0.30 to 0.81)	0.3620	6.22 (0.36)	6.07 (0.36)	-0.05 (-0.58 to 0.48)	0.8414	0.03 (-0.55 to 0.60)	0.9276
Personal achievement score	34.67 (0.35)	35.58 (0.37)	0.80 (0.08 to 1.51)	0.0291	0.88 (0.14 to 1.62)	0.0195	35.12 (0.28)	35.22 (0.29)	-0.02 (-0.73 to 0.69)	0.9535	0.20 (-0.56 to 0.96)	0.6127
Staff quality of life – SF-12 v2 Health Survey												
Physical health score	54.97 (0.32)	55.22 (0.34)	-0.35 (-1.05 to 0.34)	0.3187	-0.26 (-0.99 to 0.47)	0.4825	55.62 (0.36)	55.18 (0.36)	-0.71 (-1.40 to -0.02)	0.0449	-0.61 (-1.36 to 0.13)	0.1058
Mental health score	44.56 (0.49)	43.71 (0.52)	-0.66 (-1.64 to 0.32)	0.1883	-0.45 (-1.49 to 0.58)	0.3896	43.76 (0.44)	43.69 (0.45)	0.15 (-0.83 to 1.13)	0.7642	0.31 (-0.76 to 1.38)	0.5691

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

RESULTS

TABLE 5 Number of serious adverse events by arm across the trial

SAE	Control	Intervention
Suicide	0	2
Responses showing potential for self-harm	0	4
Stabbing incidents	0	2
Possible non-consensual sex (including age < 10 years)	6	0
Disability or long-term illness	1	0
Total	7	8

Reported at the school level (any relevant events in any student) or student level (from survey responses).
Reproduced from Bonell *et al.*¹⁷² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Regarding specific components (training, AGs and restorative practices), these were delivered with good fidelity. The fidelity of curriculum delivery was lower. Increased fidelity of delivery in the first 2 years of intervention was associated with lower rates of bullying victimisation at 24 months, but not with rates of aggression. The fidelity score for year 3 was not associated with either primary outcome (see *Appendix 1, Table 30*).

Slightly over half of staff in intervention schools were aware that the school had been taking steps to reduce bullying and aggression, with this falling slightly between years 2 and 3 (*Table 6*). About one-third of students reported being aware that the school had been taking steps to reduce bullying (*Table 7*). About half reported that, if there was trouble at school, staff responded by talking to those involved to help them get on better. About two-thirds of students reported that teachers and students got together to build better relationships or discuss their views and feelings. Other data on the process evaluation are shown in *Appendix 1, Tables 31–35*.

Many control schools implemented similar activities to those prescribed in the intervention but with very variable quality. We identified five control schools that delivered restorative practice, social and emotional skills education and consultation with students on policy. A per-protocol analysis excluding these schools showed no discernible differences in the intervention effects compared with the intention-to-treat analyses (*Table 8*).

Response rates

Some elements of the process evaluation, such as interviews with facilitators, school staff and students, as well as observations of training and AGs, achieved excellent response rates (see *Appendix 1, Table 36*). However, other elements, such as the surveys of staff delivering the curriculum or restorative practice (RP), achieved poor response rates. This reflected what, in retrospect, was an excessive data collection burden on schools. The low response rates for some elements of the process evaluation in year 3, such as the survey of AG members, also reflected the reality that in this year not all schools delivered all aspects of the intervention, rendering data collection obsolete. However, despite this, our multisource approach meant that we were still able to assess what was happening in most schools, with the exception of a few schools in year 3, where it is reasonable to assume that few, if any, intervention activities were being implemented.

TABLE 6 Staff awareness of the intervention processes

Measure	Staff responses					
	24 months			36 months		
	Overall % across intervention group	N schools where school % is > 70%	N schools with no data	Overall % across intervention group	N schools where school % is > 70%	N schools with no data
This school has recently been taking steps to reduce bullying and aggression (% yes)	58.33	6	1	54.21	4	3
I support this new work (% yes)	91.71	19	2	91.40	20	3
I understand what is meant by 'restorative practice' (% yes)	77.90	14	1	79.52	13	3
I support using restorative practice in schools (% yes)	52.85	4	1	55.38	3	3
If there is trouble at this school, staff response includes talking to those involved to help them get on better (% tick)	85.99	16	1	89.14	17	3
Teachers and students at this school get together to build better relationships (% often/sometimes)	94.55	20	1	94.06	20	3
Teachers and students at this school get together to discuss their views and feelings (% often/sometimes)	83.95	19	1	82.39	17	3

Reproduced from Bonell *et al.*¹⁷⁷ © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

TABLE 7 Student awareness of the intervention processes

Measure	Staff responses					
	24 months			36 months		
	Overall % across intervention group	<i>N</i> schools where school % is > 70%	<i>N</i> schools with no data	Overall % across intervention group	<i>N</i> schools where school % is > 70%	<i>N</i> schools with no data
This school has recently been taking steps to reduce bullying and aggression (% yes)	34.29	0	0	33.00	0	0
If there is trouble at this school, staff response includes talking to those involved to help them get on better (% tick)	53.69	0	0	50.65	0	0
Teachers and students at this school get together to build better relationships (% often/sometimes)	69.68	12	0	67.85	9	0
Teachers and students at this school get together to discuss their views and feelings (% often/sometimes)	63.48	7	0	59.84	5	0

Reproduced from Bonell *et al.*¹⁷² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

TABLE 8 Adjusted intention-to-treat and per-protocol analysis for the main primary and secondary student outcomes at 36 months

Outcome	Intention-to-treat (adjusted)		Per-protocol (adjusted)	
	Difference (95% CI)	p-value	Difference (95% CI)	p-value
GBS overall score	-0.03 (-0.06 to -0.001)	0.0441	-0.03 (-0.06 to 0.01)	0.1211
ESYTC overall score	-0.13 (-0.43 to 0.18)	0.4199	-0.12 (-0.46 to 0.22)	0.4869
PedsQL overall score	1.44 (0.70 to 2.17)	0.0001	1.02 (0.22 to 1.83)	0.0122
SDQ total difficulties score	-0.54 (-0.83 to -0.25)	0.0002	-0.35 (-0.67 to -0.04)	0.0289
SWEMWBS total well-being index	0.33 (0.00 to 0.66)	0.0487	0.31 (-0.05 to 0.66)	0.0943
CHU9D overall score	0.01 (-0.00 to 0.01)	0.0795	0.00 (-0.00 to 0.01)	0.6087
Age of sexual debut	-0.35 (-1.48 to 0.78)	0.5409	-0.28 (-1.55 to 0.98)	0.6589
MAS bullying subscale score	-0.26 (-0.57 to 0.05)	0.0976	-0.21 (-0.57 to 0.14)	0.2342

Outcome	Intention-to-treat (adjusted)		Per-protocol (adjusted)	
	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value
Ever smoked regularly	0.58 (0.43 to 0.80)	0.0009	0.66 (0.47 to 0.93)	0.0160
Ever drunk alcohol	0.72 (0.56 to 0.92)	0.0094	0.79 (0.60 to 1.03)	0.0764
Ever tried illicit drugs	0.62* (0.41 to 0.93)	0.0221	0.91* (0.63 to 1.32)	0.6258
Used any contraception at first sex	1.08 (0.50 to 2.35)	0.8410	1.29 (0.54 to 3.08)	0.5741
Use of NHS in past 12 months	0.96 (0.82 to 1.11)	0.5652	0.93 (0.79 to 1.09)	0.3685
Contact with police	0.74 (0.56 to 0.97)	0.0269	0.73 (0.55 to 0.98)	0.0369

* Proportional odds ratio.

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Baseline context of intervention schools

Motivation to participate in the trial

A key theme was the variety in schools' motivations to participate in the trial. One subtheme focused on a desire to address bullying or an interest in RP, while another subtheme was a desire to provide evidence of performance. For example, in one school, the external facilitator felt that the first year of the trial was focused more on generating evidence for Ofsted than on substantive changes that would benefit students:

Can I just say that the whole of the first year the emphasis was on just getting this Ofsted. I felt anything we did that had as an agenda.

Facilitator 9, school AH, interview, year 2

Another facilitator explained:

There's also a thing where all schools are like this; they want to make sure . . . there's an Ofsted agenda which drives all schools. And she wanted to make sure as well that she was . . . they want 'outstanding', so she wants to make sure that she's got all the information possible to make sure that they can get through that.

Facilitator 2, school AD, interview, year 2

Context of control schools

Existing relevant services and practices in control schools during the trial period

Overall, in year 1, five schools delivered some form of RP, social and emotional skills education and student consultation on policy (schools AB, AC, AP, BH and BN), rising to 11 in year 3 (schools AB, AC, AG, AJ, AN, AP, AY, BA, BB, BH and BN).

School discipline systems

A key theme reflected in data from nearly all control schools was that addressing bullying, aggression, and other forms of misbehaviour was generally not schools' main priority. In almost all schools, staff reported that improving educational attainment was their key priority. Only in one school did participants (school BG) report that addressing bullying was the school's current main priority, and this was because it had been raised by Ofsted as an area to improve.

Existing use of RP was a common theme. In year 1, 15 schools were reported to use RP to address difficulties between students (schools AA, AB, AC, AG, AI, AJ, AN, AP, AR, BA, BB, BF, BG, BH and BN), with this rising to 17 in year 3 (to also include schools AQ and AY). Staff used the term RP and described processes that brought together each party in cases of conflict, as the following quotation illustrates:

Interviewer: So just going back to bullying and aggression in particular, would you say that your school uses any restorative approaches?

Respondent: Constantly.

Interviewer: Constantly. OK.

Respondent: Yeah . . . yesterday morning we had two boys in Year 12 . . . somebody had said something funny that the other boy didn't find funny . . . He took it quite seriously, it blew out of all proportion; and my parent support advisor and I sat down with the victim . . . no, well . . . the alleged perpetrator, and said, look we need to resolve this. We need to mend this, you've got to get back into lessons together. I'm not going to be on anybody's side . . . will you come and sit and meet? They had a little bit of time. They both said some things, we said things, they shook hands.

SLT member, school AC, interview, year 1

Other quotations, such as the following, offered a more nuanced account of how restorative sessions might focus on recognising and healing social and emotional conflicts:

We have another two students who were saying unkind things to each other in class and it just kept going on . . . And we would just talk through how it made each other feel, you know, what they'd said, what they'd done. And actually get them to come up with how they would resolve . . . how they're going to resolve it, how they're going to make each other feel better and what the strategies would be. So it may be, if they're not going to be friends, how to deal with each other in the community; or if we are able to – quite often we can get them to resolve to the point where . . . they would actually be able to communicate with each other again.

Phase leader for years 7 and 8, school AR, interview, year 1

A recurring idea was that RP was an approach to behaviour management that could be used alongside more traditional, punitive measures, but that could be more effective in addressing the underlying causes of conflict. In the following quotation, a teacher discusses the school's reasons for using the technique:

It's about actually sitting down with the victims and the perpetrators and actually having those conversations whereby people are able to express how the situation or the incident or the thing that was said has made them feel you know. And have that opportunity to directly speak to those

individuals who have caused the problem. And then for the other ones to be able to respond to that and actually really understand. Because I think sometimes when you sanction and they do their sanction, it's like . . . OK, has anyone actually stopped to really understand what this has led the other person to feel like or . . . how it's affected them you know? It's easy just to give sanctions and then draw a line under it. But really for us it is about actually getting to the root of the problem and understanding and making people aware that this is not acceptable.

Head of year, school AI, interview, year 1

One subtheme was variety in how RP was used. In interviews with staff at some schools, for example in the following quotation, RP was presented as a technique that permeated staff practice in and outside classrooms, and to prevent as well as respond to conflict:

And then also all the teachers are aware of how we should treat each other and how students should treat each other and act as role models. So then if something happens in their lessons they will use restorative justice type conversations with students. And also, if there's an incident, an argument or an act of unkindness within the lesson, then they're followed up with a restorative justice meeting that's usually held by the classroom teacher. So it does permeate across the school.

SLT member, school BH, interview, year 3

However, interviews in some other schools suggested that RP was regarded more as a resource to draw on in particular circumstances:

If it's a bunch of girls who are generally friends and they can't seem to get on and there's constant bickering, we might get one of our behaviour support workers in and they do a restorative thing at lunchtime where they get the girls together and they all have a chance to say their bit about what they're unhappy about in the relationship or not. And then go on from there. So we sometimes do that although it's not that often. We probably have about four of those sessions a year with different girls.

SLT member, school BB, interview, year 1

Most of those reporting that their school used RP suggested that it had been used for some time, but a few suggested that it had started after the school had been allocated to the control group in the trial. One SLT member described how, in response to being allocated to the control group, the school had initiated its own programme of RP:

I was really, really upset that we didn't get picked as one of the schools because we're actually at a point where we're going through so much change and what you guys were offering was exactly what we need. But we're going to do it ourselves, so it'll be interesting to see . . . how we compare.

SLT member, school BG, interview, year 3

Training was another subtheme. Eight control schools were reported to have trained some staff in using RP in year 1 (schools AB, AC, AG, AI, AW, BB, BH and BL), with this rising to 13 in year 3 (schools AA, AB, AC, AJ, AQ, AR, AY, BA, BB, BG, BH, BL and BN).

The ubiquity of policies and rules was another subtheme. In years 1 and 3, 18 schools were reported to have a policy on bullying or aggression (schools AA, AB, AC, AG, AI, AJ, AN, AP, AQ, AR, AY, BA, BB, BF, BG, BH, BL and BN). In year 1, staff of nine schools reported that their school had revised its policy on bullying or aggression in the last year (schools AA, AI, AN, AP, AR, BA, BB, BF and BN); in year 3 this rose to 11 schools (schools AA, AG, AI, AN, AQ, AR, AY, BB, BF, BH and BL). Policies on bullying and aggression were sometimes part of larger policies, such as on behaviour management or 'culture for learning'. Schools varied enormously in how staff described the importance of the policies. In most schools, the policies were salient in staff accounts and were said to be reviewed regularly. However, in a few

RESULTS

schools, the policies appeared to exist but not to greatly inform practice and to be seldom reviewed, as the quotation below illustrates:

Interviewer: Do you have any formal policies related to bullying and aggression?

Respondent: Yeah. We have an anti-bullying policy. We also have a behaviour policy. And a cyber-bullying policy.

Interviewer: And how often do you review them and who does this and how long does it take?

Respondent: Yeah. Well we're not very good on policies. We're good at practice, we're better at practice than we are at policy.

SLT member, school AB, interview, year 1

All schools had rules for student behaviour. Schools varied in whether these were presented primarily in terms of positive expectations or as a list of unacceptable behaviours. However, in all schools, bullying, aggression and other forms of physical or non-physical violence were identified as unacceptable. All schools employed punitive measures. These could work as a complement or an alternative to restorative measures. Many participants reported that punitive measures and RP were used alongside one another.

Personal, social and health education

The provision of PSHE was a recurring theme in staff accounts, with diversity of provision again an important subtheme. About half of schools offered regular timetabled subject lessons and around half delivered PSHE in tutor time or on so-called 'off-timetabled' or 'drop-down' days. In those schools where provision did not occur in specifically timetabled lessons, it was a recurring subtheme, exemplified by the quotation below, that such a format of provision was not ideal:

But obviously we have . . . like the PSHE days. So rather than it being embedded . . . we have 4 PSHE days a year. Like one-off days where the curriculum is changed to focus on exactly like the well-being. But from my point of view I think it would be better run across . . . like weekly during tutor groups, so it's consistent rather than kind of . . . It seems a bit like an add-on.

Pastoral co-ordinator, school BL, interview, year 1

It was very difficult to assess the extent to which control schools were delivering social and emotional skills curricula to students because staff discussed such provision in very vague terms, generally because they were not directly involved in its delivery. Staff commonly referred in general terms to lessons addressing topics such as 'well-being', 'resilience', 'relationships' or 'social issues', but were vague about details. In years 1 and 3, there was some evidence that 13 schools provided their students with some sort of social and emotional skills curriculum (schools AB, AC, AG, AJ, AN, AP, AQ, AR, AY, BA, BB, BH and BN). Staff commonly reported that social and emotional education was integrated into a range of academic subject lessons, but this provision did not appear to be based on a formal curriculum with clear learning objectives. For example, a member of staff described how social and emotional learning occurred in lessons such as drama:

Interviewer: The social and emotional education, is that covered in PSHE?

Respondent: To some extent, yeah. There's stuff, there's definitely lessons and things about healthy relationships and interaction buddying, these kind of things are . . . are covered there.

Interviewer: Are they covered in any other subjects?

Respondent: I think that, obviously being a drama teacher, we do a lot of stuff to do with well-being and . . . kind of emotional education. But there aren't any subjects that . . . are kind of . . . responsible for teaching it.

Teacher and deputy head, school AG, interview, year 1

Student participation in decision-making

Student participation in decision-making was a recurring theme. In years 1 and 3, there was evidence that 14 schools had an active student forum (schools AB, AC, AG, AI, AJ, AP, AQ, AR, BA, BB, BF, BG, BH and BN). In years 1 and 3, 15 schools were reported to consult with students on some policy decisions (schools AA, AB, AC, AG, AI, AJ, AN, AP, AQ, AY, BA, BB, BG, BH and BN). In year 1, there was evidence that in seven schools students and staff met together to make decisions (schools AC, AJ, AP, AQ, BA, BN and BG); this rose to 10 schools in year 3 (schools AA, AB, AC, AG, AP, AQ, AY, BB, BG and BN). In year 1, there was some evidence that five schools (schools AB, AJ, AP, AR and BN) surveyed students on health- or well-being-related issues; this rose to six schools (schools AG, AJ, AR, AY, BA and BB) in year 3.

In some schools, staff reported elaborate processes for students' active participation in decision-making alongside staff. For example, a staff member in one control school reported a parallel student leadership team interacting with the staff SLT:

We have student voice, student council; so we have . . . The same where we have SLT we have JLT [junior leadership team], which is the junior leadership team. These are the students who apply for the job, through an application process. And they have last interviews and then they are selected as member of the JLT. They are attached, each of those is attached to a member of SLT if you like, so they kind of shadow them.

Head of year, school AI, interview, year 1

However, in other schools, staff reported a much more limited and what was sometimes seen as a tokenistic system for participation, as the following quotation from a staff member suggests:

We have the student voice and all those kind of school councils and stuff, which . . . I don't know; my own personal view is I don't know, they're a little bit tokenistic . . . We've got our school prefect teams. But I guess ultimately . . . they can propose quite a few things, but it rests with the Head and the senior leadership team whether those things are going to go through or not. So I think students . . . I think certain voices . . . I think they're heard. But from working, again talking from a pastoral level, and working with the more challenging students; probably they don't feel like their voice is heard.

Pastoral co-ordinator, school BL, interview, year 1

Existing relevant policies and practices in intervention schools during the trial period

School discipline systems

Another theme was staff interest in RP. At baseline, staff from five schools reported in interviews that they had been interested in moving towards using RP to address student behaviour before taking part in the trial (schools AD, AK, AM, AX and BI). A subtheme was the existing use of the approach. Seven schools (schools AH, AK, AL, AX, BD, BI and BK) were already using some form of RP. Delivery within a school was generally inconsistent:

We also have a guy who works in the internal exclusion room and wherever he has two students who've been involved in an incident together, he will sit them down together or talk about the incident just within the room, and approach it from that way. So yeah, the restorative practice does go on. But I think it's in quite isolated pockets; it's not across the whole staff.

Year lead, school AH, interview, year 1

Staff in three other schools reported holding conversations to address incidents of bullying or misbehaviour, but did not describe this as RP (schools AD, AS and BE). Some teachers used restorative approaches alongside other approaches. A teacher from school BD explained:

The approach to bullying is more, from my experience and other members of staff, that if there is an issue, having a discussion about it is the first action to take and having the students or who's involved

RESULTS

having a discussion about, you know, this is what's happened, this is how this person feels, and that, and to express themselves so that the other students can . . . hear you know what they're doing has an effect. And also they wouldn't like it. And then usually you might just set a detention, or they sometimes underline the issues where students have been friends before, they're not friends any more or stuff like that.

Behaviour manager, school BD, interview, year 1

However, another subtheme was the use of more punitive approaches. Intervention facilitators described schools AH, AT and BD as primarily managing student behaviour through quite 'disciplinary' approaches. A teacher from school AT described the school as dealing with misbehaviour 'quite severely, quite harshly'. When asked if the school had ever used RP before, the teacher replied:

No, not at all. It's more . . . I feel more accusatory, more . . . 'you have done this' and then there's been the punishment that follows: isolation and then to be expelled for a certain number of days, whatever it is.

Intervention co-ordinator, school AT, interview, year 1

The importance of written policies was another subtheme, with most schools having policies on bullying, and many schools embedding this within their broader behaviour policies. Staff in most schools reported that they were clear about the bullying policy. Occasionally, staff reported not knowing what the bullying policies were (schools AE, AM and AU) or that the policies were not implemented (schools AD and AM).

Training was another subtheme. Several schools had at least a few members of staff who had at some point been trained in RP and who used this technique individually (schools AD, AH, AK, AO, AT, AZ, BE and BM). Staff in only three schools reported having whole-school training (schools AH, AO and BC) and 10 reported receiving no training in dealing with bullying (schools AE, AF, AL, AM, AU, AW, AX, AZ, BE and BK). Some schools offered training on issues indirectly connected with bullying, such as behaviour management (schools AK, AM, AS and BE) or safeguarding (schools AD, AF, AT, BIAD, AF, BI and AT):

We have loads of training on all kinds of different things. We are inundated with training. But directly relating . . . we have child protection training, but that doesn't really cover bullying, so no, I don't think there's anything that specifically deals with bullying.

Subject teacher, school AD, interview, year 1

Personal, social and health education

The variety in provision of PSHE was a prominent theme. Most schools offered distinct PSHE lessons or incorporated such learning into tutor times, generally weekly or twice a week (schools AD, AH, AK, AM, AO, AS, AU, AW, BE and BI). A few supplemented this with teaching PSHE content in assemblies (schools AL, AX and BE) or specific days or part-days when students were taken off their normal timetable of lessons (schools AD, AH, AS, AT and AX). Some schools delivered PSHE through other subjects, including citizenship, religious education and biology (schools AE, AL, AW, AX, BE and BI). The amount of time and scheduling varied greatly, from 20 minutes per fortnight (school AW) to 20 hours per term (school AS). In some schools, teachers reported that the time spent on PSHE had contracted in recent years (schools AW and AU). There was a broad recognition that the time allocated to PSHE was insufficient (schools AD, AL, AO, AT, AW, AX, BD and BE):

I don't think you can do enough in that short period of time [20 minutes, twice a week]. I think again, it's almost kind of doing it because . . . we're obliged to cover the material. I think if it's a real focus and a genuine focus, I think more time should be devoted to it.

Head of learning, school AX, interview, year 1

Tokenism was a prominent subtheme. Teachers in some schools reported that although PSHE in their schools addressed health and well-being, this was somewhat tokenistic, with the school's overwhelming focus being on attainment:

I suppose the head teacher superficially would verbalise that, and by signing up to this project has put some resources behind it. But in practice, politically it isn't easy to get the time or the resources towards that.

Assistant head-teacher, school BD, interview, year 1

I think it comes down, unfortunately comes down to exams, it comes down to . . . It is a very short term . . . you know, if we drop this for an hour we can have an extra maths lesson or an extra English lesson.

Head of department, school AW, interview, year 1

Student participation in decision-making

Student participation in decisions was a recurring theme across schools. Staff reported in interviews that almost all schools had some kind of student council and some had more than one group for discussing student views (schools AF, AK, AL, AS, BC and BE). In some schools, students were invited to share their opinions via surveys or formal discussions with staff (schools AD, AF, AS, AU, AX and BI) or informally (schools AS, BC and BN). Three schools reported no such mechanism (schools AM, BE and BI). Schools aimed either to involve a cross-section of students, including those with a history of misbehaviour (schools AD, AK, AL, AS, AW, AX, BD and BI), or to over-represent students from lower socioeconomic groups (school AS) or those exhibiting challenging behaviour (schools AM, AX and BI).

Intervention implementation

Intervention fidelity

Fidelity per school

Overall fidelity was variable, with a reduction in the fidelity of formal intervention activities in year 3 (see *Appendix 1, Tables 37 and 38*). The median fidelity score for years 1 and 2 (maximum possible score of 8) was 6 (interquartile range 5–8), whereas for year 3 (maximum score of 4) the median was 1 (interquartile range 0–3). Two schools achieved perfect fidelity in years 1 and 2, six scored 7 out of 8 points, another six scored 6 out of 8 points and the remaining seven scored below this. Two schools achieved perfect fidelity in year 3, whereas another three scored 3 out of 4 points. In year 3, 15 schools sustained RP. Interviews with AG members and focus groups with staff in case study schools suggested that, in year 3, schools commonly incorporated what they regarded as the most useful AG functions into mainstream school structures and processes, for example involving students in decision-making and further reviewing policies to ensure that these supported RP and were integrated into existing school committees. Regarding components, training and AGs were delivered with good fidelity. The fidelity to the curriculum was lower. Schools AF, AM, BE and BI did not engage with the trial in year 3 beyond some teachers' continued use of RP. A higher fidelity score for years 1 and 2 was associated with lower rates of bullying victimisation at 24 months but not with decreased rates of aggression. The fidelity score for year 3 was not associated with either primary outcome.

Overall fidelity per facilitator

The fidelity of facilitators was judged for those elements of the intervention which they facilitated (in years 1 and 2 only), expressed as the mean score achieved 0–1 across these two years. Most facilitators achieved mean scores in the range 0.55–0.83 (*Table 9*). Two facilitators scored 1 in their joint work across three schools. One facilitator scored 0.33 working with one school and another scored 0.25 working across four schools. One facilitator was replaced after year 1 and, as a result, this individual's schools were redistributed to other facilitators.

RESULTS

TABLE 9 Overall fidelity per facilitator

Facilitator	Number of schools worked in	Mean fidelity score for their AGs			Mean score overall/1
		Achieving minimum six meetings	Reviewing policies/rules	Implementing locally decided actions	
1	3	0.33	1	1	0.77
2	2	0.5	1	1	0.83
3	1	0	1	0	0.33
4	1	1	1	0	0.66
5	4	0.25	0.5	0.75	0.25
6	2	1	0.5	1	0.83
7 replaced by 8	3	1	1	1	1
9	3	0.66	1	0.66	0.77
7 replaced by 2	1	0	1	1	0.66

Fidelity of training in restorative practice

According to the trainers' diaries, there was very good fidelity of delivery of the all-staff training (*Table 10*). Data were available for 19 of the 20 schools and, for these, fidelity was as intended, other than less than intended coverage for the three sessions on maintaining relationships after difficult conversations, plus one session in which the intended DVD was not shown.

In total, across all schools, 1878 staff attended all-staff training and 131 staff attended in-depth training. Fifteen schools met the target of having at least five staff members receive in-depth training in using RP, over 3 days (schools AD, AE, AF, AH, AK, AL, AO, AT, AU, AW, AX, AZ, BE, BC and BI). The protocol did not include assessment of the fidelity of in-depth training. A few schools (schools AF, AK, AT, BD, BI and BM) received the training late in the first year of the study, reducing the extent to which RP could be implemented in the first year of intervention. During all six observations of the all-staff training, researchers judged that participants were moderately to very enthusiastic but that the groups were too large. Participatory learning could then be achieved only through small-group or paired activities.

TABLE 10 All-staff training implementation

Training implementation	Fidelity self-reported in trainer diary, <i>n</i> (%) of schools			
	As intended	Less than intended	Not at all	Schools with no data
Covered 'what is restorative justice/practice'	19 (95)	0	0	1 (5)
Covered 'the importance of language'	19 (95)	0	0	1 (5)
Covered 'what we do to challenge bad behaviour/nature of challenge'	19 (95)	0	0	1 (5)
Covered 'the importance of emotions'	19 (95)	0	0	1 (5)
Covered 'the importance of listening'	19 (95)	0	0	1 (5)
Covered 'maintaining the relationship after difficult conversation'	16 (80)	3 (15.79)	0	1 (5)
Used PowerPoint slides	19 (95)	0	0	1 (5)
Used DVD	18 (90)	0	1 (5)	1 (5)
Facilitated paired activity	19 (100)	0	0	1 (5)
Facilitated small-group activity	19 (95)	0	0	1 (5)

Fidelity of needs assessment surveys

Needs assessment surveys were completed in all schools, with generally high response rates (*Table 11*). In year 1, 11 schools achieved student response rates of $\geq 90\%$, six achieved response rates in the range 80–89% and three achieved response rates in the range 70–79%. In year 2, these figures were 10, 7 and 3, respectively. In year 3, five schools achieved responses of $\geq 90\%$, 10 achieved responses in the range 80–89%, three achieved responses in the range 70–79% and four achieved responses $< 70\%$. According to facilitator diaries and meeting minutes, all schools in year 1 and 14 schools in year 2 reported using the needs data to inform priorities and actions. In year 3, four schools (schools AL, AO, AT and BD) reported using the needs data to inform decisions. Three more read the report and used it for discussions (schools AH, AS and AU) and the two schools that most disliked the survey reported that they found the resulting needs assessment report useless (schools AE and AU).

Fidelity of action group meetings

In the first year, 19 schools completed all six AG meetings (AGMs) as intended and only one (school AM) convened only four meetings. In the second year, 11 schools completed six AGMs (schools AD, AE, AH, AK, AL, AS, AU, AX, BC, BD and BE), four held five AGMs (AF, AT, AW and BI), three held four AGMs (schools AM, AO and BM), and one held three AGMs (school BK). Thus, 11 schools achieved the target of holding at least six AGMs per year in both years 1 and 2 (schools AD, AE, AH, AK, AL, AS, AU, AX, BC, BD and BE). Of these schools, only AE and AS maintained perfect AG fidelity into the third year of the intervention.

In year 3, several schools (schools AF, AM, BE and BI) did not hold any AGMs, whereas other schools either used AGs for quite different purposes or used other meetings to implement the intervention. Schools AK, AO and AZ began to use AGs to train students or staff in RP or mentoring. Schools AE, AK and AT used student council or student voice committees to fulfil the function of an AG. Seventeen schools succeeded in reviewing school policies and rules during year 1 or 2 (schools AD, AE, AH, AK, AL, AM, AO, AT, AU, AW, AX, AZ, BC, BE, BI, BK and BM). In year 2, four changed school rules (schools AE, AH, AM and AO). In year 3, five schools reviewed their behaviour policies (schools AD, AL, AX, BC and BD).

Most AGs reviewed their rules or policies relating to behaviour (schools AD, AF, AH, AO, AU, AW, AX, AZ, BD and BE) or rewards and sanctions (schools AX, AZ, BC and BM). A recurring theme in interviews with staff members and facilitators was that the school focused considerable energy on revising school policies informed by restorative principles (schools AH, AL, AM, AO, AU, AW, BD and BI):

The behaviour policy has been rewritten now. So restorative practice is part of it. Interactions with parents, interactions with students . . . we're going to do staff training in terms of how do you talk to parents, how do you talk to students, language used in the classrooms. All [newly qualified teachers] now have training in restorative practice when they arrive at the school . . .

Senior leader, school AW, interview, year 2

TABLE 11 Needs assessment survey implementation

Needs assessment surveys in intervention schools	Number (%) of schools		
	Year 1	Year 2	Year 3
Conducted	20	20	20
Schools with response rate in the range of			
< 70%	0 (0.0)	0 (0.0)	4 (20.0)
70–79%	3 (15.05)	3 (15.0)	1 (5.0)
80–89%	6 (30.0)	7 (35.0)	10 (50.0)
$\geq 90\%$	11 (55.0)	10 (50.0)	5 (25.0)

RESULTS

Sixteen schools enacted local decisions in years 1 and 2 (schools AD, AE, AF, AJ, AK, AL, AM, AO, AT, AU, AW, BC, BD, BE, BK and BM). In year 3, 10 schools completed locally decided actions (schools AE, AH, AK, AL, AO, AS, AT, AW, BC and BD). Common locally decided actions included cascading RP training to staff who had not attended the in-depth training, or to students so that those students could work as peer mentors or buddies (schools AH, AK, AS, AU, AW, AX, AZ, BC, BD, BI, BK and BM) or resolve conflicts themselves (schools AS, AU and BM). Others included delivering assemblies on RP (schools AE, AI, AO, AS and BM), instituting specific safe spaces on the school site (schools AF and AW) and various approaches to encouraging girls' empowerment (school BD). Safety concerns identified by students were addressed in some schools by having staff patrol hallways between classes (schools AF, BD and BI) to discourage aggressive behaviours. Some schools offered more after-school clubs (schools AE, BD, BK), and one offered drop-in services (school AS) to improve school engagement and social and emotional health among students. School BK funded new, external, specialist staff to work with students to improve mental health and well-being, including a counsellor and a boys' boxing coach. Many of these activities specifically aimed to broaden students' social circles through engaging with people in other year groups. Other schools made improvements to the physical environment, including decorating the schools with informational or motivational posters (schools AD, AO and BI) and displaying student work (school AS).

According to aggregate data from our survey, all AGs were assessed by their members in years 1 and 2 as having a good or quite good range of staff from across the school, and a good or quite good range of students of different backgrounds and academic ability (see *Appendix 1, Table 37*). All were assessed by their members in terms of it being definitely or partly true that there was someone from this school on the AG who co-ordinated it and showed leadership.

A rare theme was adaptation of, or deviation from, the intervention manual. One selective school (school AU) did not view itself as having significant problems with bullying or aggression and, therefore, used the AG primarily as a way to revise the homework policy to try to reduce the stress on students arising from receiving multiple homework assignments on the same day. They held double-meetings so that students were removed from lessons less frequently (school AU). Two schools broke the AG into subcommittees (schools AW and BD). This was done in one school so that students were less intimidated to speak in front of a large group and in the other school so that multiple groups could get more done. One school held only one AG meeting in year 2 because its student council already brought students and staff together to make collective decisions and review rules and policies (school AZ).

Schools AS and AD did not formally review their rules as part of the intervention but did incorporate the restorative sentiment into their new interpretation or enforcement of the rules:

So whilst there isn't an individual rule that has been changed, many policies and practices as a result of the project so far are being flexed and adapted and tweaked accordingly.

Deputy head-teacher, school AS, interview, year 1

Among the AGMs we observed, some were noticeably well led (schools AD, AE, AM, AO, AU and BI), whereas others seemed poorly led (schools AH, AM, BD and BK). Researchers noted that in three schools in particular, students' comments or concerns did not appear to be taken seriously (schools AH, AM and BD).

Fidelity of restorative practice

Because response rates to the surveys of those trained to deliver RP were very low, we instead assessed the extent of RP by analysing the responses to the 24-month all-staff survey question about RP. In 12 schools (all but schools AF, AH, AK, AM, BD, BE, BI and BM), at least 85% of staff reported that if there was trouble at their school, staff responded by talking to those involved to help them get on better. Qualitative research generally suggested that, in the course of the 3 years of the intervention, there was a continuing increase in the use of restorative approaches. The vast majority of staff reported in surveys that teachers and students got together sometimes or often to build better relationships, or to discuss their views and feelings.

Interviews and focus groups suggested that variability of the form and scope of the RP implemented was a key theme. The majority of schools were reported to be regularly using some form of RP (schools AD, AE, AK, AL, AO, AT, AU, AX, AZ, BD, BI and BM). In some schools, key staff members, particularly the trial champion and pastoral staff, used restorative approaches, but felt frustrated that the school was generally slow to adapt and normalise such work (schools AM, AO, AW, AX, BI and BK), often because the broader culture of the school staff was not supportive or because these schools lacked the management capacity to ensure that restorative approaches were more widely adopted. Schools AU, BD and BM used RP primarily as a strategy to manage classrooms and learning behaviour.

Fidelity of social and emotional education

Response rates to our annual survey of those responsible for delivering the curriculum were poor, so we drew on both this and interviews with curriculum leads to gauge the fidelity of curriculum delivery. This established that in year 1 only 12 schools delivered the curriculum as intended in the protocol (5 hours' total teaching time or teaching comprising unit 1 as well as one or more additional units) (schools AD, AE, AF, AH, AM, AS, AU, AW, AX, AZ, BC and BE) and in year 2 half as many did so (schools AD, AE, AH, AS, AW and AZ). In year 3, only five schools delivered this threshold of teaching time (schools AD, AL, AS, AT and BC). The curriculum was taught in various forms across the schools, including in 10- to 20-minute sections in tutor time, rarely in specific PSHE lessons, and quite often in off-timetable days focused on PSHE topics.

Many schools did not deliver the curriculum materials as they were packaged. Staff in several schools commented that they needed to adapt the curriculum because it did not address the needs of their students, or because the materials were too simplistic (schools AD, AE, AF, AK, AS and AW). There was some evidence that some schools treated the delivery of the curriculum as a box-ticking exercise. They implemented the minimum required without ensuring that implementation was strong enough to maximise its potential impact. For example, at the end of year 1, a SLT member in school AW admitted that the school had not ensured that the curriculum was delivered by the most appropriate teachers:

Our strongest teachers aren't our PSHE teachers and our ... you know ... RE teachers, they're not really our ... our best teachers in the school ... if we're really going to place importance in it, how important are we going to think about delivery of it? It's all very good going through the motions, but unless you do it properly there's no point doing it.

Assistant head-teacher, school AW, interview, year 1

Participation and reach

We drew on facilitator diaries, AG minutes and training attendance sheets to estimate the numbers of staff and students participating in training and AGs. All-staff training on average reached 103 staff members (range 34–162 staff members) per school in schools for which data were available. In-depth training on average reached 6.35 staff members (range 2–13 staff members) per school in schools for which we have data. Fifteen schools sent at least five staff members on this training (see *Appendix 1, Table 39*).

On average in year 1, 7.4 staff members (range 3–12 staff members) per school participated in AGs in schools for which data were available, with this falling to 5.4 staff members (range 2–9 members) per school in year 2. On average in year 1, 7.8 students (range 5–14 students) per school participated in AGs in schools for which we have data, with this increasing to 8.8 students (range 4–23 students) per school in year 2. We did not calculate figures on participation in year 3 because schools implemented AGs in such heterogeneous ways.

Schools were encouraged to, and most did, select a diversity of students to serve on the AG, including those with a history of misbehaviour (schools AD, AF, AL, AO, AT, AW, AX, BC, BK, BI and BM), those who struggled academically (schools AD, BC, BI, BK and BM) and those who were at risk of disengagement (AK and BC). Other schools focused recruitment on students who were high achievers

RESULTS

(schools AK, AS and AW). These schools sometimes dropped lower-attaining students from AGs. Most schools selected students across school years to participate, but some (schools AF, AL and AW) invited only those in the study cohort.

Just over one-third of staff completing the all-staff survey at the end of year 2 reported that students were involved in developing the behaviour policy or school rules, with this rising slightly by the end of year 3 (Table 12). At the end of years 2 and 3, only about one-third of students reported being aware that the school had been taking steps to reduce bullying. By the end of year 2, around one-third of students reported that they had experienced lessons about relationships or emotions, with this falling to around one-quarter by the end of the third year (Table 13). Only around one-quarter of students reported that students had been involved in developing the behaviour policy or school rules. Across both time points, only about one-sixth of students knew what was meant by the term 'restorative practices'. About half reported that, if there was trouble at school, staff responded by talking to those involved to help them get on better. About two-thirds of students reported that teachers and students got together to build better relationships or discuss their views and feelings. Focus groups with students not involved in AGs suggested that there was broader awareness of the work of the AG among students in some schools (schools AE, AF, AH, AW and BK) more than others (schools AO, AS, AT, AZ, BC and BD).

Even if the AG itself was not well known, students on AGs in some schools reported that their peers had noticed changes being made but were unaware of how these had come about (schools AD, AT and BK):

They're aware of certain changes. But they're not really aware of why people are doing this. So they see it as, why have they put up a basketball hoop? Oh, that's probably for people to play on. They don't see why . . . what caused it to happen. It's only people on the action group and some teachers are aware of it. Everyone else is oblivious to the situation.

Male, year 9 student, school BK, interview, year 1

Slightly over half of staff in intervention schools were aware that the school had been taking steps to reduce bullying and aggression, with this falling slightly between years 2 and 3 (see Table 12). In interviews with SLT members in year 3, some reported that although the name of the intervention was sometimes not known by all staff members, the push for increased use of RP was more widely known.

Reception and responsiveness

Among the staff who were aware that the intervention was occurring, the vast majority were supportive overall at both time points (see Table 12). With the increasing expectations for educational attainment, some staff resented the work involved in the intervention given its primary focus on health and well-being rather than attainment:

One of the challenges that I foresee is that . . . I've worked at a few schools and I don't think this school is unusual in that teachers feel responsible for academic progress and they don't . . . feel responsible for student well-being in the same way. And I think that's a result of accountability measures being focused around those things and so the challenge is to move that perception so that staff take ownership of all of those things. But also then to help them to understand that the two things obviously are linked.

Assistant head-teacher, school BD, interview, year 1

Commitment to the intervention was notably lower in schools where the pre-existing approach to behaviour was disciplinarian or not supportive of student participation in decision-making (schools AH, AT and BD).

TABLE 12 Staff awareness of the intervention processes

Measure	Staff responses					
	24 months			36 months		
	Overall % across intervention group	Number of schools where school % is > 70 %	Number of schools with no data	Overall % across intervention group	Number of schools where school % is > 70 %	Number of schools with no data
This school has recently been taking steps to reduce bullying and aggression (% yes)	58.33	6	1	54.21	4	3
I support this new work (% yes)	91.71	19	2	91.40	20	3
I understand what is meant by 'restorative practice' (% yes)	77.90	14	1	79.52	13	3
I support using restorative practice in schools (% yes)	52.85	4	1	55.38	3	3
If there is trouble at this school, staff response includes talking to those involved to help them get on better (% yes)	85.99	16	1	89.14	17	3
Teachers and students at this school get together to build better relationships (% often/sometimes)	94.55	20	1	94.06	20	3
Teachers and students at this school get together to discuss their views and feelings (% often/sometimes)	83.95	19	1	82.39	17	3
At this school, students were involved in developing the behaviour policy or school rules (% yes)	38.72	1	1	44.05	2	3

TABLE 13 Student awareness of the intervention processes

Measure	Staff responses					
	24 months			36 months		
	Overall % across intervention group	Number of schools where school % is > 70%	Number of schools with no data	Overall % across intervention group	Number of schools where school % is > 70%	Number of schools with no data
This school has recently been taking steps to reduce bullying and aggression (% yes)	34.29	0	0	33.00	0	0
This past year in class, we have been learning how to get on well together (% yes)	37.68	0	0	26.98	0	0
This past year in class, we have been learning how to manage our emotions (% yes)	33.28	0	0	24.52	0	0
This past year in class, we have been learning how to resolve conflict (% yes)	38.72	0	0	28.49	0	0
I understand what is meant by 'Restorative Practice' (% yes)	15.42	0	0	13.52	0	0
If there is trouble at this school, staff response includes talking to those involved to help them get on better (% yes)	53.69	0	0	50.65	0	0
Teachers and students at this school get together to build better relationships (% often/sometimes)	69.68	12	0	67.85	9	0
Teachers and students at this school get together to discuss their views and feelings (% often/sometimes)	63.48	7	0	59.84	5	0
At this school, students were involved in developing the behaviour policy or school rules (% yes)	28.84	0	0	24.08	0	0

Needs assessment surveys

The undertaking of the needs assessment surveys was acceptable to staff and students, evidenced by the high response rates across all schools. The reception given to the reports from these surveys is discussed under AGs.

Training

Those receiving in-depth training reported high satisfaction (Table 14). Participants across all intervention schools reported that they 'probably' or 'definitely' had learned useful skills, and all indicated that they would recommend the training to colleagues. Over 90% of participants across all schools reported they intended to put into practice the skills learned, felt confident putting into practice the skills learned, and rated the training as good or excellent. Almost 90% rated the training as excellent.

In interviews, staff were very positive about both the all-staff and the in-depth training, with some describing it as 'brilliant' (schools AX and AZ), 'transformational' (school AW), and helpful with behavioural issues:

I've been on two trainings this term and they're the best two trainings I've ever been on in my life. They really, really, really, really helped.

Role unknown, school AH, interview, year 1

Brilliant. Hands down, absolutely amazing. I think every single student, every single teacher should go through that, definitely. It's been absolutely phenomenal training. And from everybody that I've had that's gone through has found that incredibly useful and are really behind it.

Head of year, school AZ, interview, year 2

Only a few reports about the training were less positive. In three schools (schools AD, AK and AM), some staff were frustrated about the length of the in-depth training, regarding it as condescending to experienced teachers. Conversely, staff in some schools thought that the shorter all-staff training was too short to have a sustainable impact on staff behaviour (schools AD, AO and BK). Some staff sitting on AGs reported that choosing who should attend the in-depth training was a fraught process. A staff member in one school felt that staff were chosen without enough consideration of their roles (school BE). In two schools, an interviewee suggested that too many SLT members attended the training despite having less contact than other staff with students (schools AD and BD). In another school, a staff member suggested that too many teachers and not enough senior leaders were trained, thereby eroding the extent to which training informed school management culture (school BI). In other schools, AG members reported that trainees were selected appropriately (schools AE, AF, AH, AK, AM and BM). Schools AF, AK, AT, BD, BI and BM received the training late in the first year, thus reducing the time available to put RP into action.

TABLE 14 Participant satisfaction with the in-depth training

Measure (response option that indicates satisfaction)	Responses	
	Overall % members responding as indicated	Number of schools where > 70% of members responded as indicated
Do you feel that you learned useful skills at this training? (yes definitely or yes probably)	100	20
Do you feel confident in putting into practice the skills you have learned today? (yes)	91	20
Do you intend to put the skills you learned to use in your everyday practice? (yes)	95	20
Would you recommend this training to a colleague? (yes)	100	20
Overall, how would you rate this training? (good or excellent)	99	20

RESULTS

Many schools decided to cascade the training so that some of those receiving the in-depth training then trained other staff. Thus, all staff were trained in some schools (schools AE, AK, AO, AZ, BI and BK), whereas others focused this training on key personnel (schools AF, AM, AS and BM) or newly qualified teachers (school AW). Half of the intervention schools trained students to resolve conflict using restorative techniques (schools AE, AK, AM, AO, AS, AW, AZ, BC, BD and BK). In year 3, four schools trained student conflict mediators and anti-bullying mentors (schools AK, AO, AZ and BD). A problem noted across the process evaluation was the tendency for staff who had received the in-depth training to leave the school before the intervention was complete. This meant that schools with high rates of staff turnover had a shortage of staff who had been trained in RP.

Action group meetings

Use of needs data

Over two-thirds of AG members completing the survey reported that they worked hard on the AG because they wanted to, not because they had to (Table 15). Over 80% reported looking forward to meeting and almost 70% reported that meetings were exciting and energising. Our survey of AG members found that most rated the needs assessment report as useful for informing decisions, with the overall percentage of members rating it thus declining slightly from 91% across the 19 schools providing data in year 1 to 88% across the 17 schools providing data in year 2 to 76% among the three schools providing data in year 3. These results appeared broadly consistent across schools, with 17 averaging 70% positive ratings in year 1 and 14 doing so in year 2, and two of the three schools completing the survey in year 3.

In interviews with students and staff on AGs, as well as with facilitators, most reported that they valued the needs reports and used them to identify priorities (schools AD, AE, AF, AH, AK, AL, AM, AO, AS, AT, AU, AW, AX, BC, BE, BI, BK and BM):

And they've been drawing on the needs assessment report research to help them think about that as well. It's feeding into everything they do. So they . . . we focused on it quite a lot in meetings, pored over the data in the first few meetings. And then we still keep coming back to it when we're thinking about what next, what actions should we be doing? Yeah, so they've . . . yeah, they've taken it to heart and really want to make a difference.

Facilitator, school BM, interview, year 1

Staff in most schools suggested in interviews that they used the reports as evidence to advocate to school leadership teams that changes were necessary. In school AH, for example, the needs assessment was useful for staff advocating on behalf of students whose concerns, although previously voiced, had not been acted on by school leaders. This was also the case in school AF:

But it certainly gave me some ammunition that I can say, well this is . . . this is proven because the kids have actually said this, so we need to move forward with it.

Assistant head-teacher, school AF, interview, year 2

In some schools (schools AF, AO, BE and BM), staff expressed shock or surprise at the results, particularly about the data on student perceptions of feeling unsafe at school or feeling uncomfortable talking to staff about problems. The reports often highlighted student concerns that staff had previously never considered, particularly relating to feeling safe at school (schools AH, AW and AZ):

I've always felt it quite a safe area and a lot of the young people don't feel it's particularly safe because of the way they have to come in and out. And obviously that's being changed now . . . So I think it's really been useful for us to do that. And quite . . . eye-opening and telling as to the areas that as adults we thought we'd covered and actually, no, the young people don't think so much.

Head of year, school AZ, interview, year 1

TABLE 15 Member satisfaction with AGs

Measure (response option that indicates satisfaction)	Responses								
	Year 1			Year 2			Year 3		
	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data
Was the needs assessment report useful in helping the AG decide what actions to take? (somewhat or very)	91	17	1	88	14	3	76	2	17
Was the external facilitator useful in ensuring that all AG members could have their say? (somewhat or very) ^a	89	18	1	90	15	3	NA	NA	NA
Was the external facilitator useful in helping the AG decide what actions to take? (somewhat or very) ^a	86	16	1	88	17	3	NA	NA	NA
Was the external facilitator useful in helping to ensure that actions were actually implemented? (somewhat or very) ^a	80	13	1	79	13	3	NA	NA	NA
Do you think the LT project was a good way to ensure that students contribute to decision-making at this school? (very or quite)	95	19	1	94	16	3	100	3	17
Overall do you think the AG made good decisions about what actions to take? (% very or quite)	94	19	1	93	16	3	98	3	17
Do you think the AG made sure that these actions were implemented? (% yes)	70	10	1	72	10	3	69	1	17
NA, not applicable.									
a No external facilitator in Year 3.									

© Queen's Printer and Controller of HMSO 2019. This work was produced by Borell et al. under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts for individual use provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising, promotional or commercial reproduction should be addressed to: NHS Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

RESULTS

Another school received a report that presented evidence of considerable unmet student needs and this information was accepted as valid by school leaders:

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

Director of Inclusion, school AM, interview, year 1

There were some more negative views expressed. Students or staff in eight schools reported that the reports were difficult to understand (schools AF, AK, AM, AX, AZ, BD, BI and BK). Staff in schools AH, AO and BK misunderstood the reference group to which data on students in their school were being compared, thinking that this was all schools nationally rather than other schools in the intervention arm of the trial. During an observation, it was apparent that the lead staff member in school BK did not understand the concept of statistical significance and needed support to interpret the colour-coding system depicting this.

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

The data from the [your report] is meaningless; completely meaningless . . . It's the worst . . . it's a pointless exercise for our students.

Assistant head-teacher, school AE, interview, year 3

Other schools disagreed with some elements of the needs report (schools AD, AL, AT, AU, AW, AX and BC). As with schools BK and AE, school AX reported that the data provided through the study contradicted its internally maintained data. Staff in three schools (schools AF, AS and BD) did not share some or all of the findings with students because the report was seen as either too long and complicated or inappropriate for students to see. When year-on-year trends in such factors did not improve, staff members could sometimes feel dispirited (schools AF and BI). Such disappointments may have contributed to these schools not continuing intervention activities in the final year of the study.

Facilitators

Most AG members rated the facilitator as useful in ensuring that all members could have their say, with the overall percentage of members rating it thus being maintained at 89% across the 19 schools providing data in year 1 and 90% across the 17 schools providing data in year 2 (see *Table 15*). These results appeared broadly consistent across schools, with 18 schools averaging > 70% positive ratings in year 1 and 15 averaging this in year 2. Most AG members rated the facilitator as useful in helping AGs make decisions, with the overall percentage of members rating it thus being maintained at 86% across the 19 schools providing data in year 1 and at 88% across the 17 schools providing data in year 2. These results appeared broadly consistent across schools, with 16 schools averaging > 70% positive ratings in year 1 and 17 out of 17 averaging this in year 2. Most AG members rated the facilitator as useful in helping to ensure that decisions were implemented, with the overall percentage of members rating it thus being maintained at 80% across the 19 schools providing data in year 1 and at 79% across the 17 schools providing data in year 2. These results were less consistent across schools, with 13 averaging > 70% positive ratings in year 1 and year 2.

However, a theme across interview data was many staff members' mixed feelings about the facilitation. Some schools reported that the facilitator was relatively passive, listening and taking notes (schools AD, AS and AX), whereas others felt that the facilitator simply did not contribute much or were generally dissatisfied with the facilitator's work (schools AF and BK):

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

Role unknown, school AF, interview, year 1

One facilitator was replaced between years 1 and 2, in part because of school dissatisfaction (schools AE, AU, BE and BK). Several of these schools (schools AE, BE and BK) went on to disengage from the intervention or expressed their dissatisfaction with it, although the balance is unclear between the poor facilitation causing the disengagement or the disengagement coming first and informing the criticism. Other schools were not clear about the role of the facilitator and, therefore, had trouble setting reasonable expectations of what the facilitator should or was able to contribute to the group:

I think . . . [facilitator] has been really useful as a . . . So making sure we're doing what we're supposed to do in one respect is great. But then it's sometimes flipped into, well are we doing it right. And having [facilitator]'s presence there has always . . . are we being watched, judged or supported? We're not quite sure where that . . . what that role is.

SLT member, school AW, Interview, year 2

However, there were also positive feelings. Staff in some schools (schools AE and AT) reported that the facilitator added weight or gravitas to the meetings and reinforced the notion that the school would be held accountable for progress:

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

Staff member, school AE, interview, year 1

There were mixed reports on what difference the absence of external facilitation made to the third year of the intervention. Most interviewees suggested that external facilitation was not necessary in the final year but a few suggested that this was a significant loss:

The absence of [facilitator] has been incredibly significant because she did, she has, she was able to tie it in all the time to the agenda. And was a touchstone I suppose really for that. And then . . . so that . . . I think that was a loss.

Deputy head-teacher, school AD, interview, year 3

Student participation

Most AG members agreed that the AG was a good way to ensure that students contributed to decision-making, with the overall percentage of members rating it thus being maintained at 95% across the 19 schools providing data in year 1, at 94% across the 17 schools providing data in year 2 and at 100% among the three schools providing data in year 3 (see *Table 15*). These results appeared broadly consistent across schools, with 19 schools averaging > 70% positive ratings in year 1, 16 in year 2 and 3 out of 3 in year 3.

RESULTS

A recurring theme in student accounts was students joining the AG because they wanted to help change their school (schools AE, AF, AT, AW, BE and BK) or to have more of a voice (schools AD, AM, AT, AW and BC):

Because I felt sometimes frustrated with how the school was sometimes. It felt like a chance where I can say what I want to say now and maybe make a difference so that it wouldn't be like that anymore, I guess.

Female year-9 student, school AO, interview, year 2

Students valued the opportunity to express their views and highlight where they were unhappy, and AGs provided a calm, structured environment to do this:

Male, year 9 student, school BE, interview, year 2: I was able to speak from my own point of view not just like statistics and what was on a piece of paper. Because I was able to like . . . relate to it in a sense. And . . . like I started . . . talking about my own experiences, the teachers were acknowledging it and . . . yeah.

Interviewer: Did that feel good . . . ?

Male, year 9 student, school BE, interview, year 2: Yeah. It kind of felt good because it's like . . . it's not me having to shout at a teacher or like I want to do this, dadadada. But it was me actually being able to relate to a teacher seeing why if I'm shouting and I'm cussing it's difficult for a teacher with a class of 30 or more students.

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

That was good because we didn't feel intimidated by the teachers. Not that you felt intimidated before, but you didn't feel like . . . You could say something, all the teachers would listen. If they contradicted, they contradicted, if they agreed, they agreed. But no one disagreed, they're going to say you're wrong, but everyone put across their own views with no arguments or anything.

Student of unknown sex and year, school BE, focus group, year 2

Students could grow in confidence in the course of their involvement on the AG. This was most obvious in student accounts from schools AM and BM:

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

Student of unknown sex and year, school BM, focus group, year 1

However, a theme across some staff accounts was struggling to manage good interactions between staff and students. Facilitator and interviews suggested that although many students participated meaningfully, others remained disengaged or, at times, combative:

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

Facilitator, school BD, interview, year 2

Interviews and focus groups suggested that whereas in many schools student views were described as being taken seriously or very seriously (schools AE, AO, AS, AT, AX, AZ, BI and BM), in other schools staff took note of student views only when these were deemed 'realistic' (schools AD, AF, AK, AL, AU, AW and BE). Staff in some schools reported not really valuing student input or that student participation led to conflicts with staff (schools AD, AH and AK). During the interview with a staff member at school AH, these tensions became apparent:

They [students] want some rules to change; so they want to be able to bring their mobiles . . . There were a lot of conversations about, 'no . . . that's not [going to happen] . . .' [. . .]. And I think that they think they had a bigger input than they perhaps did.

Assistant head-teacher, school AH, interview, year 2

The above quotation suggests that students' participation was, at times, tokenistic. The same staff member continued:

And I guess the limitation is the things that they think . . . having phones will make us feel safer at school, and I know that our . . . part of our health and safety is that no, phones actually make young people more vulnerable. So we say no mobile phones at school. So there are some things which are just . . . not going to happen – so it's a brick wall. And they don't like that either. So there is conflict. I always try and say, well I'll take it to the principal, so that they don't feel like I'm dismissive.

Assistant head-teacher, school AH, interview, year 2

Taking account of student views could be a challenge to schools' normal way of working, as one teacher explained:

I think the challenges are twofold, from two different sides. The first from the teachers' side is that you want to include students but often they're not equipped to make good decisions for themselves . . . I think from the student side the students often struggle to feel that actually what they say has meaning, because it's so often just ignored. And therefore I think they can become disillusioned and not really feel it's worth contributing.

Teacher, school AD, interview, year 1

One staff member reported finding it difficult to take a facilitative approach to managing student inputs:

I'm aware my role is of a facilitator you know, and luckily I've done some courses when it comes to teaching about taking a step back; because most teachers want to fix stuff, that's what we do. And we just think we can fix everything and we'll tell you what to do and you'll do it.

Assistant head-teacher, school AF, interview, year 2

However, other schools were enthusiastic about students' participation in decisions. In school AU, staff on the AG expanded the process of revising school rules so that this involved students beyond the AG alone:

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

Assistant head-teacher, school AU, interview, year 2

RESULTS

A recurring theme was the impact of the work on mutual understanding between staff and students, as these four quotations illustrate:

Yeah, because it made me feel like maybe . . . sometimes when teachers do stuff it's not because they want to do it, it's because of the rules that maybe they have set. But then when you're listening to them and they're agreeing with what you have to say, it's like they actually do understand and you can see that what I'm saying is being understood.

Female year-9 student, school AO, interview, year 2

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

Female year-10 student, school AU, interview, year 2

Yes [it made me feel differently about teachers], a hundred per cent because it give you a different insight to what they're really like, especially if you're with teachers who haven't taught you before or something like that. Because they're not as bad as you think.

Male year-11 student, school BM, interview, year 3

I know there is a hierarchy and there are imbalances in power, but with the Learning Together project I'm sat there with a child discussing from my point of view what I think, and from their point of view what they think. We're almost like putting it in the middle of the table. How can we make that better for both of us?

Role unknown, school AU, focus group, year 1

This could lead to better relationships between staff and students:

I think that's really good for me to be in something that's quite positive, because I do deal with quite a lot of negative in my role. So for the students to actually see me in a positive light within things like this I think is good for my role. Because then they'll stop and talk to me in the corridor: 'Miss, are you going to the meeting tonight? Yeah, yeah, are you going?' So it's nice just to have that. And other people see them talking to me and go 'Oh OK, she's not telling everyone off or talking about problems. She's quite fun actually'. And I think we've had a bit of fun in the meetings, had a little bit of a laugh, haven't we, as well?

Role unknown, school BM, focus group, year 1

Decision-making

Most AG members agreed that the AG made good decisions, with the overall percentage of members rating it thus being maintained at 94% across the 19 schools providing data in year 1, 93% across the 17 schools providing data in year 2 and 98% among the three schools providing data in year 3 (see Table 15). These results appeared broadly consistent across schools, with 19 out of 19 schools providing > 70% positive ratings in year 1, 16 out of 17 in year 2 and three out of three in year 3.

Action group members were less likely to agree that the AG made sure that these decisions were implemented, with the overall percentage of members rating it thus being maintained at 70% across the 19 schools providing data in year 1, 72% across the 17 schools providing data in year 2 and 69% among the three schools providing data in year 3. These results were also less consistent across schools, with 10 out of 19 schools providing > 70% positive ratings in year 1, 10 out of 17 in year 2 and one out of three in year 3.

A common theme in interviews with facilitators and staff was concern about the AG's capacity to implement actions (schools AF, AL, AM, AT, AW, AX and BK). According to the facilitator, school AF's AG repeatedly revisited minor issues, such as the school's policy on make-up, but was largely unable to confront more substantive issues relating to student behaviour, health or well-being. In school AT, the AG proposed numerous actions that were rejected by the head teacher. The amount of work already facing teachers made implementation more difficult. One facilitator said of the individual leading on AG that he had:

... been given so much to do I think in his day-to-day role, it's just another thing for him to do. He does it, he does it with very good grace and he's obviously very well respected by the students.

Facilitator, school BE, interview, year 2

Schools were better able to implement action when SLTs were represented on action teams or otherwise supported them. In some schools, the staff lead received virtually no support from colleagues (schools AH, AL, AM, AO, AT, AW, AX, BI and BK). In school AH, the staff lead had a committed AG lead but received little tangible support from the head teacher or other teachers:

The head teacher there was completely uninterested when I came to give a talk to the senior leadership team, and it's the only school where the head has not said anything to me or taken any notice or made any effort to be slightly friendly. It was awful. But I gave a talk to the SLT, they all sat ... they were terribly ... first of all, everybody there dresses extremely formally in business suits. And the head never said a thing.

Facilitator, school AH, interview, year 2

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

Throughout the 3-year study, the majority of schools handed over leadership of the project at least once (schools AF, AH, AM, AO, AT, AU, AW, AX, BC, BD, BE, BK and BM). Many went on maternity leave, changed roles or went on sick leave. Through these transitions, schools were sometimes able to transfer the programme to effective leads (schools AT, AW and BD), but some could not.

We used an adapted version of the Learner Empowerment Scale to assess the extent to which staff and student AG members felt empowered to make decisions. The mean percentage of positive responses across all items among AG members in intervention schools for which data were available was 75% in year 1, rising to 79% in year 2 and 80% in year 3 (Table 16). In year 1, 13 out of 19 schools that provided data had a rate of positive reports > 70% across all members and items. In years 2 and 3, these figures were 12 out of 17 and two out of three, respectively.

Restorative practice

Restorative practice was widely regarded as extremely useful for dealing with behavioural problems and conflicts. It appeared acceptable to most staff and students in most schools. Over three-quarters of staff responding to the all-staff survey understood the term 'restorative practice'. Around half supported the use of RP (see Table 12).

The presence of a staff member who championed the use of RP was critical to its broader use, and hence to its reach. Schools AO and AS had such a champion. In school AS, the staff member had a strong academic background in RP. He began working with students who misbehaved in class, expanding to use it with students who had particular pastoral needs, and then to the sixth form, and eventually to the entire range of year groups for which he was responsible. He set up voluntary training for all staff and unofficially trained students in restorative techniques. Having an important middle management role in the school and being an established and well-liked teacher placed this teacher in a strong position to champion RP.

TABLE 16 Learner Empowerment Scale for AG members

Learner Empowerment Scale item (response option that indicates empowerment)	Responses								
	Year 1			Year 2			Year 3		
	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data
During AGMs, I felt confident that I could do what was expected (yes)	90	18	1	91	14	3	98	3	17
I had the power to make a difference to how things were done at AGMs (yes)	78	14	1	82	10	3	89	3	17
The AGs work fits with what I believe in (yes)	88	17	1	92	17	3	89	3	17
My participation was important to the success of the AG (yes)	83	16	1	81	14	3	86	3	17
Other people on the AG made me feel like I was not good enough (no)	92	19	1	79	15	3	89	2	17
I actively took part in the tasks undertaken on the AG (yes)	90	19	1	89	16	3	91	3	17
I usually did more work than I had to do on the AG (yes)	23	0	1	34	0	3	33	0	17
I was overwhelmed by all of the work required of the AG (no)	75	14	1	74	15	3	86	3	17
I worked hard on the AG because I wanted to, not because I had to (yes)	83	17	1	83	14	3	84	2	17
I had a choice in the way I went about doing work on the AG (yes)	86	17	1	90	17	3	86	3	17

Learner Empowerment Scale item (response option that indicates empowerment)	Responses								
	Year 1			Year 2			Year 3		
	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data
The things I had to do on the AG meant a lot to me personally (yes)	61	5	1	74	10	3	69	1	17
I liked to talk about what I was doing on the AG with friends or family (yes)	55	2	1	61	11	3	66	1	17
I felt nervous about what was expected of me on the AG (no)	76	15	1	73	12	3	80	2	17
I was able to affect the way things were done on the AG (yes)	76	12	1	84	14	3	75	1	17
I looked forward to the AGMs (yes)	73	11	1	84	13	3	91	3	17
Those leading the AG believed that they must control how I contributed to the AG (no)	63	9	1	61	4	3	79	2	17
I got positive responses when I expressed my own attitudes and ideas on the AG (yes)	85	15	1	92	17	3	84	2	17
I agree with the things that were done on the AG (yes)	90	17	1	94	16	3	93	3	17
I had the skills needed to contribute to the AG (yes)	89	18	1	91	17	3	80	2	17
My ability to contribute to the AG was under my control (yes)	85	17	1	88	13	3	82	2	17

continued

TABLE 16 Learner Empowerment Scale for AG members (continued)

Learner Empowerment Scale item (response option that indicates empowerment)	Responses								
	Year 1			Year 2			Year 3		
	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data	Overall % of members responding as indicated	Number of schools where > 70% of members responded as indicated	Number of schools with no data
Those leading the AG felt that they were always right (no)	71	11	1	69	9	3	72	2	17
I found the AG to be exciting and energising (yes)	62	17	1	71	8	3	84	2	17
I found the AG to be interesting (yes)	87	17	1	90	15	3	86	1	17
The things I did on the AG were valuable to me (yes)	75	14	1	80	13	3	78	2	17
The things I learned and did on the AG will be helpful for my future (yes)	77	13	1	86	15	3	78	1	17
This AG was one of the first times I felt like I was able to contribute to something important (yes)	43	0	1	45	3	3	27	0	17
This AG taught me how to work well together with others (yes)	56	3	1	64	3	3	75	1	17
This AG helped me better understand some of the problems in my school (yes)	84	14	1	88	15	3	91	3	17
This AG gave me the chance to do something about the problems in my school (yes)	82	15	1	92	16	3	88	2	17
Mean across all items	75	13	1	79	12	3	80	2	17

Reproduced with permission from John Wiley and Sons from Warren et al.¹²⁶ © 2019 British Educational Research Association.

In some schools, RP was used by most of those staff who had attended the in-depth training but some of these staff felt frustrated that the school was not adopting RP quickly enough (schools AM, AO, AW, AX, BI and BK). Some schools where this was the case had cultural norms that acted as barriers to implementation, such as an authoritarian ethos, or did not have adequate commitment from the SLT to ensure more widespread adoption. Schools that faced challenges, particularly in terms of poor results or Ofsted ratings, often struggled to adopt RP fully because it was not seen as a priority for the SLT to lead.

Where it was used, staff generally felt that RP allowed them to resolve problems more easily, because it allowed them to focus on restoring relationships rather than being bogged down in culpability:

And before I found we were spending a lot of time in 'Billy said (a), Johnny said (b)'. And we're arguing about (a) and (b) forever and ever and you probably never get to the end of it. And now Billy gets to give his view, Johnny gets to give his view, and then we move on to how we're going to move on and what this is going to look like in the future.

Role unknown, school AO, focus group, year 1

In other schools (schools AU, BD and BM), RP was used primarily as a strategy to manage classroom and learning behaviour. This often took the form of requiring a restorative conversation before a student who had breached expected norms of behaviour could be readmitted to the classroom.

The RP sessions benefited from their being facilitated by trained staff. Many students acknowledged this:

Well I got excluded for overheated arguments and when I came back from exclusion [the teacher] sat with me and the boy and we kind of . . . because before, if we tried to resolve it ourselves it would just end up, like the same thing that happened, we start overheating each other and then someone would get hit. But when you're in restorative justice it's more calm. Because the teacher's there, the presence of the teacher or whoever's there, it helps calm down the situation and you actually listen to them. So I reckon that helped a lot.

Student of unknown sex and year, school AO, focus group, year 2

School AO often brought in parents or family members to restorative conferences and their presence could further calm the interactions that ensued. Students reportedly felt more pressure to be more honest and mature in such meetings, not wanting to embarrass their parents:

You have to act a lot more mature. You can't just be bickering and you know . . . let it get heated in front of your parents. You have to really think about what it is you're saying and what you mean.

Female year 10 student, school AO, RP interview, year 2

Students in general reported positive experiences of RP. Several students in schools AM and BE felt that restorative conferences could improve relationships where these had been conflicting. Students did not necessarily leave conferences as friends but were better able to manage their relationships:

I thought it [restorative conference] was useful because if we wouldn't have had it there would have been many, many, many more problems would have happened. It's not like it's completely gone. It's not because there's still a bit of tension between all of us . . . It's not as big as it used to be.

Student of unknown sex and year, school BE, interview, year 1

Sessions could force perpetrators to appreciate the hurt they had caused. A student with a history of bullying perpetration learned through a difficult restorative discussion the extent to which his bullying had hurt another boy in his class:

I just felt so sorry . . . the boy was just crying. When he was speaking he could barely speak and it just reminded me why did I do that? You shouldn't have done that. It just . . . made me realise . . .

RESULTS

And when we came in, it's just like . . . at first I was laughing . . . But then when I just saw him there sitting down at this table and his eyes were all red from the tears . . . I just don't . . . it just came to me and just shocked me. That could have that had happened to me really.

Male student of unknown year, school AM, interview, year 1

After a restorative conversation, a student in school BD not only forgave the student who had been mean, but also saw that that he was capable of change and personal growth:

I could tell he kind of regretted it, what he did. But at the same time, I saw that he could actually change. Bad people are not always that bad at the same time they can always change.

Male student of unknown year, school BD, interview, year 1

Social and emotional education

A recurring theme in staff interviews was the very mixed responses to the curriculum. There were some subthemes concerning positive evaluations. Some teachers appreciated the clear structure of sessions (schools AK, AL, AS and AZ). Many students also found the lessons acceptable because they provided a good opportunity for discussions (schools AE and AL) or felt that the topics covered were highly relevant to their life (schools AK, AS and AW). Some staff commented that the curriculum as with the RP sessions could provide students with the skills needed to resolve disputes themselves:

And training for students is important. Because if we can teach the students how to monitor and measure themselves, and use this process, that's where you're going to have the biggest breakthrough. Because if students are taught the skill of sitting down and talking through problems and have a structure to do it, that's where you're going to have a win.

Role unknown, school AO, focus group, year 1

Other staff commented that although some of the skills focused on in the curriculum were basic, they were nonetheless important:

So I was given . . . the learning objective around how we can get our students to work together more cohesively in teams. And it was really, really interesting because a lot of the lesson . . . initially when I looked at the lesson plan I thought, this is quite basic stuff, because as an adult you do just think it's second nature. [. . .]. What's always really interesting to see is that some of these things are not obvious to students. Like stopping to listen to someone when they're speaking, or taking roles in groups or . . . different things like that which seem really obvious to us, which was really lovely. And also they're very easily achievable learning outcomes from the lessons. Also the students are left with a real sense of being equipped, and they left with a sense of achievement in terms of having very easily achieved a learning outcome.

Trainee teacher, school BD, interview, year 2

However, another theme was how lessons could be improved. One concern was that the materials had been sent too late in the year to be incorporated into lesson plans (schools AE, AF and AU), which was repeated in all three years of the study. Teachers also said they felt they had to adapt the lessons because they were too simplistic (schools AD, AE, AF, AK, AS and AW) or so that they could be used as evidence for Ofsted inspections (school AK):

Increasingly, the further along with the units we went the more we used our own resources [. . .] we have a lot of PowerPoints and teacher notes, but not very much to evidence the students' learning. And that's a big focus for our school you know; we have to evidence their progression.

PSHE teacher, school AK, interview, year 1

Some teachers said that they adapted materials to make the curriculum more engaging for their pupils. This was often felt necessary because the lessons were viewed as unfocused, simplistic or boring (schools AF, AK, AS and AX):

Sometimes to make it . . . I had to make it sometimes a little bit more visual or include videos or try to . . . sometimes approach it from a different . . . in a slightly different way.

Teacher, school AX, interview, year 1

They [students] felt it was a little bit repetitive and they thought that some of it at least was a bit boring for them because it was relatively straightforward. It was things that they do anyway, things they think about anyway.

Teacher, school AF, interview, year 1

Teachers found it time-consuming to adapt lessons to the needs of their classes (schools AD, AE, AS and AW):

Teachers don't have time to rewrite things and . . . it does take a lot of time. You know planning an hour lesson can take up to 3 hours. And if you're then messing around with what someone else has done as well . . . it gets very complicated.

Teacher, school AW, interview, year 2

We did have to tweak them quite a lot just to make sure that all children would be able to do it. So I would say they probably spent quite a bit of time just tweaking them.

Teacher, school AE, interview, year 2

Intermediate outcomes and mediation

Quantitative analyses examined effects of the intervention on school climate and involvement with anti-school peer groups as pre-hypothesised intermediate outcomes and potential mediators of intervention effects on our primary and secondary student outcomes. The measures of potential mediators had good response rates and reliability, with > 90% of all participants completing all items and with multi-item scales and subscales having alphas of > 70% in all cases (Table 17).

The intervention had statistically significant effects on our overall measure of a positive school climate as well as its constituent subscales (supportive relationships with teachers; sense of belonging; perception of a participative school environment; and commitment to academic values) at 36-month but not 24-month follow-up. The intervention had a statistically significant effect on our measure of involvement with anti-school peer groups at 24 months but one slightly short of statistical significance at 36 months (Table 18). We undertook a per-protocol analysis to examine the effect of additionally adjusting for our potential mediators measured at 24 months on the associations previously found between the intervention and our primary and secondary outcomes measured at 36 months. This analysis found no evidence that this additional adjustment made any difference except marginally in the case of the intervention effect on well-being, where adjustment for both school climate and for involvement with anti-school peers at 24 months removed the previously statistically marginal intervention effect at 36 months (see Appendix 1, Table 40).

TABLE 17 Mediator measures response and reliability

Measure	Response rates, <i>n</i> (%)				Internal consistency			
	Baseline		24 months		Baseline		24 months	
	Completed all items	Completed half or more of items	Completed all items	Completed half or more of items	Alpha (standardised)	Ordinal alpha	Alpha (standardised)	Ordinal alpha
BBSCQ								
Overall BBSCQ	5733 (85.99)	6635 (99.52)	5549 (8.22)	6265 (99.60)	0.9137	0.9373	0.9170	0.9371
Student sense of belonging subscale	6293 (94.39)	6613 (99.19)	5965 (94.83)	6240 (99.21)	0.7952	0.8345	0.8225	0.8576
Student commitment to academic values subscale	6519 (97.78)	6581 (98.71)	6190 (98.41)	6231 (99.06)	0.7394	0.8407	0.7732	0.8619
Student perception of supportive teacher relationships subscale	6221 (93.31)	6631 (99.46)	5935 (94.36)	6247 (99.32)	0.8804	0.9115	0.8938	0.9207
Student perception of participative school environment subscale	6396 (95.94)	6600 (99.00)	6071 (96.52)	6231 (99.06)	0.8005	0.8540	0.8313	0.8712
YPDP measure of student involvement with anti-school peer groups (single item)	6494 (97.41)	NA	6167 (98.04)	NA	NA	NA	NA	NA
BBSCQ, Beyond Blue School Climate Questionnaire; NA, not applicable.								

TABLE 18 Intervention effects on mediators at 24 and 36 months

Measure	24 months				36 months							
	Arm, mean (SE)		Unadjusted effect		Adjusted effect		Arm, mean (SE)		Unadjusted effect		Adjusted effect	
	Control	Intervention	Difference (95% CI)	p-value	Difference (95% CI)	p-value	Control	Intervention	Difference (95% CI)	p-value	Difference (95% CI)	p-value
Beyond Blue School Questionnaire												
BBSQ overall score	2.92 (0.03)	2.90 (0.03)	-0.00 (-0.02 to 0.02)	0.915	-0.00 (-0.02 to 0.02)	0.993	2.82 (0.03)	2.85 (0.03)	0.05 (0.03 to 0.07)	<0.001	0.05 (0.03 to 0.08)	<0.001
Student perception of supportive teacher relationships subscale	2.76 (0.04)	2.70 (0.04)	-0.03 (-0.06 to 0.00)		-0.02 (-0.05 to 0.01)		2.64 (0.03)	2.66 (0.03)	0.05 (0.02 to 0.08)		0.06 (0.03 to 0.09)	
Student sense of belonging	2.84 (0.03)	2.88 (0.03)	0.04 (0.01 to 0.07)		0.04 (0.01 to 0.07)		2.78 (0.02)	2.84 (0.02)	0.06 (0.03 to 0.09)		0.06 (0.03 to 0.09)	
Student perception of participative school environment subscale	2.96 (0.03)	2.91 (0.03)	-0.02 (-0.05 to 0.01)		-0.03 (-0.06 to 0.01)		2.81 (0.04)	2.82 (0.04)	0.05 (0.02 to 0.08)		0.05 (0.01 to 0.08)	
Student commitment to academic values subscale	3.51 (0.01)	3.53 (0.01)	0.00 (-0.02 to 0.03)		0.00 (-0.02 to 0.03)		3.42 (0.01)	3.46 (0.01)	0.03 (0.01 to 0.06)		0.03 (0.01 to 0.06)	
Measure	24 months				36 months							
	Arm, n (%)		Unadjusted effect		Adjusted effect		Arm, n (%)		Unadjusted effect		Adjusted effect	
	Control	Intervention	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value	Control	Intervention	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value
YPDP measure of student involvement with anti-school peer groups, single item												
<i>Friends cautioned/stopped/picked up by police</i>												
No	2203 (70.05)	2271 (75.15)	0.80 (0.66 to 0.97)	0.024	0.80 (0.66 to 0.96)	0.020	2073 (68.64)	2044 (73.21)	0.84 (0.69 to 1.02)	0.074	0.83 (0.69 to 1.01)	0.066
Yes	942 (29.95)	751 (24.85)					947 (31.36)	748 (26.79)				

Chapter 4 Economic evaluation

Methodology

Study question, selection of alternatives and form of evaluation

A cost-effectiveness analysis (CEA) was undertaken to assess the relative cost-effectiveness of the LT intervention compared with standard school-based practices for managing aggression. The economic evaluation was undertaken in conjunction with the analysis of the effectiveness of the LT intervention at 24 and 36 months since randomisation. The primary economic evaluation was a cost-consequences analysis with all main outcomes. In addition, a cost-utility analysis was conducted, with health outcomes expressed in terms of QALYs.

Resource use and outcome data collected as part of the cluster RCT were used to report cost-effectiveness at 24 and 36 months since randomisation. The CEA involved estimation of resource use, costs, HRQoL and QALYs at 24 and 36 months. In the cost-consequences analysis, we report incremental costs and incremental effects of all main outcomes. The cost-utility analysis used information on QoL collected at 24 and 36 months' follow-up to report QALYs at each time point. We compared randomised groups by reporting incremental costs, incremental QALYs and an incremental cost-effectiveness ratio (ICER). The main assumptions of the CEA were subjected to sensitivity analyses. The results are illustrated on the cost-effectiveness planes (Figures 2 and 3).

Effectiveness data and benefit measurement and evaluation

The CHU9D measure⁹⁹ was used to assess HRQoL as part of the economic evaluation. Utility values were collected using the CHU9D questionnaire at baseline and 24 and 36 months. We calculated QALYs for each participant using the 'area under the curve' approach, that is, the weighted average of time spent in the study and HRQoL.¹²⁷

However, the analysis of the pilot study data suggested that there was only a weak correlation between the CHU9D and the proposed primary outcome measures. Therefore, there was a concern that QALYs were unlikely to capture all of the intervention effects. We therefore planned to also perform a cost-consequences analysis as the 'primary analysis', as recommended by the National Institute for Health and Care Excellence (NICE)'s public health methods guidance.¹²⁸

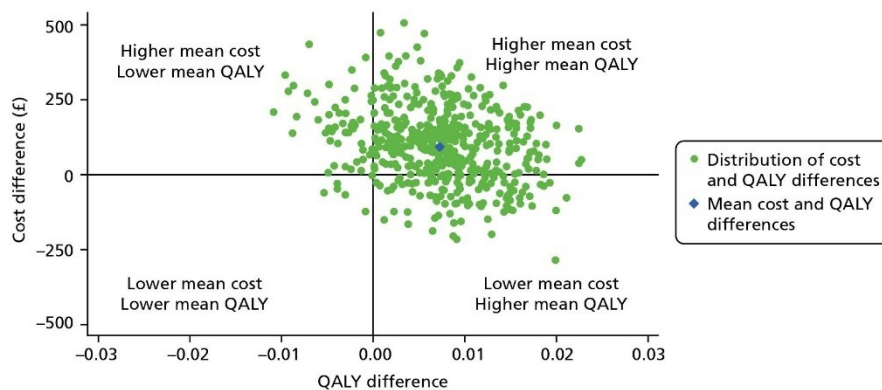


FIGURE 2 Uncertainty in the mean costs (£) and QALY differences and their distribution for intervention versus control at 24 months.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Bonell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. 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.

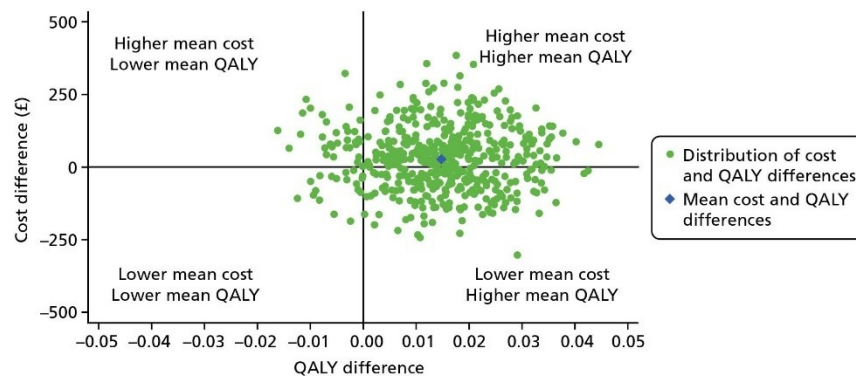


FIGURE 3 Uncertainty in the mean costs (£) and QALY differences and their distribution for intervention versus control at 36 months.

Resource use measurement and costing

The cost analyses took a public sector perspective following NICE's method guidelines.¹²⁹ NICE's guidelines for interventions with health and non-health outcomes in public sector settings¹²⁹ recommend that the base-case cost-effectiveness estimate is presented from a public sector perspective. We used a public sector perspective to cover education, NHS and police costs. The main resource use items and sources of these data are detailed in *Appendix 1, Table 41*.

The costs of the intervention (trainer, facilitator and staff costs) and costs of the NHS and police resources used by the students were considered. In calculating the total cost per school, the attribution of costs such as facilitator and staff time recognised that only a percentage of time is going to the year group under study relative to the total number of students in each school. For the main trial analysis outcomes for the 11- to 12-year-old cohort, the costs have to be a proportion of the total number of students in the school to be consistent. The only exception is that the curriculum was delivered only in that year group so will be costed at 100%.

We collected costs on the delivery of the intervention as incurred during the trial. The costs of the trainers' and facilitators' time were available from the trainers' and facilitators' invoices, which capture preparation and organisation time.

Staff time included staff time spent dealing with bullying or aggression that was identified in the teacher survey. The amount of time spent with AGs was taken from the facilitator diary and, in addition, there were questions in the AG survey about the related time involved in preparation and follow-up. The staff time involved in curriculum change was taken from the curriculum logs. The amount of time spent training was taken from the trainer diaries and the number of teachers attending was recorded on attendance sheets. In the interviews with staff, we enquired whether our training was additional or instead of other training and, if the latter, what it replaced.

Teacher salaries for each intervention and control school were obtained from the Department for Education website.¹³⁰ To estimate an hourly rate, we divided salaries by the Department for Education statutory guidance on school teachers' pay and conditions document detailing the annual hours of work. The student costs included the implications for NHS resource use in terms of visits and hospital stays, and policing costs associated with stopping children and arrests. These items were identified with specific questions in the student survey. The unit cost of inpatient stay in hospital was £298.¹³¹ The unit cost of an NHS visit was

calculated as the weighted average unit costs of general practitioner (GP) consultation and outpatient visit.^{131,132} The percentages of GP visits and outpatient visits reported in another published study¹³³ were applied. The unit costs for police attendance were £267 for those who were not arrested and £457 for those who were arrested. These unit costs were obtained from previously published work.¹³⁴ Total costs up to 24 and 36 months were calculated by combining the resource use with unit costs. Where there were missing costs at the school level we used mean imputation unless there was evidence from the process evaluation that these costs had not been incurred.

Modelling and adjustments for timing of costs and benefits

We considered the potential for longer-term modelling as part of the pilot phase and concluded that, although there are longer-term implications for bullying, given the inherent limitations of the existing evidence base such an exercise would be likely to produce cost-effectiveness estimates that are so uncertain that they would be of little practical use.

The time horizon captured costs and outcomes within the trial. We conducted the analysis at 2 years to capture the time when there was a facilitator, and at 3 years where there was no facilitator for the final year. A discount rate of 3.5% was applied to both costs and outcomes.

Statistical analysis and uncertainty

The CEA followed the intention-to-treat principle, and used multilevel linear regression models that allowed for clustering of pupils at school level¹³⁵ to report mean (95% CI) incremental costs and main outcomes of LT interventions compared with standard practice at 24 and 36 months. The cost-utility analysis also allowed for correlation between costs and QALYs¹³⁶ to report mean (95% CI) incremental costs and QALYs of LT interventions compared with standard practice at 24 and 36 months.

Both cost-consequences and cost-utility analyses reported unadjusted and adjusted differences in mean (95% CI) costs and outcomes (mentioned above) between the randomised groups. In the adjusted analysis, the following baseline variables were adjusted: baseline measures of outcomes, sex, ethnicity and SES, and school-level stratifying factors (single-sex vs. mixed-sex school; school-level deprivation; value-added strata). The differences in average costs and QALYs between the randomised groups were used to calculate the ICER of the LT intervention versus standard school-based practices. The CIs around the ICER were constructed by applying Taylor series expansion¹³⁷ on the incremental estimates of cost and QALY of LT interventions. We also reported cost-effectiveness acceptability curves by calculating the probability that, compared with standard school-based practices, the LT intervention was cost-effective, given the data, at alternative levels of willingness to pay for a QALY gain.

We also performed a number of sensitivity analyses. The sensitivity analysis explored the inclusion of police and NHS costs and the exclusion of costs of school staff time training. The implications of fidelity measures and compliance on cost-effectiveness was explored as per statistical analysis of primary outcomes.

Results

Resource use and costs

Table 19 reports the school staff time associated with delivering the intervention and dealing with bullying and aggression per student. The main time component for school staff was attending the training and curriculum development. The interviews with staff in year 1 revealed that the training was attended not as an additional training but as part of an existing inset training period, suggesting that these staff costs might not need to be included; the implications of this are further explored in the sensitivity analysis. Interestingly, in the intervention arm, school staff spent slightly more time dealing with bullying than in the control arm. The effect of this was more marked in years 1 and 2 than in year 3.

ECONOMIC EVALUATION

TABLE 19 School staff time associated with the delivery of the intervention and dealing with bullying and aggression per student

	Control, mean (SD)	Intervention, mean (SD)
A: Years 1 and 2 i.e. facilitated intervention years		
<i>School staff time (minutes)</i>		
School staff time training (year 1)	.	20.04 (6.82)
Curriculum development (year 1)	.	16.98 (9.32)
School staff time at action groups (year 1)	.	2.62 (1.44)
School staff time preparing/after action groups (year 1)	.	3.72 (2.04)
Curriculum development (year 2)	.	9.73 (9.86)
School staff time at action groups (year 2)	.	2.18 (1.58)
School staff time preparing/after action groups (year 2)	.	2.26 (1.49)
School staff time dealing with bullying and aggression (year 1 and 2)	219.30 (85.44)	224.33 (113.49)
Total school staff time years 1 and 2	219.30 (85.44)	281.85 (120.71)
B: Year 3 i.e. unfacilitated intervention year		
<i>School staff time (minutes)</i>		
Curriculum development	.	6.40 (9.43)
School staff time at action groups	.	1.87 (3.45)
School staff time preparing/after action groups	.	2.47 (4.56)
School staff time dealing with bullying and aggression	115.29 (56.04)	117.44 (56.34)
Total school staff time year 3	115.29 (56.04)	128.21 (57.70)
Total school staff time years 1 to 3	334.58 (119.99)	410.07 (163.10)
Reproduced from Bonell <i>et al.</i> ¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.		

Table 20 shows the costs of trainers, facilitators and staff associated with the delivery of the intervention and dealing with bullying and aggression per student. The average cost for trainers was £3.44 and for facilitators and staff was £5.00 and £3.36, respectively. The largest component of cost for school staff relate to curriculum delivery and staff training. Overall, in the first 3 years, the mean (SD) total cost of trainers, facilitators and school staff time was £116 (£47) in the control arm compared with £163 (£69) in the intervention arm. As expected, in the third year the difference in cost was much smaller as no training or facilitators were required and, in addition, school staff spent less time dealing with bullying. In the third year, the mean (SD) total cost of trainers, facilitators and staff was £63 (£33) in the control group and £74 (£37) in the intervention group.

Table 21 presents the student health service and police resource use and shows the related costs. Overall, the health service and police resource use is similar in both trial arms. However, slightly more students were stopped or told off by police in the control arm. Unexpectedly, more nights in hospital related to accident or injury occurred in the intervention arm, but it was not determined whether or not these were directly related to bullying. As these additional costs of nights in hospital in the intervention arm may not be directly related to bullying, this is further explored in the sensitivity analysis by excluding police and NHS costs.

TABLE 20 Trainers, facilitators and staff costs associated with the delivery of the intervention and dealing with bullying and aggression per student

Costs (£)	Control, mean (SD)	Intervention, mean (SD)
A: Years 1 and 2 i.e. facilitated intervention years		
<i>Trainer and facilitator</i>		
Trainer (year 1)	.	3.44 (1.43)
Facilitator action groups (year 1)	.	5.00 (2.28)
Facilitator action groups (year 2)	.	3.36 (1.35)
<i>School staff time costs</i>		
Staff time training (year 1)	.	10.74 (3.55)
Curriculum development (year 1)	.	9.79 (5.34)
School staff time at action groups (year 1)	.	1.43 (0.87)
School staff time preparing/after action groups (year 1)	.	2.23 (1.46)
Curriculum development (year 2)	.	5.39 (5.37)
School staff time at action groups (year 2)	.	1.16 (0.90)
School staff time preparing/after action groups (year 2)	.	1.21 (0.88)
Staff time dealing with bullying and aggression (year 1 and 2)	115.64 (46.62)	119.27 (62.75)
Total cost years 1 and 2	115.64 (46.62)	163.02 (68.78)
B: Year 3 i.e. unfacilitated intervention year		
<i>School staff time costs</i>		
Curriculum development	.	3.25 (4.80)
School staff time at action groups	.	0.91 (1.59)
School staff time preparing/after action groups	.	1.21 (2.09)
School staff time dealing with bullying and aggression	63.21 (33.49)	68.70 (36.94)
Total costs year 3	63.21 (33.49)	74.08 (36.60)
Total costs years 1 to 3	178.85 (67.42)	237.10 (97.61)

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Health outcomes

The HRQoL utility scores indicate that utility scores were slightly higher at baseline in the control arm and there is no clear pattern of a difference in utility scores between arms (see *Appendix 1, Table 42a–c*).

Cost-effectiveness

There was no significant difference between groups in staff QoL as measured with the overall CHU9D or the SF12 (*Table 22*). However, as noted earlier in the report, there was a significant effect on the primary outcomes of the GBS and ESYTC scores. The mean total costs per pupil at 24 and 36 months for the control school were £493 and £667, respectively. The mean total costs per pupil were higher for the intervention school at both 24 months (£650) and 36 months (£719). The unadjusted incremental costs were £157 (95% CI –£83 to £397) at 24 months and £62 (95% CI –£165 to £288) at 36 months. The adjusted incremental costs were £57 (95% CI –£183 to £297) at 24 months and £71 (95% CI –£73 to £214) at 36 months. Overall, the intervention is associated with higher costs, but the incremental costs are surrounded by statistical uncertainty.

ECONOMIC EVALUATION

TABLE 21 Student health service and police resource use

Resource use	Control, mean (SD)	Intervention, mean (SD)
Resource use up to 24 months		
<i>Health service and police</i>		
Visits to health service for accident or injury	1.56 (3.60)	1.51 (2.58)
Nights in hospital for accident or injury	0.59 (4.76)	1.03 (20.61)
Stopped or told off by police	0.34 (1.06)	0.31 (1.04)
Formally cautioned or arrested by police	0.11 (0.60)	0.11 (0.65)
Resource use 24 to 36 months		
<i>Health service and police</i>		
Visits to health service for accident or injury	0.60 (1.53)	0.92 (9.45)
Nights in hospital for accident or injury	0.30 (2.71)	0.57 (9.64)
Stopped or told off by police	0.18 (0.58)	0.11 (0.45)
Formally cautioned or arrested by police	0.06 (0.33)	0.04 (0.26)
Student health service and police costs (£)		
Costs up to 24 months		
Visits to health service for accident or injury	63.29 (145.70)	61.12 (104.32)
Nights in hospital for accident or injury	172.38 (1395.69)	303.60 (6037.89)
Stopped or told off by police	88.22 (278.93)	82.42 (273.50)
Formally cautioned or arrested by police	48.71 (271.14)	48.29 (292.78)
Total costs up to 24 months	372.60 (1686.06)	495.44 (6203.27)
Costs 24 to 26 months		
Visits to health service for accident or injury	23.21 (58.76)	35.31 (363.19)
Nights in hospital for accident or injury	83.09 (754.93)	158.34 (2682.37)
Stopped or told off by police	44.69 (143.45)	27.87 (111.88)
Formally cautioned or arrested by police	24.96 (139.66)	15.52 (110.07)
Total costs 24 to 36 months	175.95 (871.10)	237.04 (3107.29)
Total costs up to 36 months	483.33 (1827.44)	507.81 (3482.45)
Reproduced from Bonell et al. ¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.		

Overall, the intervention is associated with higher costs (Table 23) but the mean gain in CHU9D score is slightly higher in the intervention arm even though it has not reached a significant level (see Table 22), leading to small QALY gains. The adjusted ICERs are £13,284 (95% CI –£32,175 to £58,743) and £1875 (95% CI –£12,945 to £16,695) per QALY at 2 and 3 years, respectively. These baseline ICERs are well below the threshold at which NICE considers interventions to be cost-effective, which is £20,000 to £30,000 per QALY gained. However, assessment of the CIs around ICERs indicates that there is uncertainty in the results about whether or not the intervention is cost-effective at 2 years, whereas at 3 years it is within the threshold.

TABLE 22 Cost-consequences results

Measure	Trial arm, mean (SD)		Unadjusted effect estimate, difference (95% CI); <i>p</i> -value	Adjusted effect estimate, difference (95% CI); <i>p</i> -value
	Control	Intervention		
At 24 months				
Total costs (£)	493 (1687)	650 (6203)	157 (-83 to 397); 0.199	57 (-183 to 297); 0.642
Bullying victimisation – GBS overall score	0.42 (0.02)	0.37 (0.02)	-0.02 (-0.05 to 0.01); 0.2198	-0.02 (-0.05 to 0.01); 0.1581
Misbehaviour/delinquency – ESYTC	4.24 (0.28)	3.96 (0.28)	-0.04 (-0.34 to 0.27); 0.8113	-0.06 (-0.35 to 0.24); 0.7206
CHU9D – student HRQoL	0.86 (0.01)	0.87 (0.01)	0.00 (-0.00 to 0.01); 0.4883	0.00 (-0.00 to 0.01); 0.2640
SF-12 – staff HRQoL ^a physical health score	54.97 (0.32)	55.22 (0.34)	-0.35 (-1.05 to 0.34); 0.3187	-0.26 (-0.99 to 0.47); 0.4825
SF-12 – staff HRQoL ^a mental health score	44.56 (0.49)	43.71 (0.52)	-0.66 (-1.64 to 0.32); 0.1883	-0.45 (-1.49 to 0.58); 0.3896
At 36 months				
Total costs (£)	667 (1829)	719 (3486)	62 (-165 to 288); 0.593	71 (-73 to 214); 0.334
Bullying victimisation – GBS overall score	0.34 (0.02)	0.29 (0.02)	-0.03 (-0.06 to -0.00); 0.0395	-0.03 (-0.06 to -0.00); 0.0441
Misbehaviour/delinquency – ESYTC	4.33 (0.20)	4.04 (0.21)	-0.07 (-0.38 to 0.25); 0.6820	-0.13 (-0.43 to 0.18); 0.4199
CHU9D – student HRQoL	0.85 (0.00)	0.86 (0.01)	0.01 (-0.00 to 0.01); 0.2423	0.01 (-0.00 to 0.01); 0.0795
SF-12 – staff HRQoL ^a physical health score	55.62 (0.36)	55.18 (0.36)	-0.71 (-1.4 to 0.02); 0.0449	-0.61 (-1.36 to 0.13); 0.1058
SF-12 – staff HRQoL ^a mental health score	43.76 (0.44)	43.69 (0.45)	0.15 (-0.83 to 1.13); 0.7642	0.31 (-0.76 to 1.38); 0.5691

^a These are subscales of the SF-12.

TABLE 23 Cost-utility results

Measure	Trial arm, mean (SD)		Incremental effect (unadjusted), mean (95% CI)	Incremental effect ^a (adjusted), mean (95% CI)
	Control	Intervention		
At 24 months				
Total costs (£)	493 (1687)	650 (6203)	108 (-119 to 335)	96 (-151 to 343)
QALY	1.6833 (0.1710)	1.6834 (0.1710)	0.0106 (-0.0113 to 0.0325)	0.0072 (-0.0043 to 0.0188)
ICER			10,214 (-20,190 to 40,616)	13,284 (-32,175 to 58,743)
At 36 months				
Total costs	667 (1829)	719 (3485)	46 (-154 to 246)	28 (-187 to 242)
QALY	2.4858 (0.2496)	2.4937 (0.2473)	0.0240 (-0.0097 to 0.0578)	0.0148 (-0.0057 to 0.0353)
ICER			1905 (-5938 to 9748)	1875 (-12,945 to 16,695)

^a The incremental effects are reported after adjusting for baseline variables.

Graphically we present the uncertainty in cost-effectiveness acceptability curves in *Figures 4 and 5*. At 2 years, there is a 65% probability that the intervention is cost-effective at a willingness-to-pay threshold of £20,000 per QALY. At 3 years, there is a 90% probability that the intervention is cost-effective at the £20,000 per QALY threshold. Overall, there is some uncertainty at 2 years that the intervention is cost-effective, but at 3 years there is a high degree of certainty that it is.

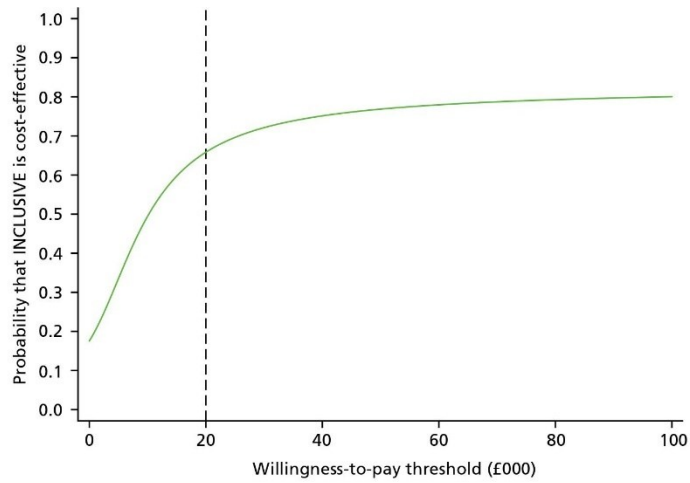


FIGURE 4 Cost-effectiveness acceptability curve, reporting the probability that the intervention is cost-effective (at 24 months) at alternative willingness to pay for a QALY gain.

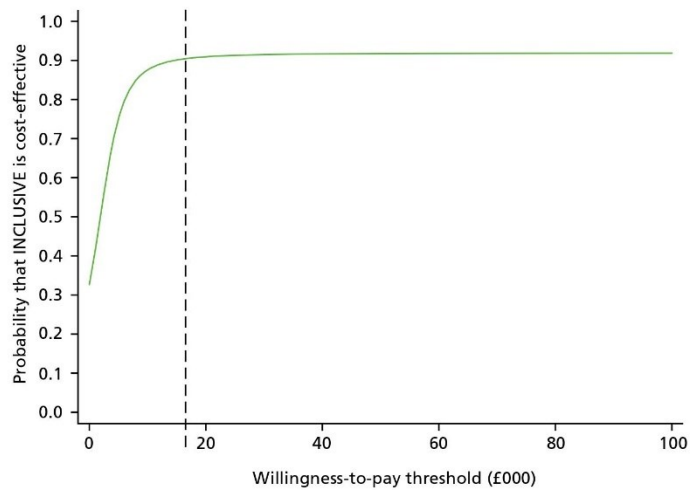


FIGURE 5 Cost-effectiveness acceptability curve, reporting the probability that the intervention is cost-effective (at 36 months) at alternative willingness to pay for a QALY gain.

In the sensitivity analysis we explored the impact of not including teacher training time, as this was provided as part of routine inset days (Table 24). This had little impact on results and conclusions. Again, there was some uncertainty about whether or not the increased hospital days in the intervention arm were related to the trial, but exclusion of NHS and police costs did not have a significant impact on results. The fidelity analysis shows (Table 25) that a higher fidelity score was associated with lower costs and QALYs in the unadjusted analysis. When other baseline variables are adjusted, higher fidelity score was associated with lower costs but higher QALYs, although the results were surrounded by statistical uncertainty.

TABLE 24 Sensitivity analysis reporting costs, QALYs and ICERs at 36 months

Measure	Trial arm, mean (SD)		Incremental effect (unadjusted), mean (95% CI)	Incremental effect (adjusted), mean (95% CI)
	Control	Intervention		
Excluding staff time training costs				
Total costs (£)	667 (1829)	709 (3486)	35 (-164 to 235)	28 (-187 to 242)
QALY	2.4858 (0.2496)	2.4937 (0.2473)	0.0241 (-0.0097 to 0.0577)	0.0148 (-0.0057 to 0.0353)
ICER			1476 (-6393 to 9345)	1175 (-13,470 to 15,820)
Excluding NHS and police costs				
Total costs (£)	184 (66)	215 (84)	17 (12 to 22)	32 (28 to 36)
QALY	2.4858 (0.2496)	2.4937 (0.2473)	0.0129 (-0.0035 to 0.0294)	0.0106 (-0.0009 to 0.0221)
ICER			1314 (-371 to 2999)	3008 (-275 to 6291)
Alternative discount rate				
Total costs (£)	681 (1862)	735 (3605)	47 (-158 to 252)	29 (-191 to 249)
QALY	2.5509 (0.2562)	2.5591 (0.2538)	0.0247 (-0.0099 to 0.0593)	0.0152 (-0.0059 to 0.0362)
ICER			1918 (-5907 to 9743)	1917 (-12,900 to 16,734)

TABLE 25 Effect of fidelity score on costs and QALYs at 36 months

Measure	Analysis, mean (95% CI)	
	Unadjusted	Adjusted
Total costs (£)	-16 (-92 to 61)	-13 (-73 to 46)
QALY	-0.0007 (-0.0112 to 0.0098)	0.0043 (-0.0017 to 0.0102)

Chapter 5 Discussion

We present here the first randomised trial of restorative approaches to reducing bullying and aggression and to promote student health in schools, within a multicomponent whole-school intervention engaging students in school decision-making, and providing restorative practice and social and emotional skills education. LT resulted in a very broad range of benefits for behaviour and health outcomes when delivered as a universal intervention over 3 years.

Primary and secondary results

Learning Together reduced student reports of bullying victimisation, both overall and in terms of teasing and rumour-spreading, compared with schools continuing their standard practice. We did not identify a reduction in student reports of aggression across the whole sample. Additionally, LT appeared to have larger beneficial impacts on a wide range of important secondary outcomes among students, ranging from improved psychological function, well-being and QoL to reductions in police contact, smoking and alcohol and drug use. The effects seen for bullying and other continuous outcomes by the third year approximate 0.1 SD, which are potentially highly important at the population level. The impact of LT on risk behaviours at 36 months was also substantial. We found intervention effects both in the whole sample and in those with higher levels of bullying or aggression at baseline, implying that the intervention worked to curtail existing bullying and aggression (secondary prevention) as well as prevent new bullying (primary prevention).

We found a small effect of the intervention on bullying at 36 months (as hypothesised for our primary outcome) but not at 24 months, and we found a similar strengthening of effects over time for most secondary outcomes, suggesting that this was a broad intervention effect. We had anticipated that this would be the case, reflecting the time needed for changes made in the first 2 years to be translated into organisational change within schools. Although many schools did not deliver formal intervention components so well in year 3 as earlier, our process evaluation suggested that by year 3 schools had integrated some of the AG functions into mainstream school structures and processes, enhancing their ability to implement whole-school change. This strengthening of intervention effects with time is consistent with evidence from the Gatehouse Project, in which effects increased among student cohorts subsequently moving through intervention schools after the initial 2-year intervention was completed.⁶⁸

We found that although LT was delivered as a universal intervention, exploratory analyses suggest that it was most powerful for students with higher baseline levels of bullying or aggressive behaviours, regardless of the prevalence of bullying or aggression in their school. This is reassuring in that LT appears to benefit those who most need it regardless of the level of problems in a school, and reflects similar findings in other school aggression programmes,⁹³ but is also likely to reflect greater room for improvement among those with higher baseline problems. We also found LT to have greater effects for boys than for girls for secondary psychological and behavioural outcomes, although not for primary outcomes.

There are few other studies to which we can appropriately compare our findings. Our study is unique for the breadth of positive outcomes across bullying, well-being, QoL, psychological problems and substance use. We did not find an effect of our intervention on perpetration of aggressive behaviours, which is contrary to the Ayan Aba study,⁶¹ although consistent with the Gatehouse study⁶⁸ and findings from reviews that school-based studies consistently have stronger effects for victimisation than perpetration.⁵⁷ Among comparable trials, the Gatehouse Project identified no impact on bullying or psychological problems, in contrast to our findings,⁶⁰ although it identified similar impacts on risky behaviours, including substance use.⁶⁸ As predicted by the theory of human functioning and school organisation,⁵⁹ which underpinned our intervention theory of change, intervention effects were concentrated on risk behaviours that may function as overt manifestations of anti-school rebelliousness, such as bullying, smoking and drunkenness, rather than on outcomes that are less manifest, such as sexual health outcomes.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Bonell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. 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.

DISCUSSION

We hypothesised that the intervention would increase the age at sexual debut and increase use of contraception because there is evidence that engagement with school is protective against early sexual debut and lack of contraception use.¹³⁸ However, we found no effects on these outcomes, perhaps because our intervention did not explicitly address sexual health, or because, unlike bullying and substance use, sexual behaviours occur off the school site and in private. We did not assess sexual debut or contraception use at baseline at the request of the ethics committee because of the age of the baseline cohort. It is possible that, despite good balance generally in the trial between the intervention and control arms, there were chance imbalances in these outcomes that might have obscured intervention effects on these outcomes.

Intervention cost and cost-effectiveness

The intervention was low cost, falling into the 'very low cost' category for school interventions according to the Educational Endowment Foundation guidance.¹³⁹ The costs of trainers, facilitators and school staff were an additional £58 per pupil in the intervention compared with control schools over the 3 years, and approximately £47 if LT training costs were regarded as replacing other teacher training rather than being additional. In terms of cost per QALY, at 2 years there was uncertainty about the cost-effectiveness (ICER £13,284, 95% CI -£32,175 to £58,743), but over 3 years the intervention was cost-effective (ICER £1875, 95% CI -£12,945 to £16,695).

Trial context

Schools recruited to the trial were commonly already taking actions to implement RP, student participation in decisions, and/or social and emotional skills education, but these were not being implemented as part of a coherent, guided intervention. Despite reporting the value of RP, most schools used discipline systems focused on traditional punitive responses and only a few schools had staff trained in RP. Schools allocated to the control group did, in some cases, act to implement elements of what they perceived the intervention to involve, in particular RP. However, only a handful of control schools implemented RP, student participation in decisions and social and emotional skills education, and none did so as part of a coherent intervention.

Fidelity and acceptability

The intervention was implemented with variable fidelity and acceptability between components. The training, needs reports and AGs were well delivered and generally acceptable. AGs generally succeeded in reviewing policies and rules, and this was perceived as useful. There were mixed views about the facilitators. On balance, they appeared useful to keep schools focused on scheduling meetings and delivering action, but they may not have needed to attend every meeting. Meetings became less frequent after facilitators were withdrawn, but this may have reflected the mainstreaming of their function (see below). The curriculum was delivered patchily and there were more negative views than positive from teachers. Although data on implementation were poor, interviews and surveys suggest that RP, particularly in the form of conferences, was widely delivered and acceptable. However, time constraints and perceived threats to authority were listed as barriers. The intervention was developed with lower formal fidelity in year 3 but there was some evidence that schools prioritised the components that they most valued and adapted and mainstreamed these components into normal school policies and systems.

Determinants of implementation

The intervention was delivered more completely when led by a member of staff with sufficient authority and support to make decisions and drive delivery. In many, but not all, cases, this required that the staff member was on the school's SLT. Fidelity, in terms of both success in delivering specific components and

ensuring that activity was focused on improving relationships and supporting students' overall well-being, also appeared stronger when the intervention could build on a pre-existing school culture that was at least starting to address such issues.

Reach and awareness

There was low awareness of the intervention among both staff and students not directly involved in implementation. Staff and students who were involved in delivery did not view this as a barrier to the intervention achieving impacts, as what counted was piecemeal actions to improve the school environment whether or not these were branded as part of the intervention.

Intermediate outcomes

There was some evidence that the intervention did improve students' sense of a positive school climate, but as this did not manifest until year 3 it was not possible to conclude that this mediated intervention effects on primary and secondary outcomes that also did not manifest until year 3. There was some evidence that the intervention reduced student contact with anti-school peer groups in year 2 but not year 3 and with no evidence that this mediated intervention effects on primary or secondary outcomes. Qualitative data showed how intervention components and processes improved empathy between students and students and staff.

Strengths and limitations

We undertook a well-conducted, large, cluster RCT of an intervention of proven feasibility (LT) in a group of schools that are larger, are more ethnically and culturally diverse, and have notably greater levels of student deprivation than the English average. Participating schools were representative of the approximately 500 schools initially approached and all schools were retained in the trial. Our follow-up was sufficiently lengthy to allow time both for intervention effects to develop and for investigation of persistence of intervention effects after the end of the facilitated intervention. Student participation was high. Our outcome research team and intervention team remained independent throughout the trial and blinding of lead researchers was maintained. Outcomes were assessed using age-appropriate validated instruments. Although self-report outcomes can be open to recall bias, baseline data were collected before randomisation, instruments were used with standardised recall periods and actions at the school level are very unlikely to have biased reporting between intervention and control arms. Analysis used an intention-to-treat approach and appropriate mixed-effects models.

Our data are subject to a number of limitations. In terms of generalisability, our trial was carried out in a representative sample of urban and periurban settings in and around London. Although our study did not include schools in rural England, our process evaluation identified no factors that might suggest that implementation or effects would be different in such schools. The absence of students at some waves may have introduced bias. However, given that non-responders are more likely to have experienced bullying or behaviour problems, this is likely to have underestimated the intervention effect. The large number of secondary outcomes investigated necessitated multiple statistical testing. However, to mitigate this, we restricted ourselves to testing only prespecified secondary and subgroup analyses. A sensitivity analysis allowing for a very stringent Bonferroni correction to the *p*-values led to little change in our conclusions; the evidence of an effect of the intervention remained for QoL, psychological total difficulties, ever smoked regularly, been offered or tried illicit drugs and number of times really drunk. Subgroup analyses were underpowered. We have not yet assessed the impact of the intervention on staff outcomes or on educational outcomes (attainments, attendance and exclusions), as these rely on routine administrative data which will be available 1–2 years after the end of the trial. We used the GBS, a well-established tool, to measure the occurrence of bullying victimisation in schools. The measure aligns with the WHO definition

DISCUSSION

of bullying described in *Chapter 1*.⁹ It focuses on different forms of abuse committed face to face or online, as well as the frequency of victimisation and the hurt caused. The tool aligns less well with some other definitions of bullying, such as that of Olweus,¹¹ which focuses on repeat victimisation from the same perpetrator(s) as well as power imbalances between perpetrator(s) and victim. We used the ESYTC school misbehaviour subscale to assess broader forms of violence and aggression at school.⁷² It should be noted that, as well as assessing forms of behaviour that clearly involve interpersonal aggression, the measure includes items on some behaviours that might best be regarded as anti-social, such as arriving late for classes, refusing to do work, wandering around school in class time and cheating when doing homework or tests.

Self-report of student outcomes, although providing the most useful data for the majority of outcomes, has limitations. There is evidence that peer reports identify a higher prevalence of bullying perpetration than self-reports,¹⁴⁰ but we are unaware of any evidence that peer reports identify a higher prevalence of bullying victimisation than self-reports. The use of multiple respondents may have improved the accuracy of mental health data. Parental report or the use of linked routine administrative data is likely to have improved the accuracy of policy contacts and health service use. Our process evaluation was both broad in collecting data from all schools and deep in exploring some processes in depth in case study schools. It also enabled a longitudinal evaluation of implementation across 3 years. Although response rates for process data collection were generally high, some aspects, such as the surveys of teachers delivering the curriculum, had lower response rates. Despite this, our multisource approach meant that we were still able to assess what was happening in most schools, with the exception of a few schools in year 3, when it is reasonable to assume that few, if any, intervention activities were being implemented.

Fidelity of implementation of LT was variable. Some schools in the control arm implemented activities that resembled some elements of LT. However, only five control schools implemented what might be regarded as three of the key elements of our intervention (RP, social and emotional skills education and student participation in decision-making) and implemented these with less support than was the case in intervention schools. A per-protocol analysis excluding these control schools found similar intervention effects. A sensitivity analysis excluding the six case study schools selected for more intensive process evaluation data collection and showed no discernible differences in intervention effects from the intention-to-treat analyses, although CIs were wider as these analyses were underpowered.

The economic evaluation also had some limitations. We have not incorporated the staff outcomes (HRQoL) into the QALY estimates, although we do report physical and mental health scores of staff in the cost-consequences analysis. This was because it is difficult to know how staff and student QoL should be combined. It is possible that we have not captured all of the costs incurred in the control arm, as these were difficult to capture without changing the control arm status. Finally, we have only conducted a within-trial analysis and have not modelled the longer-term impact of bullying on health outcomes, costs and QoL.

Chapter 6 Conclusions

Learning Together offers the potential for broad improvements in behaviour and health to be promoted in secondary schools and provides strong support for further development of restorative approaches in such settings. Positive effects were found, despite the intervention being delivered with variable fidelity. It may be that, for this sort of organisational change intervention, traditional fidelity of 'form' (what intervention components were delivered) was less important than overall 'fidelity of function' (whether or not overall the intervention achieved benefits in the ways theorised, albeit locally appropriate).¹⁴¹ Our findings are particularly encouraging given that many of the control schools were delivering broadly similar activities, including RP and student involvement in decision-making. This suggests that the intervention packaged and promoted these activities more effectively than most schools could do on their own.

We have demonstrated the effectiveness and cost consequences of the LT intervention for a broad range of key public health targets for young people. At a time when young people's mental health is a major public health concern internationally,^{42,142} countries such as the UK⁴² and Australia¹⁴² have identified schools as a key policy platform for improving mental health. Universal school environment interventions such as LT are likely to be one of the most efficient ways of promoting mental health and well-being and simultaneously addressing other health harms in adolescence, owing to their potential for modifying population-level risk as well as their wide reach across health outcomes and likely sustainability.¹⁴³

Implications for research and practice

Our study adds to the evidence that whole-school approaches to preventing bullying and aggression and promoting student health are feasible to implement and have positive effects on a range of outcomes in a broad range of high-, middle- and low-income settings.^{57,60,61,144} LT offers the potential for broad improvements in behaviour and health in secondary schools and, as the first RCT of school-based RP, provides strong support for further development of restorative approaches in secondary schools. The results are important for public health policy in that a single, very low-cost intervention had an impact on a clustered set of outcomes of public health importance including bullying, mental health, well-being and QoL, as well as on the use of tobacco, alcohol and drugs. The findings are also of broader scientific interest in that they provide the first experimental evidence that it is possible to promote multiple health outcomes by transforming the school environment and increasing educational engagement rather than via health services or traditional health education, as we theorised.

The wider value of LT must be examined in further trials in diverse settings, such as nations within the UK that have different education systems (e.g. Scotland). We recommend that some future studies examine the effectiveness of interventions that do not include a curriculum component in order to assess whether or not whole-school elements alone are sufficient to achieve health outcomes. We also recommend that other studies examine the effectiveness of an intervention that incorporates a strengthened curriculum element, where more thought is given to ensuring that lessons are culturally appropriate and can be delivered within busy school timetables. Curriculum components may prove more feasible to deliver in schools or school systems that have dedicated time in school timetables for personal, social and/or health education. We further recommend that future studies explore whether external facilitation is essential to initiate implementation or training for schools on implementing whole-school change would be an alternative. We recommend that, in future studies, training be repeated annually to compensate for staff turnover.

We are currently undertaking further analyses to investigate which elements of the intervention may have been the most beneficial. The poor fidelity for the curriculum element suggests this was less likely to have contributed significantly to the benefits detected, although it is possible that schools were already implementing social and emotional learning in other ways.

© Queen's Printer and Controller of HMSO 2019. This work was produced by Bonell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. 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.

CONCLUSIONS

Future work will examine staff and educational outcomes, explore the sustainability of the intervention within LT intervention schools once the trial was complete, and examine hypothesised mechanisms by which the intervention had an impact on outcomes and whether or not these varied between schools. Work is also needed to establish the generalisability of our findings. However, given that participating schools were representative of those invited to participate and included a good range in terms of attainment, deprivation and inspectorate ratings, this suggests that LT may achieve similar effects in other schools in England and beyond. Given the extent to which control schools were implementing quite similar activities to those included in the intervention, effects of LT might be greater in settings and countries where schools are less involved in such activities. Future work is needed to examine broader implementation of LT and evaluate any modification of the intervention.

Chapter 7 Other information

Patient and public involvement

This trial has strongly benefited from service user and public involvement. Our trial involved young people from the National Children's Bureau young researchers' group (YRG) in advising on intervention and research methods.

We consulted with the YRG on refinements to the intervention and trial methods during the set-up phase. This ensured that the intervention remained acceptable and appropriate to young people, as they were in the pilot. Three meetings took place with the YRG on these issues. Note that training and support to the YRG was routinely provided by the National Children's Bureau. At the trial summation, the YRG was again consulted about the meaning of the results to students and schools, and the most effective ways to disseminate the findings to schools and charities. One group's feedback on this is provided in *Appendix 2*.

School AGs during the intervention ensured that implementation in their school was appropriate to local needs that were identified by members and the survey of student needs. This included ensuring that revisions to policies and rules built on existing work, deciding which curriculum modules to deliver in each year, and implementing locally decided actions aiming to improve relationships and student participation (e.g. cascading restorative practice training to staff who had not attended or to student peer mentors).

The study student advisory group comprised students and staff drawn from intervention schools in the pilot who had signalled their willingness to undertake this. This was used to consolidate learning from the pilot, particularly regarding implementation issues.

We convened an advisory group of research and service users. This included representatives who were key policy-makers and clinicians. This group met twice during the study and provided an opportunity for the research team to consult about the research design and methods for data collection, choice of outcomes and methods for data analyses. The advisory group also aided in interpreting the initial findings and developing the dissemination strategies.

Research governance

Trial documentation

Relevant trial documentation will kept for a minimum of 15 years.

Trial registration and conduct

The pilot study was registered with controlled-trials.com (ISRCTN88527078) and the full trial was similarly registered. Note that as the trial was not carried out in clinical settings and did not use clinical samples or a medicinal product, there was no requirement to comply with the The Medicines for Human Use (Clinical Trials) Regulations 2004.¹⁴⁵ We followed the MRC's *Guidelines on Good Clinical Practice in Clinical Trials*.¹⁴⁶ Note that the CI and the majority of the other investigators have been trained in Good Clinical Practice for clinical trials. The protocol for this trial can be found online.¹

Sponsor

This paper presents independent research funded by NIHR in England under its Public Health Research Board (12/153/60) and the Education Endowment Foundation.

Trial Steering Committees

The trial was overseen by a TSC, including an independent chairperson, at least two other independent members, and a patient and public involvement representative, and an investigator representative of each institution involved in the research. The TSC formed for the pilot trial agreed to continue for the full trial (chairperson: Professor Rona Campbell, Bristol University; other members listed in the uploaded pilot trial report). Observers from the Health Technology Assessment (HTA) programme were invited to all TSC meetings. The TSC met 6-monthly throughout the pilot. A monitoring schedule covering the roles and responsibilities of the researcher, project team, management committee and TSC for monitoring recruitment, data quality, compliance, safety and ethics was developed and agreed.

Data Monitoring and Ethics Committee

A DMC was established independent of the investigators and of the TSC, but reporting to the TSC and (via the TSC) to the sponsors and the HTA programme. This committee comprised an independent chairperson, a senior statistician and at least one other senior academic independent of the investigators. They met approximately yearly during the study. The DMC monitored data for quality and completeness. Data quality, follow-up and trial monitoring were facilitated through the development of a trial-specific database, including validation, verification, monitoring and compliance reports and follow-up report functionalities. The DMC examined the results of an interim analysis at 24 months to consider any potential harms.

Study management

Russell Viner directed the study together with Chris Bonell as co-director. The intervention and research teams were functionally independent. The research team was managed by Russell Viner with Chris Bonell and an experienced full-time trial manager based at the UCL Institute of Child Health. Chris Bonell will direct the process evaluation. An executive of Russell Viner, Chris Bonell, Adam Fletcher, Meg Wiggins and Elizabeth Allen met monthly with the trial manager. The trial manager had day-to-day responsibility for the conduct of the trial and the operations of the research team, and reported to the study executive and ultimately to Russell Viner. The investigator group formed the Scientific Steering Committee for the study and met four-monthly throughout the trial. Responsibility for data integrity and analysis were held by the CTU at the LSHTM.

The intervention team was managed by Meg Wiggins at the Institute of Education, together with an intervention manager who directed day-to-day operation of the intervention and co-ordinated the educational facilitators. Note that Meg Wiggins formed part of the Scientific Steering Committee.

Acknowledgements

Miranda Perry played a key role in developing the intervention and co-ordinated its implementation in the pilot and Phase III trials.

We are grateful to the staff and students of participating schools for their dedication to the intervention and completion of the outcome surveys and process evaluation surveys and interviews. We are very grateful for the advice and support of our TSC and DMC.

We acknowledge the work of, and mourn the loss of, Dr Farah Jamal during the trial, whose death at the age of 30 years was a tragic loss for public health research.

Contributions of authors

Chris Bonell (Professor, Public Health Sociology) co-led development of the LT intervention, conceptualised the study, led design of the trial, led the research and writing of the manuscript. He directed the study, together with Russell M Viner. Chris Bonell conducted the trial's progress review, including staff and resources. He ensured that the trial was administered in a financially responsible manner. The research team was managed by Russell M Viner with Chris Bonell. Chris Bonell led the process evaluation and intervention teams. Chris Bonell planned and led the process evaluation.

Elizabeth Allen (Professor, Medical Statistics) contributed to the trial design, and was responsible for randomisation of schools, for entry and cleaning of all quantitative data, and for all quantitative analyses.

Emily Warren (Research Fellow, Behaviour and Intervention Use) was responsible for process evaluation.

Jennifer McGowan (Research Associate, Health Psychology) conducted the review of the progress of the trial, looking at staff and resources issues. She reported to the NIHR on study's progress and achievements. She had day-to-day responsibility for the conduct of the trial and the operations of the research team, and writing of the final report.

Leonardo Bevilacqua (Research Assistant, Developmental Psychopathology) conducted schools recruitment, dissemination of surveys in schools, overall management of the trial, and led the process evaluation activities.

Farah Jamal (Social Researcher, School Environment) conducted schools recruitment, surveys in schools, overall management of the trial, and led the process evaluation activities.

Zia Sadique (Lecturer, Economic Evaluation) conducted the economic evaluation with Rosa Legood.

Rosa Legood (Associate Professor, Health economics) conducted overviews of progress, advice, implementation, publications and dissemination strategy. She was responsible for the economic evaluation.

Meg Wiggins (Research Officer, Education) managed the intervention team at the Institute of Education. She was part of the Scientific Steering Committee.

Charles Opondo (Honorary Assistant Professor, Medical Statistics) conducted the qualitative statistical analysis with Elizabeth Allen.

ACKNOWLEDGEMENTS

Anne Mathiot (Trial Manager) conducted the review of the progress of the trial, looking at staff and resources issues. She reported to the NIHR on study's progress and achievements, before passing the role to Jennifer McGowan.

Joanna Sturgess (Research Fellow, Clinical Trials) was responsible for randomisation of schools, for entry and cleaning of all quantitative data, and for the quantitative analyses. She managed the trial data and contributed to the analysis.

Sara Papparini (Honorary Research Fellow, Social and Environmental Health) contributed to the fieldwork and analysis within the process evaluation.

Adam Fletcher (Senior Lecturer, Public Health Improvement) conducted overviews of the intervention progress, advice, implementation, publications and dissemination strategy.

Miranda Perry (Education Consultant) was responsible for the information required for delivering the intervention in the schools.

Grace West (Research Assistant, Developmental Psychology) conducted schools recruitment, dissemination of surveys in schools, overall management of the trial, and led the process evaluation activities.

Tara Tancred (Assistant Professor, Public Health) contributed to the fieldwork and analysis within the process evaluation.

Stephen Scott (Professor, Child Health and Behaviour) conducted overview of the intervention progress, implementation, publications and dissemination strategy.

Diana Elbourne (Professor, Health-Care Evaluation) contributed to the trial design, was responsible for randomisation of schools, for entry and cleaning of all quantitative data and for quantitative analyses. She also conducted overviews of the intervention progress, advice, implementation, publications and dissemination strategy.

Deborah Christie (Professor, Paediatric and Adolescent Psychology) conducted overviews of the intervention progress, advice, implementation, publications and dissemination strategy.

Lyndal Bond (Professor, Population Health and Evaluation) conducted overviews of the intervention progress, advice, implementation, publications and dissemination strategy.

Russell M Viner (Professor, Adolescent Health) co-led development of the LT intervention, conceptualised the study, led design of the trial, led the research and led the writing of the manuscript. He directed the study, together with Chris Bonell as co-director. Russell M Viner conducted the trial's progress review, including staff and resources. He ensured that the trial was administered in a financially responsible manner. The research team was managed by Russell M Viner with Chris Bonell. Russell M Viner led the outcome evaluation team.

All authors contributed to writing the manuscript, with the exception of **Farah Jamal**.

Publications

Bonell C, Allen E, Christie D, Elbourne D, Fletcher A, Grieve R, *et al*. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial. *Trials* 2014;**15**:381.

- Hargreaves D, Bevilacqua L, Shackleton N. Association between ethnic minority status at school and bullying and mental health outcomes in early adolescence: cross-sectional analysis of data from the INCLUSIVE study. *Lancet* 2015;**386**:S42.
- Jamal F, Fletcher A, Shackleton N, Elbourne D, Viner R, Bonell C. The three stages of building and testing mid-level theories in a realist RCT: a theoretical and methodological case-example. *Trials* 2015;**16**:466.
- Bonell C, Shackleton N, Fletcher A, Jamal F, Allen E, Mathiot A, *et al.* Student- and school-level belonging and commitment and student smoking, drinking and misbehaviour. *Health Educ J* 2016;**76**:206–20.
- Bonell C, Warren E, Fletcher A, Viner R. Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle *et al.* *Trials* 2016;**17**:478.
- Shackleton N, Hale D, Bonell C, Viner RM. Intraclass correlation values for adolescent health outcomes in secondary schools in 21 European countries. *SSM Popul Health* 2016;**2**:217–25.
- Bonell C, Mathiot A, Allen E, Bevilacqua L, Christie D, Elbourne D, *et al.* Initiating change locally in bullying and aggression through the school environment (INCLUSIVE) trial: update to cluster randomised controlled trial protocol. *Trials* 2017;**18**:238.
- Shackleton N, Fletcher A, Jamal F, Markham W, Aveyard P, Mathiot A, *et al.* A new measure of unhealthy school environments and its implications for critical assessments of health promotion in schools. *J Crit Pub Health* 2017;**27**:248–62.
- Bonell C, Allen E, Warren E, McGowan J, Bevilacqua L, LeGood R, *et al.* A multi-component school environment intervention reduces bullying and risky behaviour and improves mental health and quality of life: findings from the INCLUSIVE cluster randomised controlled trial. *J Adolesc Health* 2018;**62**:S9.
- Bonell C, Moore G, Warren E, Moore L. Are randomised controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services. *Trials* 2018;**19**:238.
- Fantaguzzi C, Allen E, Miners A, Christie D, Opondo C, Sadique Z, *et al.* Health-related quality of life associated with bullying and aggression: a cross sectional study in English secondary schools. *Eur J Health Econ* 2018;**19**:641–51.
- McGowan J, Bonell C, Allen E, Warren E, Aswathikutty A, Bevilacqua L, *et al.* The relationship between school-level factors and adolescent student well-being: cross-sectional findings from the INCLUSIVE trial. *J Adolesc Health* 2018;**62**:S14–15.
- Bonell C, Allen E, Warren E, McGowan J, Bevilacqua L, Jamal F, *et al.* Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial. *Lancet* 2018;**392**:2452–64.
- Bonell C, Allen E, Opondo C, Warren E, Elbourne D, Sturgess J, *et al.* Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention. *J Epidemiol Community Health* 2019;**73**:455–64.
- Bonell C, Beaumont E, Dodd M, Elbourne D, Bevilacqua L, Mathiot A, *et al.* The effects of school environments on student risk behaviours: evidence from longitudinal study of secondary schools in England. *J Epidemiol Community Health* 2019;**73**:502–8.

ACKNOWLEDGEMENTS

Bonell C, Allen E, Opondo C, Warren E, Elbourne DR, Sturgess J, *et al.* Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention. *J Epidemiol Community Health* 2019;**73**:455–64.

Warren E, Bevilacqua L, Opondo C, Allen E, Mathiot A, West G, *et al.* Action groups as a participative strategy for leading whole-school health promotion: results on implementation from the INCLUSIVE trial in English secondary schools. *Br Educ Res J* 2019;**45**:979–1000.

Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. Access to anonymised data may be granted following review.

References

1. Bonell C, Allen E, Christie D, Elbourne D, Fletcher A, Grieve R, *et al*. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial. *Trials* 2014;**15**:381. <https://doi.org/10.1186/1745-6215-15-381>
2. World Health Organization. *Global Status Report on Violence Prevention 2014*. Geneva: World Health Organization; 2014.
3. Department of Health and Social Care. *Public Mental Health Priorities: Investing in the Evidence. Annual Report of the Chief Medical Officer 2013*. London: Department of Health and Social Care; 2014.
4. 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. <https://doi.org/10.1136/bmj.323.7306.191>
5. Krug EG, Mercy JA, Dahlberg LL, Zwi AB. The world report on violence and health. *Lancet* 2002;**360**:1083–8. [https://doi.org/10.1016/S0140-6736\(02\)11133-0](https://doi.org/10.1016/S0140-6736(02)11133-0)
6. McKee M, Raine R. Riots on the streets. *BMJ* 2011;**343**:d5248. <https://doi.org/10.1136/bmj.d5248>
7. Fletcher A, Gardner F, McKee M, Bonell C. The British government's Troubled Families Programme. *BMJ* 2012;**344**:e3403. <https://doi.org/10.1136/bmj.e3403>
8. 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 29 September 2017).
9. Craig WM, Henderson K, Murphy JG. Prospective teachers' attitudes toward bullying and victimization. *School Psychol Int* 2000;**21**:5–21. <https://doi.org/10.1177/0143034300211001>
10. Newman RS, Murray B, Lussier C. Confrontation with aggressive peers at school: students' reluctance to seek help from the teacher. *J Educ Psychol* 2001;**93**:398–410. <https://doi.org/10.1037/0022-0663.93.2.398>
11. Olweus D. Bullying at school: basic facts and effects of a school based intervention program. *J Child Psychol Psychiatry* 1994;**35**:1171–90. <https://doi.org/10.1111/j.1469-7610.1994.tb01229.x>
12. Brooks F, Magnusson J, EKlemera E, Chester K, Spencer N, Smeeton N. *HBSC England National Report 2014*. Hatfield: University of Hertfordshire; 2015.
13. Nansel TR, Craig W, Overpeck MD, Saluja G, Ruan WJ, Health Behaviour in School-Aged Children Bullying Analyses Working Group. Cross-national consistency in the relationship between bullying behaviors and psychosocial adjustment. *Arch Pediatr Adolesc Med* 2004;**158**:730–6. <https://doi.org/10.1001/archpedi.158.8.730>
14. Radford L, Corral S, Bradley C, Fisher H, Bassett C, Howat N, Collishaw S. *Child Abuse and Neglect in the UK Today*. London: National Society for the Prevention of Cruelty to Children; 2011.
15. Bevilacqua L, Shackleton N, Hale D, Allen E, Bond L, Christie D, *et al*. The role of family and school-level factors in bullying and cyberbullying: a cross-sectional study. *BMC Pediatr* 2017;**17**:160. <https://doi.org/10.1186/s12887-017-0907-8>
16. 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. <https://doi.org/10.1186/1471-2458-11-440>
17. Chamberlain T, George N, Golden S, Walker F, Benton T. *Tellus4 National Report*. London: DCSF; 2010.

REFERENCES

18. Home Office. Research, Development and Statistics Directorate. Offending Surveys and Research, National Centre for Social Research, BMRB. Social Research. (2008). *Offending, Crime and Justice Survey, 2006. 2nd Edition*. 2006. URL: <https://beta.ukdataservice.ac.uk/datacatalogue/doi/?id=6000#!#1> (accessed 29 September 2017).
19. Wolke D, Woods S, Stanford K, Schulz H. Bullying and victimization of primary school children in England and Germany: prevalence and school factors. *Br J Psychol* 2001;**92**:673–96. <https://doi.org/10.1348/000712601162419>
20. 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 & Dentistry; 2003.
21. 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. <https://doi.org/10.2105/AJPH.87.12.1997>
22. National Center for Educational Statistics. *Indicators of School Crime and Safety: 2007*. Washington, DC: National Center for Educational Statistics; 2007.
23. Arseneault L. Annual research review: the persistent and pervasive impact of being bullied in childhood and adolescence: implications for policy and practice. *J Child Psychol Psychiatry* 2018;**59**:405–21. <https://doi.org/10.1111/jcpp.12841>
24. 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. <https://doi.org/10.1542/peds.2005-2388>
25. Hair EC, Park MJ, Ling TJ, Moore KA. Risky behaviors in late adolescence: co-occurrence, predictors, and consequences. *J Adolesc Health* 2009;**45**:253–61. <https://doi.org/10.1016/j.jadohealth.2009.02.009>
26. 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. <https://doi.org/10.1136/bmj.319.7206.344>
27. Kaltiala-Heino R, Rimpelä M, Rantanen P, Rimpelä A. Bullying at school – an indicator of adolescents at risk for mental disorders. *J Adolesc* 2000;**23**:661–74. <https://doi.org/10.1006/jado.2000.0351>
28. Juvonen J, Graham S, Schuster MA. Bullying among young adolescents: the strong, the weak, and the troubled. *Pediatrics* 2003;**112**:1231–7. <https://doi.org/10.1542/peds.112.6.1231>
29. Copeland WE, Wolke D, Angold A, Costello EJ. Adult psychiatric outcomes of bullying and being bullied by peers in childhood and adolescence. *JAMA Psychiatry* 2013;**70**:419–26. <https://doi.org/10.1001/jamapsychiatry.2013.504>
30. Lereya ST, Copeland WE, Costello EJ, Wolke D. Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries. *Lancet Psychiatry* 2015;**2**:524–31. [https://doi.org/10.1016/S2215-0366\(15\)00165-0](https://doi.org/10.1016/S2215-0366(15)00165-0)
31. 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. <https://doi.org/10.1111/1469-7610.00629>
32. Bond L, Carlin JB, Thomas L, Rubin K, Patton G. Does bullying cause emotional problems? A prospective study of young teenagers. *BMJ* 2001;**323**:480–4. <https://doi.org/10.1136/bmj.323.7311.480>

33. Arseneault L, Bowes L, Shakoor S. Bullying victimization in youths and mental health problems: 'much ado about nothing'? *Psychol Med* 2010;**40**:717–29. <https://doi.org/10.1017/S0033291709991383>
34. Fisher HL, Moffitt TE, Houts RM, Belsky DW, Arseneault L, Caspi A. Bullying victimisation and risk of self harm in early adolescence: longitudinal cohort study. *BMJ* 2012;**344**:e2683. <https://doi.org/10.1136/bmj.e2683>
35. Winsper C, Lereya T, Zanarini M, Wolke D. Involvement in bullying and suicide-related behavior at 11 years: a prospective birth cohort study. *J Am Acad Child Adolesc Psychiatry* 2012;**51**:271–82.e3. <https://doi.org/10.1016/j.jaac.2012.01.001>
36. van der Wal MF, de Wit CA, Hirasing RA. Psychosocial health among young victims and offenders of direct and indirect bullying. *Pediatrics* 2003;**111**:1312–7. <https://doi.org/10.1542/peds.111.6.1312>
37. Glew GM, Fan MY, Katon W, Rivara FP, Kernic MA. Bullying, psychosocial adjustment, and academic performance in elementary school. *Arch Pediatr Adolesc Med* 2005;**159**:1026–31. <https://doi.org/10.1001/archpedi.159.11.1026>
38. Woods S, Wolke D. Direct and relational bullying among primary school children and academic achievement. *J School Psychol* 2004;**42**:135–55. <https://doi.org/10.1016/j.jsp.2003.12.002>
39. Crick NR, Ostrov JM, Werner NE. A longitudinal study of relational aggression, physical aggression, and children's social-psychological adjustment. *J Abnorm Child Psychol* 2006;**34**:131–42. <https://doi.org/10.1007/s10802-005-9009-4>
40. Price LH, Kao HT, Burgers DE, Carpenter LL, Tyrka AR. Telomeres and early-life stress: an overview. *Biol Psychiatry* 2013;**73**:15–23. <https://doi.org/10.1016/j.biopsych.2012.06.025>
41. Steer A. *Learning Behaviour: Lessons Learned. A Review of Behaviour Standards and Practices in Our Schools*. London: Department for Children, Families and Schools; 2009.
42. Department of Health and Social Care and Department for Education. *Transforming Children and Young People's Mental Health Provision: A Green Paper*. London: Department of Health and Social Care and Department for Education; 2017.
43. Department for Education. *Preventing and Tackling Bullying: Advice for Headteachers, Staff and Governing Bodies*. London: Department for Education; 2017.
44. 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*. York: Joseph Rowntree Foundation; 2002.
45. Baron RA, Richardson DR. *Perspectives in Social Psychology. Human Aggression*. 2nd edn. New York, NY: Plenum Press; 1994.
46. Olweus D. *Bullying at School: What We Know and What We Can Do*. Hoboken, NJ: Wiley-Blackwell; 1993.
47. 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. <https://doi.org/10.1002/cbm.799>
48. Brame B, Nagin DS, Tremblay RE. Developmental trajectories of physical aggression from school entry to late adolescence. *J Child Psychol Psychiatry* 2001;**42**:503–12. <https://doi.org/10.1111/1469-7610.00744>
49. Boxer P, Edwards-Leeper L, 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. <https://doi.org/10.1891/vivi.2003.18.6.691>

REFERENCES

50. Steer A. *Learning Behaviour: Lessons learned – A Review of Behaviour Standards and Practices in our Schools*. London: DCSF; 2009.
51. Parsonage M. *Diversion: A Better Way for Criminal Justice and Mental Health*. London: Sainsbury Centre for Mental Health; 2009.
52. Department for Education and Skills (DfES). *White Paper: The Future of Higher Education*. London: DfES; 2003.
53. Department for Children, Families and Schools (DfCFS). *Your Child, Your Schools, Our Future: Building a 21st Century Schools System*. London: The Stationery Office; 2009.
54. Department of Health and Social Care. *Healthy Lives, Brighter Futures*. London: HMSO; 2009.
55. Vreeman RC, Carroll AE. A systematic review of school-based interventions to prevent bullying. *Arch Pediatr Adolesc Med* 2007;**161**:78–88. <https://doi.org/10.1001/archpedi.161.1.78>
56. Smith JD, Schneider B, Smith PK, Ananiadou K. The effectiveness of whole-school antibullying programs: a synthesis of evaluation research. *Sch Psychol Rev* 2004;**33**:547–60.
57. Langford R, Bonell CP, Jones HE, Poulidou T, Murphy SM, Waters E, et al. The WHO health promoting school framework for improving the health and well-being of students and their academic achievement. *Cochrane Database Syst Rev* 2014;**4**:CD008958. <https://doi.org/10.1002/14651858.CD008958.pub2>
58. Bonell C, Fletcher A, McCambridge J. Improving school ethos may reduce substance misuse and teenage pregnancy. *BMJ* 2007;**334**:614–16. <https://doi.org/10.1136/bmj.39139.414005.AD>
59. 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. [https://doi.org/10.1016/S0277-9536\(02\)00120-X](https://doi.org/10.1016/S0277-9536(02)00120-X)
60. 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. <https://doi.org/10.1136/jech.2003.009449>
61. 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. <https://doi.org/10.1001/archpedi.158.4.377>
62. Viner RM, Ozer EM, Denny S, Marmot M, Resnick M, Fatusi A, Currie C. Adolescence and the social determinants of health. *Lancet* 2012;**379**:1641–52. [https://doi.org/10.1016/S0140-6736\(12\)60149-4](https://doi.org/10.1016/S0140-6736(12)60149-4)
63. 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).
64. Dahlgren G, Whitehead M. *Policies and Strategies to Promote Social Equity in Health*. Stockholm: Institute for Future Studies; 1991.
65. Marmot MG. Creating healthier societies. *Bull World Health Organ* 2004;**82**:320.
66. 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. <https://doi.org/10.1016/j.jadohealth.2004.11.012>
67. Buck D, Frosini F. *Clustering of Unhealthy Behaviours Over Time*. London: The King's Fund; 2012.
68. 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. <https://doi.org/10.2105/AJPH.2004.047399>

69. Langford R, Bonell C, Komro K, Murphy S, Magnus D, Waters E, *et al.* The Health Promoting Schools Framework: known unknowns and an agenda for future research. *Health Educ Behav* 2017;**44**:463–75. <https://doi.org/10.1177/1090198116673800>
70. 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. <https://doi.org/10.1177/109019810102800310>
71. Glover S, Butler H. Facilitating Health Promotion Within School Communities. In Moodie R, Hulme A, editors. *Hands-on Health Promotion*. Melbourne, VIC: IP Communications; 2004.
72. 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 Adolesc Health* 2010;**47**:555–63. <https://doi.org/10.1016/j.jadohealth.2010.04.011>
73. 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* 2009;**110**:252–72. <https://doi.org/10.1108/09654281011052628>
74. 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). <https://doi.org/10.3310/hta19530>
75. Morrison B. Restorative Justice in Schools. In Elliot E, Gordon RM, editors. *New Directions in Restorative Justice: Issues, Practices, Evaluation*. Cullompton: Willan Publishing; 2005.
76. Bazemore G. Young people, trouble, and crime restorative justice as a normative theory of informal social control and social support. *Youth Soc* 2001;**33**:199–226. <https://doi.org/10.1177/0044118X01033002004>
77. Pepi C. Children without childhoods: a feminist intervention strategy utilizing systems theory and restorative justice in treating female adolescent offenders. *Women Ther* 1998;**20**:85–101. https://doi.org/10.1300/J015v20n04_07
78. Farrington DP, Ttofi MM. School-based programs to reduce bullying and victimization. *Campbell Syst Rev* 2009;**6**:1–148. <https://doi.org/10.1037/e528362010-001>
79. *National Evaluation of the Restorative Justice in Schools Programme*. London: YJD, Youth Justice Board for England and Wales; 2004.
80. Lloyd G, Kane J, McCluskey G, Stead J, Riddell S. *Restorative Approaches in Scottish Schools: Transformations and Challenges*. Edinburgh: Scottish Executive Education Department; 2006.
81. Skins L, Du Rose N, Hough M. *An Evaluation of Bristol RAiS*. London: ICPR, King's College London; 2009.
82. Minnesota Department of Children, Families & Learning. *In-School Behavior Intervention Grants: A Three-Year Evaluation of Alternative Approaches to Suspensions and Expulsions. Report to the Legislature*. Roseville, MN: Minnesota Department of Children, Families & Learning; 2001.
83. Shaw G. Restorative practices in Australian schools: changing relationships, changing culture. *Conflict Resolution Q* 2007;**25**:127–35. <https://doi.org/10.1002/crq.198>
84. 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 Comp Criminol* 2011;**55**:846–62. <https://doi.org/10.1177/0306624X10374638>
85. Farrington DP, Ttofi MM. *School-Based Programs to Reduce Bullying and Victimization, Systematic Review for the Campbell Collaboration Crime and Justice Group*. Oslo: Campbell Collaboration Crime and Justice Group; 2009. <https://doi.org/10.4073/csr.2009.6>

REFERENCES

86. Buckley S, Maxwell GM. *Respectful Schools: Restorative Practices in Education: A Summary Report*. Office of the Children's Commissioner and the Institute of Policy Studies, School of Government, Victoria University, Wellington; 2007.
87. Youth Justice Board. *National evaluation of the restorative justice in schools programme*. London: Youth Justice Board Whose Justice; 2004.
88. Sumner MD, Silverman CJ, Frampton ML. *School-Based Restorative Justice as an Alternative to Zero-Tolerance Policies: Lessons From West Oakland*. Berkeley, CA: Thelton E Henderson Center for Social Justice; 2010.
89. Kokotsaki D, White C, Hopkins B. Capturing change: a review of the implementation of Restorative Approaches and its outcomes within a local authority in North East England. *Online Educ Res J* 2014;**5**:151.
90. Bitel M. *The Evaluation of Year 3 of Leicestershire Mediation Services Conflict Resolution in Schools Programme*. London: Charities Evaluation Services; 2001.
91. Youth Justice Board for England and Wales. *National Evaluation of the Restorative Justice in Schools Programme*. London: Youth Justice Board for England and Wales; 2004.
92. Durlak JA, Weissberg RP, Dymnicki AB, Taylor RD, Schellinger KB. The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Dev* 2011;**82**:405–32. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>
93. Segawa E, Ngwe JE, Li Y, Flay BR. Evaluation of the effects of the Aban Aya Youth Project in reducing violence among African American adolescent males using latent class growth mixture modeling techniques. *Eval Rev* 2005;**29**:128–48. <https://doi.org/10.1177/0193841X04271095>
94. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;**348**:g1687. <https://doi.org/10.1136/bmj.g1687>
95. Great Britain. *Data Protection Act 1998*. London: The Stationery Office; 1998.
96. 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 Sch Health* 2007;**77**:75–9. <https://doi.org/10.1111/j.1746-1561.2007.00170.x>
97. Smith DJ. *School Experience and Delinquency at Ages 13 to 16*. Edinburgh: University of Edinburgh, Centre for Law and Society; 2006.
98. Varni JW, Burwinkle TM, Seid M. The PedsQL 4.0 as a school population health measure: feasibility, reliability, and validity. *Qual Life Res* 2006;**15**:203–15. <https://doi.org/10.1007/s11136-005-1388-z>
99. 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. <https://doi.org/10.2165/11587350-000000000-00000>
100. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry* 2000;**177**:534–9. <https://doi.org/10.1192/bjp.177.6.534>
101. 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. <https://doi.org/10.1186/1471-2458-11-487>
102. Bridges S, Gill V, Omole T, Sutton R, Wright V. *Smoking, drinking and drug use among young people in England in 2010*. London: NHS Information Centre; 2011.

103. Stephenson JM, Strange V, Forrest S, Oakley A, Copas A, Allen E, et al. Pupil-led sex education in England (RIPPLE study): cluster-randomised intervention trial. *Lancet* 2004;**364**:338–46. [https://doi.org/10.1016/S0140-6736\(04\)16722-6](https://doi.org/10.1016/S0140-6736(04)16722-6)
104. Hamburger ME, Basile KC, Vivolo AM. *Measuring Bullying Victimization, Perpetration, and Bystander Experiences: A Compendium of Assessment Tools*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2011.
105. Boyce W, Torsheim T, Currie C, Zambon A. The family affluence scale as a measure of National Wealth: validation of an adolescent self-report measure. *Soc Indic Res* 2006;**78**:473–87. <https://doi.org/10.1007/s11205-005-1607-6>
106. Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V.I. *Health Qual Life Outcomes* 2007;**5**:54. <https://doi.org/10.1186/1477-7525-5-54>
107. Maslach C, Jackson SE. The measurement of experienced burnout. *J Occupational Behav* 1981;**2**:15. <https://doi.org/10.1002/job.4030020205>
108. Department for Education. *School and College Performance Tables: Statements of Intent*. London: Department for Education; 2017.
109. McLennan D, Noble M, Davies J, Garatt E, Dibben C. *The English Indices of Deprivation 2010: Technical Report*. London: Ministry of Housing, Communities and Local Government; 2011.
110. May C. Towards a general theory of implementation. *Implement Sci* 2013;**8**:18. <https://doi.org/10.1186/1748-5908-8-18>
111. Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. *Process Evaluation of Complex Interventions: UK Medical Research Council (MRC) Guidance*. London: MRC Population Health Science Research Network; 2014.
112. 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 beyond blue schools research initiative. *J Child Psychol Psychiatry* 2010;**51**:199–209. <https://doi.org/10.1111/j.1469-7610.2009.02136.x>
113. Wiggins M, Bonell C, Sawtell M, Austerberry H, Burchett H, Allen E, Strange V. Health outcomes of youth development programme in England: prospective matched comparison study. *BMJ* 2009;**339**:b2534. <https://doi.org/10.1136/bmj.b2534>
114. Green J, Thorogood N. *Qualitative Methods for Health Research*. London: Sage; 2009.
115. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ* 2015;**350**:h1258. <https://doi.org/10.1136/bmj.h1258>
116. Frymier AB, Shulman GM, Houser M. The development of a learner empowerment measure. *Communication Education* 1996;**45**:181–99. <https://doi.org/10.1080/03634529609379048>
117. Mytton J, DiGiuseppi C, Gough D, Taylor R, Logan S. School-based secondary prevention programmes for preventing violence. *Cochrane Database Syst Rev* 2006;**3**:CD004606. <https://doi.org/10.1002/14651858.CD004606.pub2>
118. Hale DR, Patalay P, Fitzgerald-Yau N, Hargreaves DS, Bond L, Görzig A, et al. School-level variation in health outcomes in adolescence: analysis of three longitudinal studies in England. *Prev Sci* 2014;**15**:600–10. <https://doi.org/10.1007/s1121-013-0414-6>
119. Tobler AL, Komro KA, Dabroski A, Aveyard P, Markham WA. Preventing the link between SES and high-risk behaviors: 'value-added' education, drug use and delinquency in high-risk, urban schools. *Prev Sci* 2011;**12**:211–21. <https://doi.org/10.1007/s1121-011-0206-9>

REFERENCES

120. Currie C, Molcho M, Boyce W, Holstein B, Torsheim T, Richter M. Researching health inequalities in adolescents: the development of the Health Behaviour in School-Aged Children (HBSC) family affluence scale. *Soc Sci Med* 2008;**66**:1429–36. <https://doi.org/10.1016/j.socscimed.2007.11.024>
121. Green J, Thorogood N. *Qualitative Methods for Health Research*. London: Sage; 2004.
122. Bonell C, Allen E, Warren E, McGowan J, Bevilacqua L, Jamal F, *et al*. Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial. *Lancet* 2018;**392**:2452–64. [https://doi.org/10.1016/S0140-6736\(18\)31782-3](https://doi.org/10.1016/S0140-6736(18)31782-3)
123. Ofsted. *Low-Level Disruption in Classrooms: Below the Radar*. London: Ofsted; 2014.
124. Angus M. *Pipeline Project: Trajectories of Classroom Behaviour and Academic Progress : A Study of Student Engagement with Learning*. Mount Lawley, WA: Edith Cowan University; 2010.
125. Lasher S, Baker C. *Bullying: Evidence from the Longitudinal Study of Young People in England 2, Wave 2*. London: Department of Education; 2015.
126. Warren E, Bevilacqua L, Opondo C, Allen E, Mathiot A, West G, *et al*. Action groups as a participative strategy for leading whole-school health promotion: results on implementation from the INCLUSIVE trial in English secondary schools. *Br Educ Res J* 2019;**45**:979–1000. <https://doi.org/10.1002/berj.3547>
127. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ* 2005;**14**:487–96. <https://doi.org/10.1002/hec.944>
128. National Institute for Health and Care Excellence. *Methods for the Development of NICE Public Health Guidance (3rd Edition)*. London: NICE; 2012.
129. National Institute for Health and Care Excellence. *Developing NICE Guidelines: The Manual*. London: NICE; 2014.
130. Department for Education. URL: www.gov.uk/government/organisations/department-for-education (accessed 29 September 2017).
131. Department of Health and Social Care. *NHS Reference Costs 2015–2016*. London: Department of Health and Social Care; 2016.
132. Curtis L, Burns A. *Unit Costs of Health and Social Care*. Canterbury: Personal Social Services Research Unit, University of Kent; 2016.
133. The King's Fund. *Improving the Quality of Care in General Practice*. London: The King's Fund; 2011.
134. Heslin M, Callaghan L, Barrett B, Lea S, Eick S, Morgan J, *et al*. Costs of the police service and mental healthcare pathways experienced by individuals with enduring mental health needs. *Br J Psychiatry* 2017;**210**:157–64. <https://doi.org/10.1192/bjp.bp.114.159129>
135. Grieve R, Nixon R, Thompson SG, Normand C. Using multilevel models for assessing the variability of multinational resource use and cost data. *Health Econ* 2005;**14**:185–96. <https://doi.org/10.1002/hec.916>
136. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Econ* 2004;**13**:461–75. <https://doi.org/10.1002/hec.843>
137. Willan AR, Briggs AH. *Statistical Analysis of Cost-effectiveness Data*. Chichester: Wiley; 2006. <https://doi.org/10.1002/0470856289>

138. Bonell C, Allen E, Strange V, Copas A, Oakley A, Stephenson J, Johnson A. The effect of dislike of school on risk of teenage pregnancy: testing of hypotheses using longitudinal data from a randomised trial of sex education. *J Epidemiol Community Health* 2005;**59**:223–30. <https://doi.org/10.1136/jech.2004.023374>
139. Higgins S, Matsipataki M, Kokotsaki D, Coe R, Major LE, Coleman RE. *The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit: Technical Appendices*. London: Educational Endowment Foundation; 2013.
140. Branson CE, Cornell DG. A comparison of self and peer reports in the assessment of middle school bullying. *J App School Psychol* 2009;**25**:5–27. <https://doi.org/10.1080/15377900802484133>
141. Hawe P, Shiell A, Riley T. Complex interventions: how 'out of control' can a randomised controlled trial be? *BMJ* 2004;**328**:1561–3. <https://doi.org/10.1136/bmj.328.7455.1561>
142. Scott S, Kearney J. *\$100m Funding Boost for School Mental Health Programs and Headspace centres*. ABC News Online, 2018. URL: www.abc.net.au/news/2018-01-08/more-dollars-for-kids-mental-health/9310172 (accessed 27 September 2017).
143. Bonell C, Wells H, Harden A, Jamal F, Fletcher A, Thomas J, et al. The effects on student health of interventions modifying the school environment: systematic review. *J Epidemiol Community Health* 2013;**67**:677–81. <https://doi.org/10.1136/jech-2012-202247>
144. Shinde S, Weiss HA, Varghese B, Khandeparkar P, Pereira B, Sharma A, et al. Promoting school climate and health outcomes with the SEHER multi-component secondary school intervention in Bihar, India: a cluster-randomised controlled trial. *Lancet* 2018;**392**:2465–77.
145. Great Britain. *The Medicines for Human Use (Clinical Trials) Regulations 2004*. 2004. URL: www.legislation.gov.uk/ukxi/2004/1031/contents/made (accessed 27 September 2017).
146. Medical Research Council. *MRC Guidelines on Good Clinical Practice in Clinical Trials*. London: Medical Research Council; 1998.

Appendix 1 Tables and figures

TABLE 26 Comparison of Inclusive study schools with non-recruited schools and average for secondary schools in England

	England average	Inclusive trial schools				Non-recruited schools				Comparison of recruited and non-recruited schools
		N	Mean (SD)	Median	Range	N	Mean (SD)	Median	Range	p-value
School IDACI (mean)	Not available	40	0.262 (0.198)	0.239	0.028–0.698	427	0.251 (0.186)	0.199	0.007–0.765	0.5*
School population size (mean)	939	40	1081 (324)	1075	446–1786	427	1073 (379)	1058	60–2405	0.9
Special educational need (SEN) (%)	2.8	40	5.2 (4.4)	5.9	0.03–14.4	422	5.6 (5.9)	5.1	0.1–42.7	0.5*
English as a foreign language (%)	13.60	40	34.7 (25.9)	29.5	3.3–90.3	424	30.5 (25.0)	23.3	0–92.8	0.2*
Eligible for free school meals (%)	16.30	40	36.8 (19.4)	38.1	3.9–75.8	427	34.7 (20.8)	31.2	1.6–94.0	0.3*
Number with ≥ 5 GCSEs at A to C grade (%)	59.30	40	60.6 (14.5)	56.5	35–99	399	63.4 (16.0)	63.0	14–100	0.3
Value added (best 8) (median score)	1000	40	1013 (24)	1014	925–1064	399	1015 (22)	1016	949–1081	0.3
Student absence (annual) (%)	5.8	40	5.2 (1.2)	5.1	2.7–9.1	419	5.0 (1.0)	4.9	2.1–10.5	0.3
OFSTED rating Outstanding	12%	13	32.5%			144	33.8%			
Good	60%	23	57.5%			194	45.5%			
Requires improvement	15%	3	7.5%			68	16.0%			
Inadequate	13%	0	0%			3	0.7%			
Not available	–	1	2.5%			17	4.0%			
Combined either outstanding or good	72%	40	90%			426	79.3%			0.04
School sex makeup		40	%			427	%			
Boys only		2	5			28	6.6			0.9
Mixed sex		32	80			356	83.4			
Girls only		6	15			43	10.1			

IDACI, Index of Deprivation Affecting Children.

* Median equality test used.

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license

Ofsted ratings were not available for one INCLUSIVE school and a number of non-recruited schools for a variety of reasons, including being new schools that had not yet been rated.

Data sources for England averages

Because recruitment was undertaken in 2014, we used the most recently available data for 2013/14. Data for England averages were taken from Department for Education statistics (www.gov.uk/government/organisations/department-for-education/about/statistics, accessed 24 April 2017) and www.compare-school-performance.service.gov.uk/ (accessed 25 July 2017).

TABLE 27 Demographic characteristics of participants at baseline, 24 months and 36 months

Outcome	Baseline		24 months		36 months	
	Control n = 3347	Intervention n = 3320	Control n = 3195	Intervention n = 3095	Control n = 3087	Intervention n = 2873
Age, mean (SD)	11.75 (0.44)	11.76 (0.43)	13.72 (0.59)	13.76 (0.45)	14.75 (1.20)	14.70 (0.82)
Sex, n (%)						
Male	1639 (49.85)	1464 (44.88)	1580 (49.45)	1357 (43.84)	1462 (47.36)	1232 (42.88)
Female	1649 (50.15)	1804 (55.20)	1543 (48.29)	1675 (54.12)	1521 (49.27)	1563 (54.40)
Ethnicity, n (%)						
White British	1391 (41.47)	1221 (37.32)	1265 (39.59)	1028 (33.21)	1127 (36.51)	966 (33.62)
White other	291 (8.78)	273 (8.34)	308 (9.64)	285 (9.21)	322 (10.43)	273 (9.50)
Asian/Asian British	859 (25.92)	786 (24.02)	837 (26.20)	762 (24.62)	816 (26.43)	687 (23.91)
Black/Black British	384 (11.59)	535 (16.35)	365 (11.42)	489 (15.80)	377 (12.21)	473 (16.46)
Chinese/Chinese British	11 (0.33)	35 (1.07)	14 (0.44)	29 (0.94)	15 (0.49)	26 (0.90)
Mixed ethnicity	238 (7.18)	224 (6.85)	231 (7.23)	229 (7.40)	217 (7.03)	191 (6.65)
Other	140 (4.22)	198 (6.05)	145 (4.54)	232 (7.50)	170 (5.51)	219 (7.62)
Religion, n (%)						
None	983 (29.59)	787 (23.99)	1037 (32.46)	808 (26.11)	1043 (33.79)	802 (27.92)
Christian	1073 (32.30)	1173 (35.76)	923 (28.89)	1019 (32.92)	871 (28.22)	900 (31.33)
Jewish	9 (0.27)	13 (0.40)	11 (0.34)	17 (0.55)	15 (0.49)	18 (0.63)
Muslim/Islam	878 (26.46)	817 (24.91)	843 (26.38)	774 (25.01)	810 (26.24)	726 (25.27)
Hindu	90 (2.71)	176 (5.37)	79 (2.47)	149 (4.81)	83 (2.69)	144 (5.01)
Sikh	71 (2.14)	88 (2.68)	69 (2.16)	87 (2.81)	70 (2.27)	81 (2.82)
Do not know	145 (4.36)	126 (3.84)	148 (4.63)	114 (3.68)	91 (2.95)	91 (3.17)
Other	73 (2.20)	100 (3.05)	57 (1.78)	85 (2.75)	69 (2.24)	67 (2.33)

continued

APPENDIX 1

TABLE 27 Demographic characteristics of participants at baseline, 24 months and 36 months (continued)

Outcome	Baseline		24 months		36 months	
	Control n = 3347	Intervention n = 3320	Control n = 3195	Intervention n = 3095	Control n = 3087	Intervention n = 2873
Family structure, n (%)						
Two parents	2393 (71.91)	2369 (72.05)	2200 (68.86)	2137 (69.05)	2073 (67.15)	1958 (68.15)
Single mothers	604 (18.15)	626 (19.04)	627 (19.62)	649 (20.97)	624 (20.21)	584 (20.33)
Single fathers	37 (1.11)	56 (1.70)	52 (1.63)	60 (1.94)	51 (1.65)	62 (2.16)
Reconstituted	246 (7.39)	204 (6.20)	237 (7.42)	182 (5.88)	240 (7.77)	177 (6.16)
Other	48 (1.44)	33 (1.00)	63 (1.97)	37 (1.20)	65 (2.11)	56 (1.95)
At least one parent in work, n (%)						
No	298 (8.73)	233 (7.18)	187 (5.85)	200 (6.46)	179 (5.80)	155 (5.40)
Yes	2437 (74.03)	2381 (73.35)	2648 (82.88)	2529 (81.71)	2612 (84.61)	2423 (84.34)
Do not know	566 (11.19)	632 (19.47)	317 (9.92)	313 (10.11)	242 (7.84)	246 (8.56)
Housing tenure, n (%)						
One rented from the council or housing association	474 (14.41)	559 (17.26)	523 (16.37)	612 (19.77)	575 (18.63)	629 (21.89)
One rented from a landlord	391 (11.88)	396 (12.23)	382 (11.96)	428 (13.83)	438 (14.19)	423 (14.72)
One owned by my family	1451 (44.10)	1273 (39.31)	1569 (49.11)	1354 (43.75)	1511 (48.95)	1312 (45.67)
Other	62 (1.88)	59 (1.82)	55 (1.72)	45 (1.45)	41 (1.33)	30 (1.04)
Do not know	912 (27.72)	951 (29.37)	616 (19.28)	600 (19.39)	467 (15.13)	422 (14.69)
Family affluence scale, mean (SD)	6.03 (1.84)	6.10 (1.81)	6.13 (1.77)	6.16 (1.76)	6.03 (1.81)	6.17 (1.75)

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

TABLE 28 Staff outcomes at baseline

	Control, mean (SD)	Intervention, mean (SD)	Overall, mean (SD)
Staff Stress and Burnout – Maslach Burnout Inventory			
Emotional exhaustion score	22.49 (11.28)	22.17 (11.36)	22.34 (11.32)
Depersonalisation score	5.52 (5.07)	5.48 (5.48)	5.51 (5.08)
Personal achievement score	35.30 (7.14)	35.69 (6.99)	35.48 (7.07)
Staff QoL – SF-12 v2 Health Survey			
Physical health score	55.22 (7.31)	56.01 (6.32)	55.59 (6.88)
Mental health score	44.64 (9.63)	44.91 (9.62)	44.77 (9.62)

TABLE 29 Subgroup analyses by sex, baseline bullying and baseline aggression for key primary and secondary analyses at 36 months

Measure	Main adjusted effect		By sex			By baseline bullying			By baseline aggression		
	Effect size*	p-value	Boys (95% CI)	Girls (95% CI)	<i>p</i> _{interaction}	Low (95% CI)	High (95% CI)	<i>p</i> _{interaction}	Low (95% CI)	High (95% CI)	<i>p</i> _{interaction}
Continuous outcomes											
GBS overall score	-0.03 (-0.06 to -0.001)	0.0441	-0.04 (-0.08 to 0.001)	-0.03 (-0.06 to 0.01)	0.6113	0.15 (0.12 to 0.18)	-0.41 (-0.45 to -0.36)	< 0.0001	0.01 (-0.03 to 0.05)	-0.06 (-0.10 to -0.02)	0.0024
ESYTC overall score	-0.13 (-0.43 to 0.18)	0.4199	-0.33 (-0.73 to 0.06)	0.04 (-0.32 to 0.39)	0.0890	0.02 (-0.34 to 0.37)	-0.17 (-0.63 to 0.29)	0.4422	0.71 (0.33 to 1.10)	-0.65 (-1.03 to -0.27)	< 0.0001
PedsQL overall score	1.44 (0.70 to 2.17)	0.0001	3.85 (2.89 to 4.80)	-0.41 (-1.28 to 0.46)	< 0.0001	0.34 (-0.52 to 1.19)	3.93 (2.81 to 5.04)	< 0.0001	-0.14 (-1.13 to 0.85)	2.70 (1.74 to 3.65)	< 0.0001
SDQ total difficulties score	-0.54 (-0.83 to -0.25)	0.0002	-1.29 (-1.67 to -0.92)	0.04 (-0.30 to 0.39)	< 0.0001	-0.08 (-0.42 to 0.26)	-1.61 (-2.05 to -1.17)	< 0.0001	0.34 (-0.05 to 0.72)	-1.31 (-1.69 to -0.94)	< 0.0001
SWEMWBS total well-being index	0.33 (0.00 to 0.66)	0.0487	1.32 (0.89 to 1.74)	-0.42 (-0.81 to -0.04)	< 0.0001	0.14 (-0.24 to 0.53)	0.93 (0.43 to 1.43)	0.0034	-0.18 (-0.62 to 0.26)	0.78 (0.36 to 1.21)	0.0001
Age at sexual debut (years)	-0.35 (-1.48 to 0.78)	0.5409	-0.63 (-1.99 to 0.74)	0.01 (-1.51 to 1.53)	0.4879	-0.14 (-1.34 to 1.05)	-0.46 (-1.89 to 0.98)	0.7085	-0.63 (-2.40 to 1.15)	-0.42 (-1.58 to 0.73)	0.8308
Modified aggression scale	-0.26 (-0.57 to 0.05)	0.0978	-0.53 (-0.89 to -0.18)	-0.03 (-0.37 to 0.31)	0.0029	-0.18 (-0.52 to 0.17)	-0.22 (-0.62 to 0.18)	0.8100	0.03 (-0.33 to 0.39)	-0.25 (-0.60 to 0.10)	0.0933
Categorical outcomes											
Ever smoked	0.58 (0.43 to 0.80)	0.0009	0.33 (0.22 to 0.50)	0.87 (0.60 to 1.25)	< 0.0001	0.66 (0.46 to 0.96)	0.52 (0.33 to 0.80)	0.2887	0.93 (0.58 to 1.49)	0.46 (0.32 to 0.68)	0.0053
Ever drunk alcohol	0.72 (0.56 to 0.92)	0.0094	0.52 (0.38 to 0.70)	0.95 (0.71 to 1.26)	0.0002	0.81 (0.61 to 1.07)	0.57 (0.41 to 0.81)	0.0541	1.08 (0.78 to 1.49)	0.56 (0.41 to 0.75)	0.0009

continued

TABLE 29 Subgroup analyses by sex, baseline bullying and baseline aggression for key primary and secondary analyses at 36 months (*continued*)

Measure	Main adjusted effect		By sex			By baseline bullying			By baseline aggression		
	Effect size*	p-value	Boys (95% CI)	Girls (95% CI)	p _{interaction}	Low (95% CI)	High (95% CI)	p _{interaction}	Low (95% CI)	High (95% CI)	p _{interaction}
Been offered illicit drugs	0.51 [†] (0.36 to 0.73)	0.0003	0.44 [†] (0.29 to 0.68)	0.57 (0.38 to 0.86)	0.2350	0.46 [†] (0.31 to 0.68)	0.58 [†] (0.37 to 0.93)	0.2895	0.62 [†] (0.40 to 0.96)	0.50 [†] (0.34 to 0.74)	0.3137
Used any contraception at first sex	1.08 (0.50 to 2.35)	0.8410	1.02 (0.39 to 2.69)	1.19 (0.36 to 3.90)	0.8409	0.55 (0.20 to 1.57)	2.63 (0.61 to 11.35)	0.0905	0.18 (0.02 to 1.45)	1.35 (0.45 to 4.13)	0.1003
Use of NHS in past 12 months	0.96 (0.82 to 1.11)	0.5652	0.96 (0.79 to 1.17)	0.95 (0.80 to 1.14)	0.9590	0.96 (0.81 to 1.16)	1.02 (0.80 to 1.29)	0.6717	1.12 (0.90 to 1.38)	0.93 (0.76 to 1.14)	0.1556
Contact with police in past 12 months	0.74 (0.56 to 0.97)	0.0269	0.62 (0.45 to 0.85)	0.93 (0.66 to 1.31)	0.0371	0.66 (0.48 to 0.92)	0.79 (0.53 to 1.17)	0.4280	0.93 (0.58 to 1.47)	0.67 (0.49 to 0.93)	0.1917

*Effects are difference (95% CI) for continuous outcomes and odds ratios (95% CI) for categorical outcomes.

[†]Proportional odds ratio.

Low and high baseline bullying were defined on the basis of the GBS. High was defined as at least weekly experience of bullying or being upset by it, low was defined as less than weekly experience of bullying and not being upset by it. Low and high aggression were defined on the basis of the ESYTC school misbehaviour subscale, with high levels of behaviour problems defined as scores > 0 and low levels defined as scores of 0.

Reproduced from Bonell *et al.*¹²⁷ © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Bullying

Low and high bullying subgroups were defined on the basis of baseline bullying experience based on the GBS. High was defined as frequent – at least weekly – experience of bullying or being upset by it; low or medium was defined as rare – less than weekly – experience of bullying and not being upset by it.

Aggression

Low and high aggression subgroups were defined on the basis of baseline behaviour problems based on the ESYTC, with high levels of behaviour problems defined as ESYTC scores greater than zero and low levels defined as scores of zero.

TABLE 30 Effect of fidelity score on the primary outcomes at 24 months and 36 months

Measure	24 months				36 months			
	Unadjusted (95% CI)	p-value	Adjusted (95% CI)	p-value	Unadjusted (95% CI)	p-value	Adjusted (95% CI)	p-value
GBS overall score	-0.02 (-0.04 to -0.003)	0.0206	-0.02 (-0.04 to -0.003)	0.0220	-0.001 (-0.01 to 0.01)	0.7433	-0.001 (-0.01 to 0.01)	0.8019
ESYTC overall score	-0.14 (-0.34 to 0.07)	0.1949	-0.12 (-0.32 to 0.08)	0.2297	0.02 (-0.08 to 0.11)	0.7600	0.02 (-0.07 to 0.12)	0.5985

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

TABLE 31 Process evaluation response rates

			Year 1	Year 2	Year 3	Total
All schools	Interviews	Facilitators	8/8	8/8	NA	16/16
		AGM members	28/40	41/40	28/40	97/120
		Senior leadership team	36/40	NA	35/40	71/80
		Other staff	66/80	NA	NA	66/80
		Curriculum deliverers	6/20	10/20	8/20	24/60
	Observations	AGMs	10/10	9/10	7/10	26/30
		All-staff training	9/10	NA	NA	9/10
	Surveys	Curriculum	20/60	21/60	2/20	43/140
		In-depth training satisfaction	129/100 ^a	NA	NA	129/100
		AGM members	228/240 ^b	184/240 ^b	32/240 ^b	444/720
		Staff delivering RP	189/300 ^c	82/300 ^c	34/100 ^d	305/700
	Diaries	Facilitators	115/120 ^e	101/120 ^e	NA	118/240
		Trainers	18/20 ^f	NA	NA	18/20
	Minutes	AGMs	115/120	101/120	NA	216/240
Case study schools		FGDs	Students	12/12	10/12	8/12
	Staff		6/6	5/6	4/6	15/18
	Interviews	Students involved in RP	12/12	6/12	6/12	24/36

a Denominator assumes 5 trainees per intervention school but this varied between schools and averaged > 5.

b Denominator assumes 12 AGM members per intervention school.

c Denominator assumes 5 staff per intervention school per term.

d Denominator assumes 5 staff per intervention school per year.

e Denominator assumes 6 meetings per year for 20 schools.

f Denominator assumes one all-staff training session for 20 schools.

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

APPENDIX 1

TABLE 32 Overall fidelity in years 1–2 and 3

Intervention element		N schools implementing this element	
		Year 1/2	Year 3
Action groups	<i>Minutes or diaries indicate min. 6 meetings in years 1 and 2</i>	11	3
	<i>Minutes or diaries indicate review of policies/rules in year 1 or 2</i>	17	NA
	<i>Minutes or diaries indicate implementation of locally decided actions in years 1 and 2</i>	16	8
	<i>Survey of members indicated good range of students and staff members (mean score for two measures at or below 2)</i>	20	NA
	<i>Survey of members indicated well led (mean score at or below 2)</i>	20	NA
Curriculum	<i>Surveys or interviews indicates five hours/> 1 unit delivered in years 1 and 2</i>	9	5
RP	<i>Attendance logs indicated at least five staff received in-depth training</i>	15	NA
	<i>Staff survey indicates at least 85% staff report that if there is trouble at this school, staff respond by talking to those involved to help them get on better</i>	12	15
Median fidelity score		6/8	1/4

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

TABLE 33 All-staff training implementation

Training implementation	Fidelity self-reported in trainer diary			Schools with no data
	N (%) of schools			
	As intended	Less than intended	Not at all	
Covered 'what is restorative justice/practice'	19 (95)	0	0	1 (5)
Covered 'the importance of language'	19 (95)	0	0	1 (5)
Covered 'what we do to challenge bad behaviour/nature of challenge'	19 (95)	0	0	1 (5)
Covered 'the importance of emotions'	19 (95)	0	0	1 (5)
Covered 'the importance of listening'	19 (95)	0	0	1 (5)
Covered 'maintaining the relationship after difficult conversation'	16 (80)	3 (15)	0	1 (5)
Used Power point slides	19 (95)	0	0	1 (5)
Used DVD	18 (90)	0	1 (5)	1 (5)
Facilitated paired activity	19 (95)	0	0	1 (5)
Facilitated small group activity	19 (95)	0	0	1 (5)

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

TABLE 34 Participant satisfaction with in-depth training

Measure (response option which indicates satisfaction)	Responses	
	Overall % members responding as indicated	N schools where > 70% of members responded as indicated
Do you feel you learnt useful skills at this training? (yes definitely or yes probably)	100%	20
Do you feel confident in putting into practice the skills you have learnt today? (yes)	91%	20
Do you intend to put the skills you learnt to use in your everyday practice? (yes)	95%	20
Would you recommend this training to a colleague? (yes)	100%	20
Overall, how would you rate this training? (good or excellent)	99%	20

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

TABLE 35 Member satisfaction with action groups

Measure (response option which indicates satisfaction)	Responses								
	Year 1			Year 2			Year 3		
	Overall % members responding as indicated	N schools where > 70% of members responded as indicated	N schools with no data	Overall % members responding as indicated	N schools where > 70% of members responded as indicated	N schools with no data	Overall % members responding as indicated	N schools where > 70% of members responded as indicated	N schools with no data
Was the needs-assessment report useful in helping the action group decide what actions to take? (somewhat or very)	91%	17	1	88%	14	3	76%	2	17
Was the external facilitator useful in ensuring that all action group members could have their say? (somewhat or very)	89%	18	1	90%	15	3	NA	NA	NA
Was the external facilitator useful in helping the action group decide what actions to take? (somewhat or very)	86%	16	1	88%	17	3	NA	NA	NA
Was the external facilitator useful in helping to ensure that actions were actually implemented? (somewhat or very)	80%	13	1	79%	13	3	NA	NA	NA
Do you think the LT Project was a good way to ensure students contribute to decision-making at this school? (very or quite)	95%	19	1	94%	16	3	100%	3	17
Overall do you think the action group made good decisions about what actions to take? (% very or quite)	94%	19	1	93%	16	3	98%	3	17
Do you think the action group made sure that these actions were implemented? (% yes)	70%	10	1	72%	10	3	69%	1	17

* No external facilitator in Year 3.
 Reproduced from Bonell et al.¹⁷² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

TABLE 36 Response rates

			Year 1	Year 2	Year 3	Total
All schools	Interviews	Facilitators	8/8	8/8	NA	16/16
		AGM members	28/40	41/40	28/40	97/120
		Senior leadership team	36/40	NA	35/40	71/80
		Other staff	66/80	NA	NA	66/80
		Curriculum deliverers	6/20	10/20	8/20	24/60
	Observations	AGMs	10/10	9/10	7/10	26/30
		All-staff training	9/10	NA	NA	9/10
	Surveys	Curriculum	20/60	21/60	2/20	43/140
		In-depth training satisfaction	129/100 ^a	NA	NA	129/100
		AGM members	228/240 ^b	184/240 ^b	32/240 ^b	444/720
	Diaries	Staff delivering RP	189/300 ^c	82/300 ^c	34/100 ^d	305/700
		Facilitators	115/120 ^e	3/120 ^e	NA	118/240
		Trainers	18/20 ^f	NA	NA	18/20
	Minutes	AGMs	115/120	101/120	NA	216/240
Case study schools	FGDs	Students	12/12	10/12	8/12	30/36
		Staff	6/6	5/6	4/6	15/18
	Interviews	Students involved in RP	12/12	6/12	6/12	24/36

a Denominator assumes 5 trainees per intervention school but this varied between schools and averaged > 5.

b Denominator assumes 12 AGM members per intervention school.

c Denominator assumes 5 staff per intervention school per term.

d Denominator assumes 5 staff per intervention school per year.

e Denominator assumes 6 meetings per year for 20 schools.

f Denominator assumes one all-staff training session for 20 schools.

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

TABLE 37 Overall fidelity in years 1–2

School	AGs					Curriculum	RP		Overall score/8
	Minutes or diaries indicated minimum of six meetings in years 1 and 2	Minutes or diaries indicated review of policies/rules in year 1 or 2	Minutes or diaries indicated implementation of locally decided actions in years 1 and 2	Survey of members indicated good range of students and staff members (mean score for two measures at or below 2)	Survey of members indicated well led (mean score at or below 2)	Surveys or interviews indicated five hours/> 1 unit delivered in years 1 and 2	Attendance logs indicated at least five staff received in-depth training	Staff survey indicated at least 85% of staff report that if there is trouble at this school, staff respond by talking to those involved to help them get on better	
BE	1	1	1	1	1	0	1	0	6
AW	0	1	1	1	1	1	1	1	7
AO	0	1	1	1	1	0	1	1	6
AZ	0	1	0	1	1	1	1	1	6
AX	1	1	0	1	1	1	1	1	7
AS	1	0	0	1	1	1	0	1	5
AD	1	1	1	1	1	1	1	1	8
AK	1	1	1	1	1	0	1	0	6
BD	1	0	1	1	1	0	0	0	4
AU	1	1	1	1	1	0	1	1	7
AF	0	0	1	1	1	1	1	0	5
AH	1	1	0	1	1	1	1	0	6
AM	0	1	1	1	1	1	0	0	5
BK	0	1	1	1	1	0	0	1	5
AE	1	1	1	1	1	1	1	1	8
BI	0	1	1	1	1	0	1	0	5
BM	0	1	1	1	1	0	0	0	4
AT	0	1	1	1	1	0	1	1	6
BC	1	1	1	1	1	0	1	1	7
AL	1	1	1	1	1	0	1	1	7

TABLE 38 Overall fidelity in year 3

School	AGs		Curriculum	RP	Overall score
	Interviews indicate minimum of six meetings in year 3	Interviews indicate implementation of locally decided actions in year 3	Surveys or interviews indicates five hours/> 1 unit delivered in year 3	Staff survey indicates at least 85% of staff report that if there is trouble at this school, staff respond by talking to those involved to help them get on better	
BE	0	0	0	1	1
AW	1	1	0	1	3
AO	0	1	0	1	2
AZ	0	0	0	1	1
AX	0	0	0	1	1
AS	1	1	1	1	4
AD	0	0	1	1	2
AK	0	1	0	1	2
BD	0	1	0	1	2
AU	0	0	0	1	1
AF	0	0	0	1	1
AH	0	1	0	1	2
AM	0	0	0	0	0
BK	0	0	0	0	0
AE	0	0	0	0	0
BI	0	0	0	0	0
BM	0	0	0	0	0
AT	0	1	1	1	3
BC	1	1	1	1	4
AL	0	1	1	1	3

TABLE 39 Staff and students participating in intervention activities

School	<i>N</i> staff participating in all-staff training	<i>N</i> staff participating in in-depth training	<i>N</i> staff in total participating in AGM (year 1)	<i>N</i> staff in total participating in AGM (year 2)	<i>N</i> students in total participating in AGM (year 1)	<i>N</i> students in total participating in AGM (year 2)
BE	84	7	11	7	13	12
AW	58	5	5	3	6	6
AO	90	7	7	8	8	8
AZ	Unknown	6	12	Unknown	12	7
AX	Unknown	6	7	5	8	7
AS	93	5	7	7	10	10
AD	131	8	6	6	6	7
AK	153	8	6	7	5	5
BD	162	5	6	Unknown	7	7
AU	93	2	8	3	8	8
AF	150	7	3	8	14	14
AH	145	9	9	2	6	8
AM	80	4	9	2	9	9
BK	60	3	Unknown	4	Unknown	10
AE	34	13	6	6	9	23
BI	Unknown	5	6	3	7	10
BM	88	4	11	5	5	9
AT	124	5	8	5	9	4
BC	Unknown	8	7	7	6	7
AL	103	8	7	9	6	5
Overall mean	103	6.25	7.4	5.4	8.1	8.8

TABLE 40 Intervention effects on student primary and secondary outcomes at 36 months before and after adjusting for Beyond Blue School Climate Questionnaire and anti-school peer group mediators at 24 months

Measure	Arm		Unadjusted effect		Adjusted effect (baseline covariates)		Adjusted effect (baseline covariates plus BBSQ)		Adjusted effect (baseline covariates plus YPDP)	
	Control, mean (SE)	Intervention, mean (SE)	Difference (95% CI)	p-value	Difference (95% CI)	p-value	Difference (95% CI)	p-value	Difference (95% CI)	p-value
Continuous student outcomes										
GBS overall score	0.34 (0.02)	0.29 (0.02)	-0.03 (-0.06 to -0.00)	0.039	-0.03 (-0.06 to -0.00)	0.044	-0.04 (-0.08 to -0.01)	0.007	-0.04 (-0.07 to -0.01)	0.016
ESYTC overall score	4.33 (0.20)	4.04 (0.21)	-0.07 (-0.38 to 0.25)	0.684	-0.13 (-0.43 to 0.18)	0.421	-0.20 (-0.52 to 0.13)	0.229	-0.12 (-0.44 to 0.20)	0.469
PedsQL overall score	78.82 (0.54)	80.65 (0.55)	1.16 (0.41 to 1.90)	0.002	1.44 (0.70 to 2.17)	< 0.001	1.42 (0.64 to 2.21)	< 0.001	1.26 (0.49 to 2.03)	0.001
SDQ total difficulties score	12.20 (0.18)	11.51 (0.19)	-0.51 (-0.80 to -0.22)	< 0.001	-0.54 (-0.83 to -0.25)	< 0.001	-0.62 (-0.92 to -0.31)	< 0.001	-0.54 (-0.85 to -0.24)	< 0.001
SWEMWBS total well-being index	22.88 (0.19)	23.32 (0.19)	0.27 (-0.06 to 0.60)	0.115	0.33 (0.00 to 0.66)	0.048	0.27 (-0.08 to 0.62)	0.141	0.23 (-0.13 to 0.58)	0.210
CHU9D overall score	0.85 (0.00)	0.86 (0.01)	0.01 (-0.00 to 0.01)	0.244	0.01 (-0.00 to 0.01)	0.080	0.00 (-0.00 to 0.01)	0.153	0.00 (-0.00 to 0.01)	0.217
Age of sexual debut	13.11 (0.43)	12.54 (0.49)	-0.58 (-1.97 to 0.81)	0.416	-0.35 (-1.48 to 0.78)	0.541	-0.24 (-1.48 to 1.00)	0.703	-0.40 (-1.65 to 0.84)	0.525
MAS bullying subscale score	2.75 (0.21)	2.33 (0.21)	-0.28 (-0.84 to 0.29)	0.334	-0.26 (-0.57 to 0.05)	0.097	-0.29 (-0.57 to -0.00)	0.047	-0.16 (-0.45 to 0.13)	0.279

TABLE 40 Intervention effects on student primary and secondary outcomes at 36 months before and after adjusting for Beyond Blue School Climate Questionnaire and anti-school peer group mediators at 24 months (*continued*)

Measure	Arm		Unadjusted effect		Adjusted effect (baseline covariates)		Adjusted effect (baseline covariates plus BBSQ)		Adjusted effect (baseline covariates plus YPDP)	
	Control, n (%)	Intervention, n (%)	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value
Categorical student outcomes										
Ever smoked										
No	2293 (77.70)	2318 (84.17)	0.59 (0.43 to 0.81)	0.001	0.58 (0.43 to 0.80)	0.001	0.55 (0.39 to 0.78)	0.001	0.57 (0.40 to 0.81)	0.002
Yes	658 (22.30)	436 (15.83)								
Ever drunk alcohol										
No	1677 (56.43)	1735 (62.43)	0.75 (0.58 to 0.79)	0.029	0.72 (0.56 to 0.92)	0.009	0.72 (0.55 to 0.95)	0.019	0.74 (0.56 to 0.97)	0.028
Yes	1295 (43.57)	1044 (37.57)								
Ever been really drunk										
No	788 (53.14)	721 (61.21)	0.50 (0.29 to 0.87)	0.014	0.47 (0.31 to 0.71)	< 0.001	0.42 (0.27 to 0.65)	< 0.001	0.47 (0.31 to 0.71)	< 0.001
Yes	695 (46.86)	457 (38.79)								
Been offered illicit drugs										
No	1913 (64.41)	1997 (72.54)	0.52 (0.34 to 0.79)	0.002	0.51 (0.36 to 0.73)	< 0.001	0.44 (0.30 to 0.64)	< 0.001	0.51 (0.36 to 0.73)	< 0.001
Yes, but did not try them	744 (25.05)	567 (20.60)								
Yes, and tried them	313 (10.54)	189 (6.87)								
Used contraception at last sex										
No	64 (23.10)	36 (21.95)	1.18 (0.56 to 2.48)	0.658	1.08 (0.50 to 2.35)	0.841	1.60 (0.65 to 3.91)	0.305	1.34 (0.56 to 3.22)	0.511
Yes	213 (76.90)	128 (78.05)								

Measure	Arm		Unadjusted effect		Adjusted effect (baseline covariates)		Adjusted effect (baseline covariates plus BBSQ)		Adjusted effect (baseline covariates plus YPDP)	
	Control, n (%)	Intervention, n (%)	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value
Use of the NHS in past 12 months										
No	1605 (53.22)	1472 (52.59)	0.96 (0.83 to 1.12)	0.639	0.96 (0.82 to 1.11)	0.565	0.96 (0.81 to 1.13)	0.600	0.99 (0.84 to 1.17)	0.916
Yes	1411 (46.78)	1327 (47.41)								
Contact with police										
No	2626 (86.52)	2485 (88.43)	0.75 (0.57 to 0.99)	0.041	0.74 (0.56 to 0.97)	0.027	0.66 (0.49 to 0.88)	0.004	0.72 (0.54 to 0.95)	0.022
Yes	409 (13.48)	325 (11.57)								

TABLE 41 Items and sources (tool) of resource use items and HRQoL

Measure sub-components	Tool	Questions
Trainer time	Costed based on trainer invoices	
Facilitator cost	Costed based on facilitator invoices	
School staff time		
Training	Trainer attendance sheet	Questions 20 and 21
AGM	Facilitator diary forms	
Preparation of AGM	AG questionnaire	
Staff time dealing with bullying	Staff survey	
Curriculum delivery	Curriculum log	
Delivery of bullying strategy in control schools	Staff interviews	
Staff costs	Salaries from routinely collected data	
Other resource use		
NHS resource use	Questions in the student survey	Questions 34 and 35
Police resource use	Questions in the student survey	Questions 31 and 32
Other resource costs	PSSRU, police data	
Health-related quality of life (HRQoL)		
Student HRQoL	CHU-9D in student survey	Question 50
Staff HRQoL	SF12 in teacher survey	Questions 23 to 29

TABLE 42 Student HRQoL utility scores at baseline

Measure sub-components	1, N (%)	2, N (%)	3, N (%)	4, N (%)	5, N (%)
Control					
Worried	2249 (70.28)	567 (17.72)	212 (6.63)	100 (3.13)	72 (2.25)
Sad	2508 (78.35)	406 (12.68)	152 (4.75)	75 (2.34)	60 (1.87)
Pain	2163 (67.59)	640 (20.00)	236 (7.38)	99 (3.09)	62 (1.94)
Tired	983 (30.67)	1127 (35.16)	500 (15.60)	332 (10.36)	263 (8.21)
Annoyed	2356 (73.53)	493 (15.39)	162 (5.06)	103 (3.21)	90 (2.81)
Sleep	1890 (59.06)	729 (22.78)	321 (10.03)	146 (4.56)	114 (3.56)
School work	2176 (68.11)	679 (21.25)	239 (7.48)	64 (2.00)	37 (1.16)
Daily routine	2646 (83.05)	390 (12.24)	104 (3.26)	25 (0.78)	21 (0.66)
Activities	2197 (69.11)	512 (16.11)	187 (5.88)	142 (4.47)	141 (4.44)
Intervention					
Worried	2093 (68.09)	605 (19.68)	203 (6.60)	109 (3.55)	64 (2.08)
Sad	2322 (75.64)	429 (13.97)	151 (4.92)	100 (3.26)	68 (2.21)
Pain	2114 (68.88)	591 (19.26)	226 (7.36)	90 (2.93)	48 (1.56)
Tired	938 (30.44)	1092 (35.44)	469 (15.22)	319 (10.35)	263 (8.54)
Annoyed	2159 (70.14)	514 (16.70)	194 (6.30)	102 (3.31)	109 (3.54)
Sleep	1868 (60.71)	692 (22.49)	276 (8.97)	133 (4.32)	108 (3.51)
School work	2103 (68.57)	661 (21.55)	204 (6.65)	64 (2.09)	35 (1.14)
Daily routine	2587 (84.54)	343 (11.21)	86 (2.81)	21 (0.69)	23 (0.75)
Activities	2150 (70.58)	456 (14.97)	195 (6.40)	132 (4.33)	113 (3.71)

1 indicates 'no problems', 5 indicates 'lots of problems'.

APPENDIX 1

TABLE 43 Student HRQoL utility scores at 24 months

Measure sub-components	1, N (%)	2, N (%)	3, N (%)	4, N (%)	5, N (%)
Control					
Worried	2074 (66.18)	579 (18.47)	237 (7.56)	164 (5.23)	80 (2.55)
Sad	2366 (75.54)	407 (12.99)	162 (5.17)	120 (3.83)	77 (2.46)
Pain	2177 (69.51)	560 (17.88)	237 (7.57)	93 (2.97)	65 (2.08)
Tired	815 (26.01)	949 (30.28)	556 (17.74)	463 (14.77)	351 (11.20)
Annoyed	2072 (66.09)	563 (17.96)	225 (7.18)	130 (4.15)	145 (4.63)
Sleep	1765 (56.43)	754 (24.10)	318 (10.17)	165 (5.27)	126 (4.03)
School work	2099 (67.28)	666 (21.35)	236 (7.56)	77 (2.47)	42 (1.35)
Daily routine	2542 (81.61)	385 (12.36)	114 (3.66)	44 (1.41)	30 (0.96)
Activities	2117 (67.98)	455 (14.61)	193 (6.20)	163 (5.23)	186 (5.97)
Intervention					
Worried	1929 (64.62)	539 (18.06)	269 (9.01)	168 (5.63)	80 (2.68)
Sad	2162 (72.57)	402 (13.49)	200 (6.71)	129 (4.33)	86 (2.89)
Pain	2073 (69.56)	514 (17.25)	260 (8.72)	82 (2.75)	51 (1.71)
Tired	800 (26.87)	870 (29.22)	535 (17.97)	424 (14.24)	348 (11.69)
Annoyed	1960 (65.77)	522 (17.52)	229 (7.68)	141 (4.73)	128 (4.30)
Sleep	1660 (55.84)	721 (24.25)	324 (10.90)	183 (6.16)	85 (2.86)
School work	2029 (68.22)	645 (21.69)	211 (7.09)	67 (2.25)	22 (0.74)
Daily routine	2440 (82.13)	361 (12.15)	115 (3.87)	34 (1.14)	21 (0.71)
Activities	2062 (69.47)	428 (14.42)	186 (6.27)	140 (4.72)	152 (5.12)

1 indicates 'no problems', 5 indicates 'lots of problems'.

TABLE 44 Student HRQoL utility scores at 36 months

Measure sub-components	1, N (%)	2, N (%)	3, N (%)	4, N (%)	5, N (%)
Control					
Worried	1847 (62.40)	553 (18.68)	290 (9.80)	180 (6.08)	90 (3.04)
Sad	2101 (71.05)	401 (13.56)	213 (7.20)	143 (4.84)	99 (3.35)
Pain	2057 (69.59)	521 (17.63)	208 (7.04)	94 (3.18)	76 (2.57)
Tired	705 (23.87)	808 (27.36)	567 (19.20)	429 (14.53)	444 (15.04)
Annoyed	1956 (66.15)	505 (17.08)	225 (7.61)	135 (4.57)	136 (4.60)
Sleep	1553 (52.55)	735 (24.87)	355 (12.01)	171 (5.79)	141 (4.77)
School work	1846 (62.83)	657 (22.36)	280 (9.53)	85 (2.89)	70 (2.38)
Daily routine	2387 (81.19)	346 (11.77)	132 (4.49)	45 (1.53)	30 (1.02)
Activities	1881 (67.66)	390 (13.32)	182 (6.22)	150 (5.12)	225 (7.68)
Intervention					
Worried	1591 (58.15)	537 (19.63)	266 (9.72)	205 (7.49)	137 (5.01)
Sad	1929 (70.50)	368 (13.45)	203 (7.42)	129 (4.71)	107 (3.91)
Pain	1916 (70.11)	444 (16.25)	211 (7.72)	91 (3.33)	71 (2.60)
Tired	671 (24.52)	784 (28.64)	472 (17.25)	403 (14.72)	407 (14.87)
Annoyed	1796 (65.64)	487 (17.80)	189 (6.91)	123 (4.50)	141 (5.15)
Sleep	1518 (55.44)	653 (23.85)	323 (11.80)	146 (5.33)	98 (3.58)
School work	1756 (64.89)	603 (22.28)	244 (9.02)	68 (2.51)	35 (1.29)
Daily routine	2222 (82.27)	326 (12.07)	89 (3.30)	37 (1.37)	27 (1.00)
Activities	1919 (71.23)	345 (12.81)	159 (5.90)	127 (4.71)	144 (5.35)

1 indicates 'no problems', 5 indicates 'lots of problems'.

APPENDIX 1

TABLE 45 Proportion of missing values for all primary and secondary outcomes at baseline, 24 months and 36 months

Outcome	Baseline		24 months		36 months	
	Control n = 3347	Intervention n = 3320	Control n = 3195	Intervention n = 3095	Control n = 3087	Intervention n = 2873
GBS overall score	9.3%	9.4%	7.8%	8.0%	9.0%	8.5%
ESYTC overall score	5.6%	6.4%	5.4%	5.0%	5.1%	5.2%
PedsQL overall score	2.6%	5.1%	1.3%	2.6%	3.1%	4.1%
SDQ total difficulties score	2.7%	4.2%	2.0%	2.6%	2.7%	3.0%
SWEMWBS total well-being index	3.9%	6.8%	2.1%	3.3%	4.0%	4.9%
CHU9D overall score	7.1%	9.8%	4.2%	5.7%	7.1%	7.8%
Age of sexual debut	NA	NA	12.6%	10.5%	12.5%	13.7%
Ever smoked	2.2%	3.6%	3.6%	4.7%	4.4%	4.1%
Ever drunk alcohol	3.3%	4.3%	2.2%	3.1%	3.7%	3.3%
Been offered illicit drugs	4.0%	4.8%	2.7%	3.7%	3.8%	4.2%
Truancy	4.2%	4.5%	2.8%	3.5%	3.0%	3.0%
Exclusion from school	1.3%	2.3%	1.2%	1.6%	1.9%	2.4%
Used contraception at last sex	NA	NA	18.8%	14.3%	9.2%	9.7%
Use of NHS in past 12 months	1.8%	2.9%	1.5%	2.0%	2.3%	2.6%
Contact with police	1.0%	1.1%	0.9%	1.4%	1.7%	2.2%

Reproduced from Bonell *et al.*¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

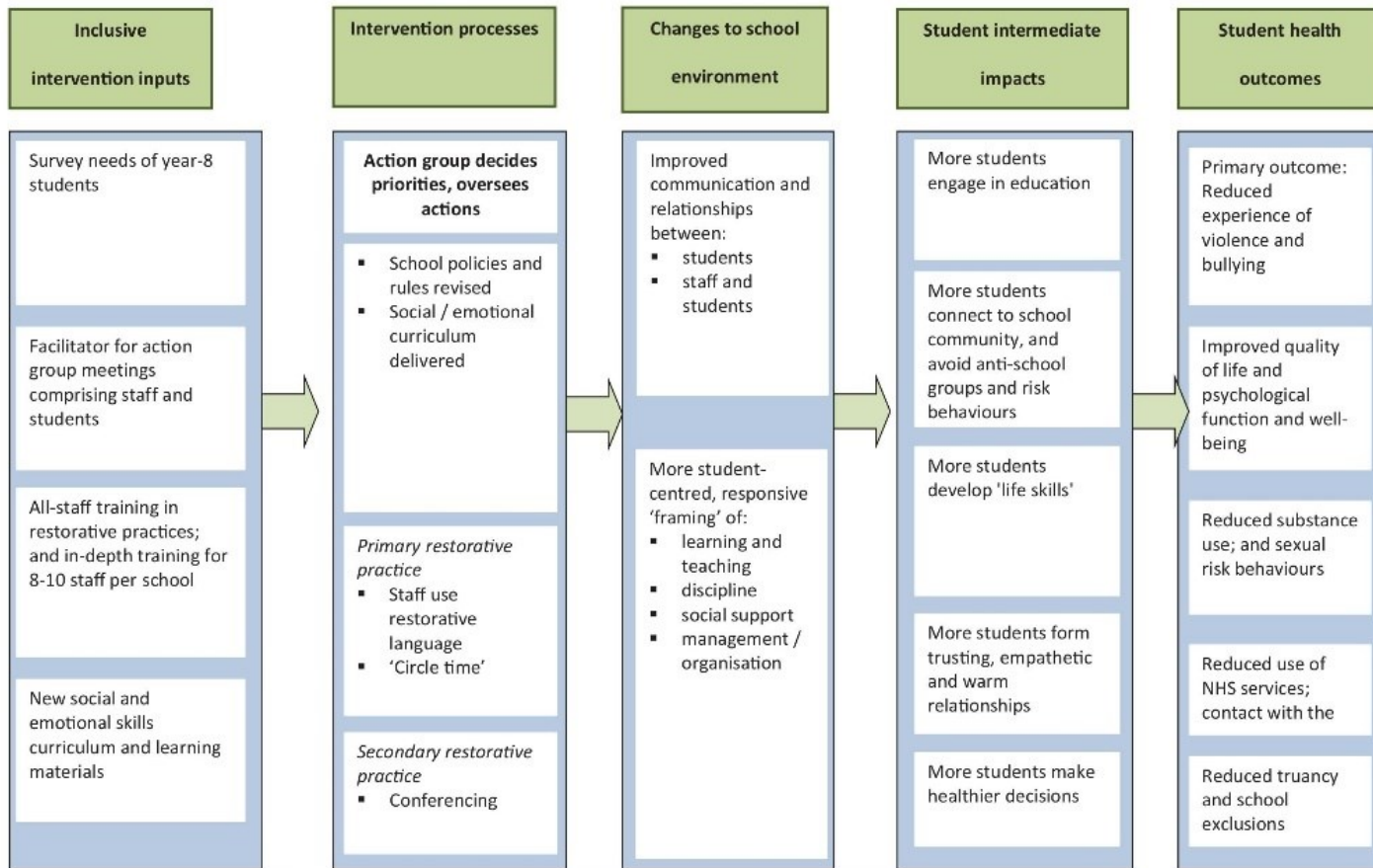


FIGURE 6 The INCLUSIVE logic model. Reproduced from Bonell et al.¹²² © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Appendix 2 The young researchers group feedback example

School Bullying Research by UCL

Introduction

Following UCL research presentation on school bullying intervention programme, ICYP (children and young people) Engagement Committee were tasked with exploring 4 questions through group discussion. These were

1. Do you think the intervention was effective?
2. Why do you think it had the results that it did?
3. What do the findings mean for schools?
 - a. For students
 - b. For teachers/administration team
4. What methods should we use to present these findings to the public?

Children, Young People and Worker Participation

Young people and parent/carers / youth workers / paediatricians took part in the session.

Young people were aged 11 – 24, mixed gender, including those with bullying experience, health care experience and from a range of ethnic backgrounds including black British, white British, Asian. Young people attending were from RCPCH &Us (ICYP Engagement Committee Members), a children's hospital youth forum and an African community group.

Adults are members of the ICYP Engagement Committee, representing parent/carers, paediatricians and a youth worker.

Consultation Responses

Group one (YP x 8)

1. You can't tell whether it was the study that was effective or whether the schools just got better over time
You need to understand impact across all years
2. No comment
3. No comment
4. Celebrity endorsement (someone who has lived experience)
Billboard advert
Teachers being aware of results
Teachers presenting to class / school in an engaging way

Social media – Instagram adverts, snapchat discover page, YouTube advert before video

Engaging (visual)

TES (times educational sector)

Student involvement in sharing results

Group two (YP x 8)

Comments

- Intervention doesn't always work
 - Punishment is sometimes needed and if the intervention didn't work
 - Punishment like detention / missing free time / ongoing / coming in on a Saturday, sitting outside the head, sit in the staff room in lunch
 - When told off seen as cool or people laugh
 - Health conditions can be targets for bullies
 - Part of the restorative meeting needs to involve adequate education around that condition. How will this individual situation be catered for?
1. Want to see numbers to see if it is a change in policy
Intervention / punishment - results didn't seem clear cut which was best
Punishment is quicker but doing the intervention takes time & skills
Focused on physical/emotional bullying not cyber bullying (could this be something to investigate). What about LGBTQ+
 2. Children had an understanding of what "bullying" means which is good!
Was it whole school training?
Might be good for students in the long term? Increases in quality of life?
 3. a) how do we know young people are telling the truth about illicit activities?
b) maybe need to re-think reflex jerk action if punishment
How much will this cost (may deter schools)? What training is required?
 4. More detail on what was "taught" what the intervention was – time frame, info over children involved, what questionnaire was used

Group three (parent/carer x 2, paediatricians x 3, youth worker x 1)

1. As the questionnaire evaluated?
Was it intervention or cultural change (tracking or bullying)
Was it effective equally across all schools?
Was baseline question same as final

2. Bringing bullying as an agenda? Cultural change not agenda
Any intervention involving students empowers them
Self-selecting
Real life effects
3. Need students to “buy-in” / willingness
Teaching how to demonstrate emotion
Recognise reason behind bullying
Do students want this?
Motivations for school to participate
How is group bullying managed?
4. needs more tangible. Not numbers not stats v all
Case studies / stories
Rebrand “restorative justice” (has a criminality) formal, not child focused
Peer = repairing relationships / friends
(2 tier; 1= friends. 2=relationships)

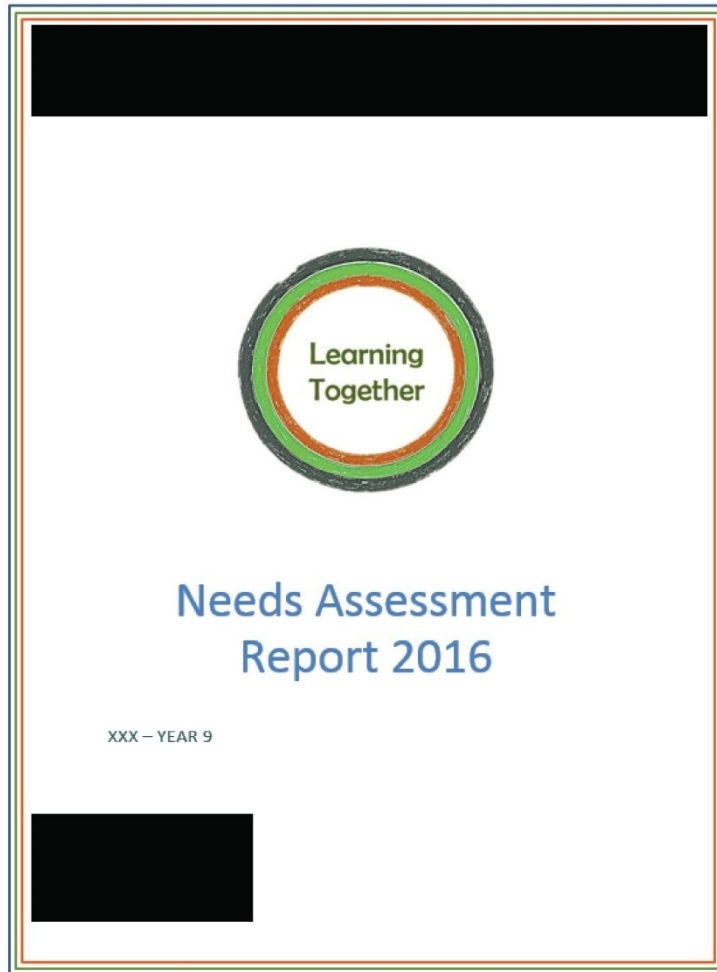
Further analysis of the views should be undertaken by the project lead and the project board. Data shared with UCL Research team. Additional comment may have been observed by LB the UCL researcher in attendance at the session. Raw data flipcharts available on request from RCPCH &Us.

Further exploration of themes or extension of topics covered can be requested through the children and young people’s engagement team for inclusion in future Roadshows and surveys.

Contact details



Appendix 3 Example school report, year 3 (anonymised)



The average of schools delivering the intervention in this trial, not the national average. By definition, for each score, half of the schools will be above the average and half will be below it. So most schools will score worse than average on some measures. It is also worth bearing in mind that the schools participating in this project are almost certainly better schools than the national average because they had the capacity and initiative to want to participate, so the average that we are working with will be higher than normal.

As well as looking at whether schools rank above or below the average for particular questions, schools should also look at the absolute frequency of the answers. For example, even if a school has a lower rate of cyber-bullying than average they may still decide to prioritise this issue if they judge the absolute level requires action. These judgements are inevitably subjective and up to schools to decide with the support of their external facilitator.

The scores are meant to be a guide to what schools might want to focus on in their work. They are not a judgement on the school. Schools may be challenged by these scores. We hope that schools reflect on the results and make good decisions based on them. In some cases schools may exercise the “precautionary principle” and act on data even where they do not chime with their experiences of the school.

Your report focuses on results for students as they came to the end of Year 9. It does not compare the results with data from last year, when the students were nearing the end of Year 7. This is because it would be impossible to unpick whether a change from the previous year reflected the impact of Learning Together or the natural effects of the students getting older. For example, we might expect to see an increase in some behaviours, such as smoking, as students get older. Such rising trends shouldn’t be treated as evidence that Learning Together is not working or that more generally things in the school are getting worse. So we think it is more important to judge the needs based on the figures this year and how they compare with the other schools.

Summary for XXX School

The responses of XXX School Year 9 students to the LEARNING TOGETHER survey are summarised in this report. It is bespoke feedback for XXX School and provides a comparison with Year 9 students in all other schools delivering the Learning Together programme. We will not release this report to any other party. However, you are welcome to share it with others if you wish. All feedback is summarised at the school level to protect individual students' privacy.

Overview of areas examined

In this report, several areas of school environment are examined:

1. School connectedness

This section explores students' sense of being engaged with and connected to the school. More specifically, we asked questions about:

- Whether students feel different from their peers in the school, whether students can be themselves at the school and whether students feel they belong to the school

2. Safety at School

This section includes statements about safety within the school environment. It also includes questions on bullying, students' misbehaviour and questions about the presence of clear and fair rules within the school. More specifically, we asked questions about:

a) Feeling of safety

- Whether students feel safe at school

b) Aggression towards students

- Whether students are threatened or physically hurt

c) Students' misbehaviour

- Whether students are cheeky to a teacher regularly, threaten a teacher regularly and whether they hit/kick a teacher regularly

d) Emotional bullying

- Whether students are left out of things at school, have rumours spread about them at school and whether they are teased

e) Rules/norms

- Whether their school have rules that are written down somewhere, whether teachers at school try to make sure that students obey rules and whether teachers at school are fair in dealing with students
- Whether they think that rules are fair at their school

3. Interpersonal relationships

This section explores how students perceive their relationships with other students and teaching staff. More specifically, we asked:

a) Relationship with other students

- Whether most other students accept them as they are, whether most students in their class are kind/helpful, whether most students in their class enjoy being together, whether they are encouraged to express their views in their class and whether other students in their school take their opinions seriously
- Whether they have argued with friends at school recently, whether they have a friend at school they can tell when they are angry/upset, whether they have a school friend they can share their happiness with, whether they have a school friend they can trust and whether they have a school friend who understand how they feel

b) Relationship with teachers

- Whether they like most of the teachers and whether they feel they can approach teachers with things that are on their mind

4. Strengths and difficulties

This section looks at students' capacity for relationships and practical reasoning. It also includes questions about the importance of education for students and how they perceive the future. It included:

a) Students' capacity for relationships

- Whether students usually share with others, whether they are helpful if someone is hurt, upset or feeling ill, whether they are kind to younger children, whether they regularly volunteer to help others, whether they fight a lot and can make people do what they want
- Whether there is at least one teacher or adult at school they can talk to about a problem

b) *Students' capacity for practical reasoning*

- Whether they are able to make their own mind about things, whether they think clearly and whether they deal with problems well
- Whether they get very angry and often lose their temper

c) *Future aspirations*

- Whether they try hard in school, whether doing well in school is important to them, whether continuing or completing their education is important to them and whether they feel they are successful in school
- Whether they are optimistic about the future

5. Teaching and Learning

This section looks at students' perception of teachers' attitudes. More specifically, questions concerned:

- Whether teachers at this school believe all students can learn, whether students' ideas are listened and valued, whether their school really care about them as individuals and whether most teachers listen to what they have to say

Overview

This report presents information derived from surveys completed by Year 9 students between March and July 2016. XXX School was randomly chosen along with 19 other secondary schools to act as the intervention group which includes half the sample.

In XXX School, 180 Year 9 students completed the LEARNING TOGETHER survey. Of the students, 49% were male and 51% were female.

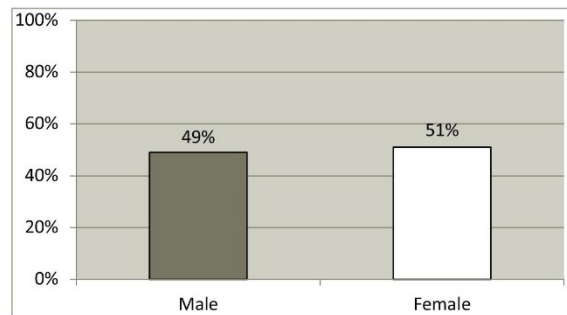


Figure 1: The percentage of male and female Year 9 pupils surveyed at XXX School

The Analysis

For each measure we compared XXX School with the other 19 LEARNING TOGETHER intervention schools' average. Where small numbers of students are involved caution must be exercised when interpreting the data. The green squares indicate areas where your school is doing well and red squares areas where more work/improvement is required. Black squares indicate scores that are similar to the other LEARNING TOGETHER intervention schools. Pie charts show values within your school. However, when these numbers are black that means that your school is similar to the other LEARNING TOGETHER intervention schools. When they are green/red a difference exists between your school and the other LEARNING TOGETHER intervention schools.

Interpretation

The LEARNING TOGETHER data in this feedback report informs XXX School about their Year 9 students compared to other participating Year 9 students across Greater London. This report does not provide information on why a difference might exist between the number of students in XXX School and those of the LEARNING TOGETHER intervention schools average. This feedback needs to be interpreted in the context of other information about the students involved.

1. School connectedness

Male students:

- 39.1% of Year 9 male students agreed with the statement that they feel different from most other students at XXX School
- 72.4% of Year 9 male students agreed with the statement that they can really be themselves at this school
- 85.1% of Year 9 male students agreed with the statement that they feel they belong at this school

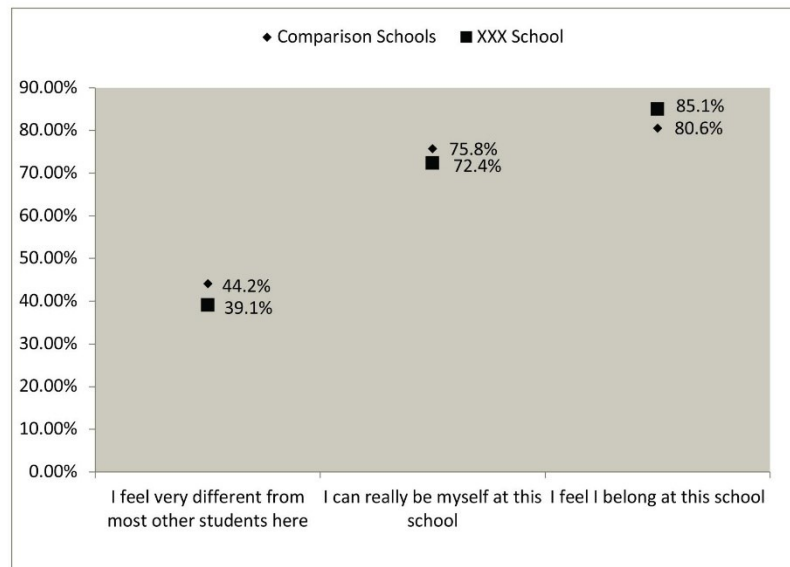


Figure 2a: Response percentages of Year 9 male students to school connectedness items.

Female students:

- 46.7% of Year 9 female students agreed with the statement that they feel different from most other students
- 65.2% of Year 9 female students agreed with the statement that they can really be themselves at this school
- 80.4% of Year 9 female students agreed with the statement that they feel they belong at this school

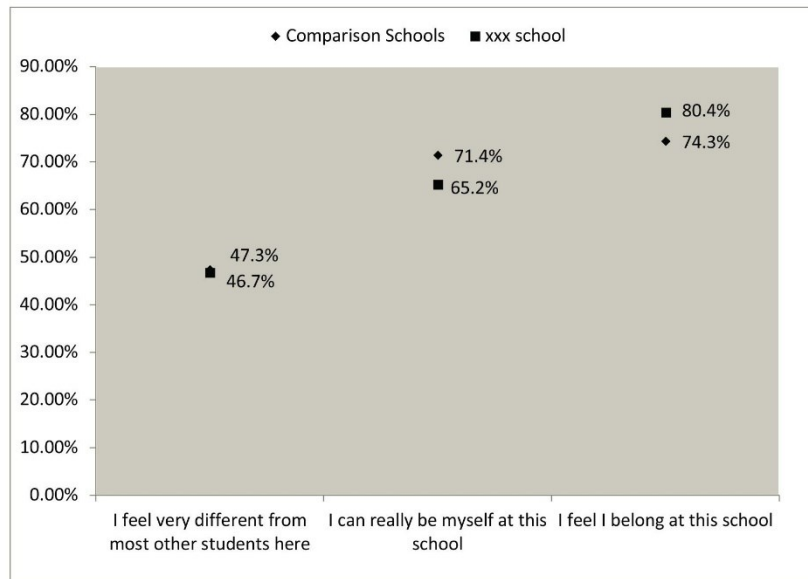


Figure 2b: Response percentages of Year 9 female students to school connectedness items.

2. Safety at School

a) Feeling of safety

In XXX School 11.9% of Year 9 male students responded “no/somewhat” to the following question: “Do you feel safe at this school?”

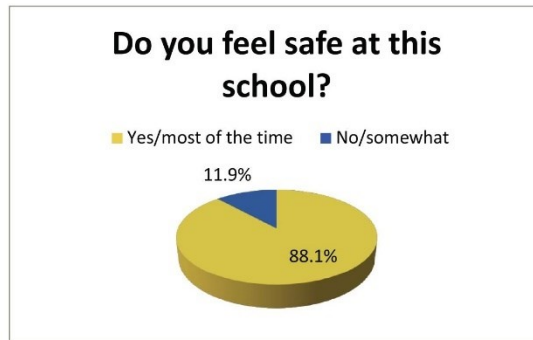


Figure 3a: Percentage of Year 9 male students who said they feel safe/unsafe at XXX School

In XXX School 9.8% of Year 9 female students responded “no/somewhat” to the following question: “Do you feel safe at this school?”

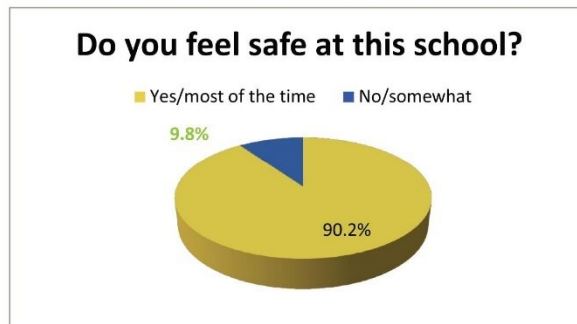


Figure 3b: Percentage of Year 9 female students who said they feel safe/unsafe at XXX School

b) Aggression towards students

Male students:

- 23% of Year 9 male students responded “Yes” to the question whether they have been victim of bullying

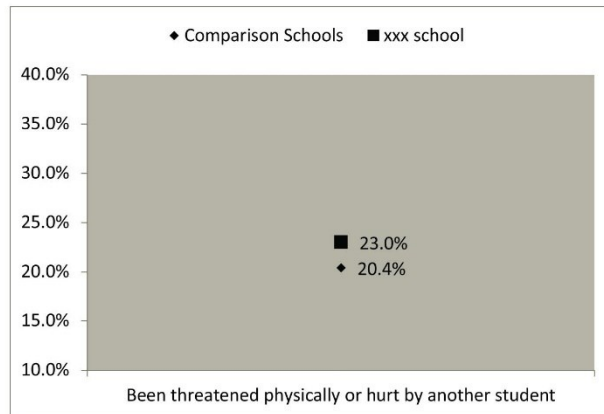


Figure 4a: Percentage of Year 9 male students who reported being threatened or physically hurt in the three months prior to the survey.

Female students:

- 2.2% of Year 9 female students responded “Yes” to the question whether they have been victim of bullying

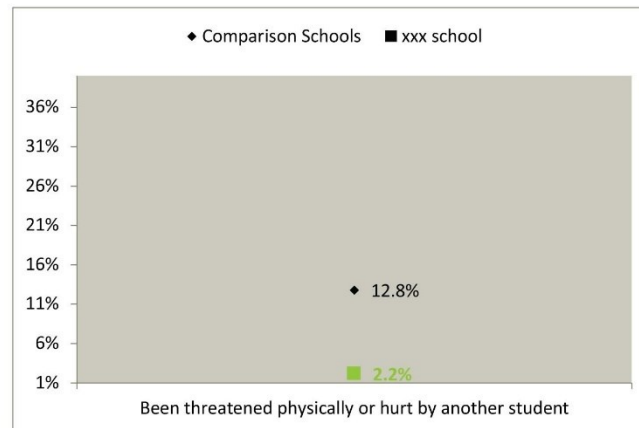


Figure 4b: Percentage of Year 9 female students who reported being threatened or physically hurt in the three months prior to the survey.

c) Students' misbehaviour

Male students:

- 33.7% of Year 9 male students agreed with the statement that they have been cheeky to a teacher regularly
- 1.2% of Year 9 male students agreed with the statement that they have threatened a teacher regularly
- 1.2% of Year 9 male students agreed with the statement that they have hit/kicked a teacher regularly

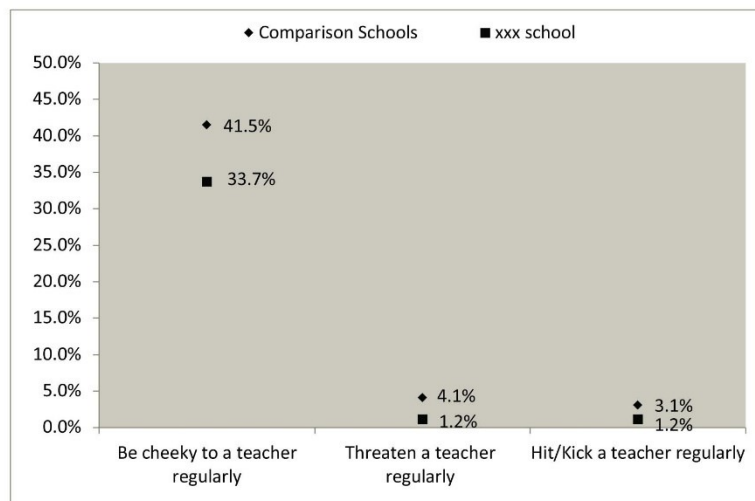


Figure 5a: Percentage of Year 9 male students who have been cheeky, have threatened or have been physically abusive to teaching staff in the three months prior to the survey.

Female students:

- 30.4% of Year 9 female students agreed with the statement that they have been cheeky to a teacher regularly
- 0% of Year 9 female students agreed with the statement that they have threatened a teacher regularly
- 1.1% of Year 9 female students agreed with the statement that they have hit/kicked a teacher regularly

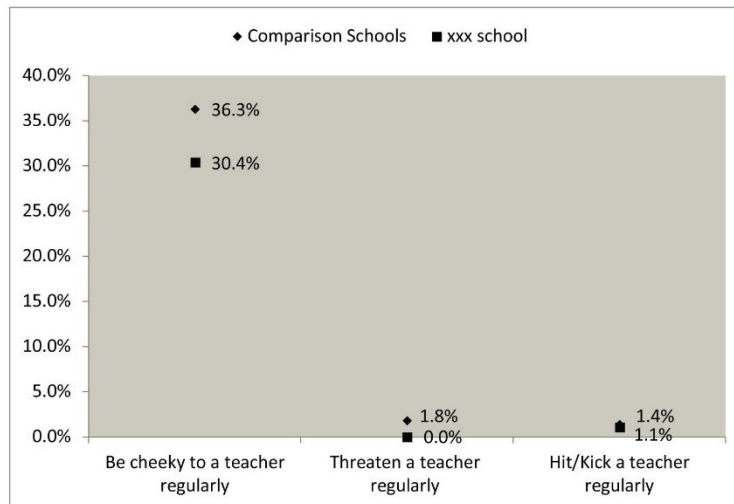


Figure 5b: Percentage of Year 9 female students who have been cheeky, have threatened or have been physically abusive to teaching staff in the three months prior to the survey.

d) Emotional Bullying

Male students:

- 20.7% of Year 9 male students agreed with the statement that they have been deliberately left out of things at school
- 29.4% of Year 9 male students agreed with the statement that they have had rumours spread about them at school
- 47.7% of Year 9 male students agreed with the statement that they have been teased/called names at school

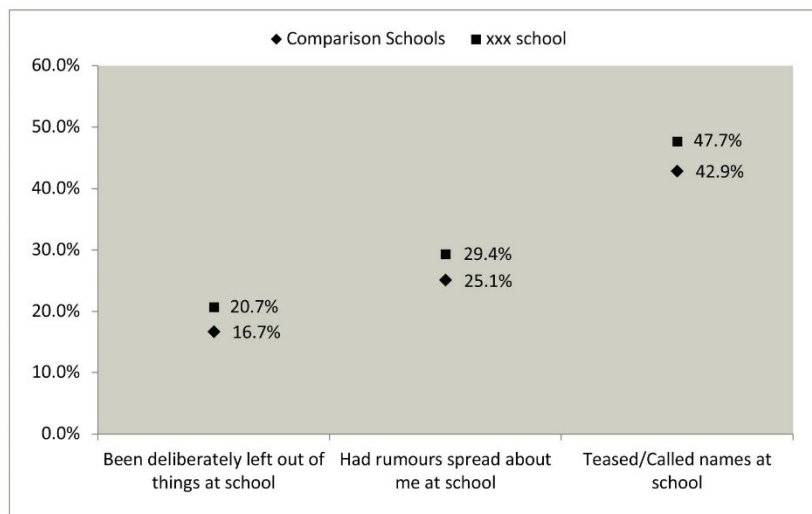


Figure 6a: Percentage of Year 9 male students who said that they had been left out of things, had rumours spread about them, or had been target of verbal abuse in the three months prior to the survey.

Female students:

- 28.3% of Year 9 female students agreed with the statement that they have deliberately left out of things at school
- 34.8% of Year 9 female students agreed with the statement that they have had rumours spread about them at school
- 38% of Year 9 female students agreed with the statement that they have been teased/called names at school

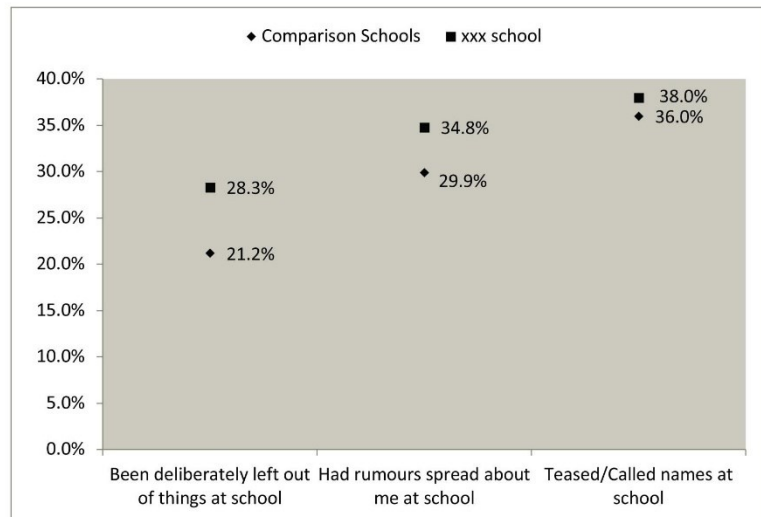


Figure 6b: Percentage of Year 9 female students who said that they had been left out of things, had rumours spread about them, or had been target of verbal abuse in the three months prior to the survey.

e) Rules/Norms

Male students:

- 51.8% of Year 9 male students at XXX School answered “no” or ”do not know” to the question “Does this school have rules written down somewhere?”
- 22.4% of Year 9 male students at XXX School answered “none” or ”some” to the question “Do teachers at this school try to make sure that students obey the rules?”
- 29.5% of Year 9 male students at XXX School disagreed with the statement “The teachers at this school are fair in dealing with students”

	XXX School n (%)	Comparison Schools n (%)
<i>Does this school have rules that are written down somewhere? (no/don't know)</i>	44 (51.8)	595 (50)
<i>Do teachers at this school try to make sure that students obey the rules? (none/some)</i>	19 (22.4)	390 (32.9)
<i>The teachers at this school are fair in dealing with students (disagree)</i>	26 (29.5)	452 (36)

Table 1a: Proportion of Year 9 male students who did not feel that their teachers were fair in dealing with students and ensuring that students obeyed the rules.

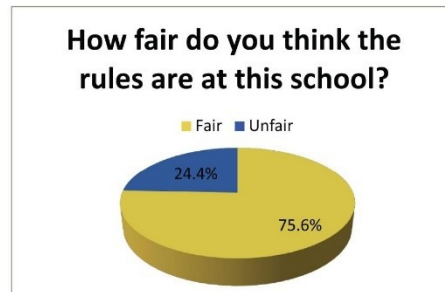


Figure 7a: Percentage of Year 9 XXX School male students who thought that the rules in their school are fair/unfair

Female students:

- 58.7% of Year 9 female students at XXX School answered “no”/“do not know” to the question “Does this school have rules written down somewhere?”
- 31.5% of Year 9 female students at XXX School answered “none”/“some” to the question “Do teachers at this school try to make sure that students obey the rules?”
- 18.5% of Year 9 female students at XXX School disagreed with the statement “The teachers at this school are fair in dealing with students”

	XXX School <i>n</i> (%)	Comparison Schools <i>n</i> (%)
<i>Does this school have rules that are written down somewhere (no/don't know)</i>	54 (58.7)	788 (49.7)
<i>Do teachers at this school try to make sure that students obey the rules (none/some)</i>	29 (31.5)	607 (38.2)
<i>The teachers at this school are fair in dealing with students (disagree)</i>	17 (18.5)	665 (41)

Table 1b: Proportion of Year 9 female students who did not feel that their teachers were fair in dealing with students and ensuring that students obeyed the rules.



Figure 7b: Percentage of XXX School Year 9 female students who thought that the rules in their school are fair/unfair

3. Interpersonal relationships

a) Relationships with other students

Male students:

- 19.5% of Year 9 male students agreed with the statement that most other students do not accept them as they are
- 19.5% of Year 9 male students agreed with the statement that most students in their classes are not kind/helpful
- 10.5% of Year 9 male students agreed with the statement that most students in their classes do not enjoy being together
- 18.6% of Year 9 male students agreed with the statement that they are not encouraged to express their own views in the their classes
- 24.4% of Year 9 male students agreed with the statement that other students in the school do not take their opinions seriously

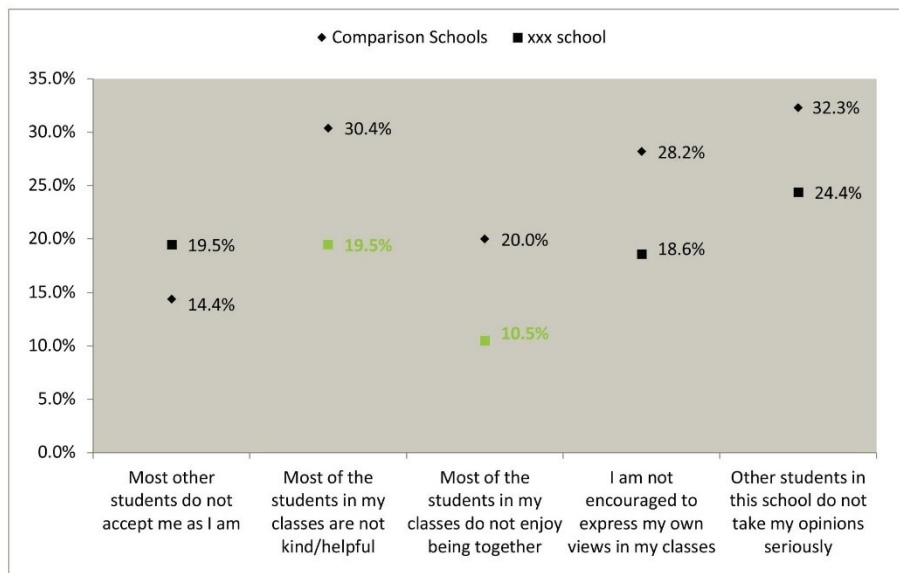


Figure 8a: Response percentages of Year 9 male students to interpersonal relationships items.

Female students:

- 27.2% of Year 9 female students agreed with the statement that most other students do not accept them as they were
- 22.8% of Year 9 female students agreed with the statement that most students in their classes are not kind/helpful
- 10.9% of Year 9 female students agreed with the statement that most students in their classes do not enjoy being together
- 20.7% of Year 9 female students agreed with the statement that they are not encouraged to express their own views in their classes
- 35.9% of Year 9 female students agreed with the statement that other students in the school do not take their opinions seriously

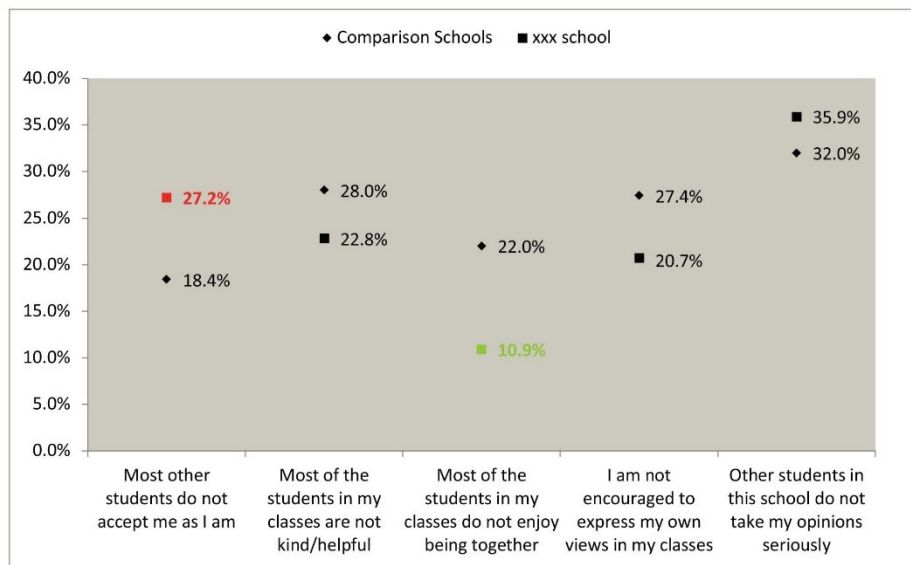


Figure 8b: Response percentages of Year 9 female students to interpersonal relationships items.

Male students:

- 25.3% of Year 9 male students agreed with the statement that they argued with friends at school recently
- 20.2% of Year 9 male students agreed with the statement that when they are angry/upset they do not have a friend at school they can tell
- 3.6% of Year 9 male students agreed with the statement that they do not have a friend at school they can share their happiness with
- 19% of Year 9 male students agreed with the statement that they do not have a school friend they can trust
- 13.1% of Year 9 male students agreed with the statement that they have a school friend who understand how they feel

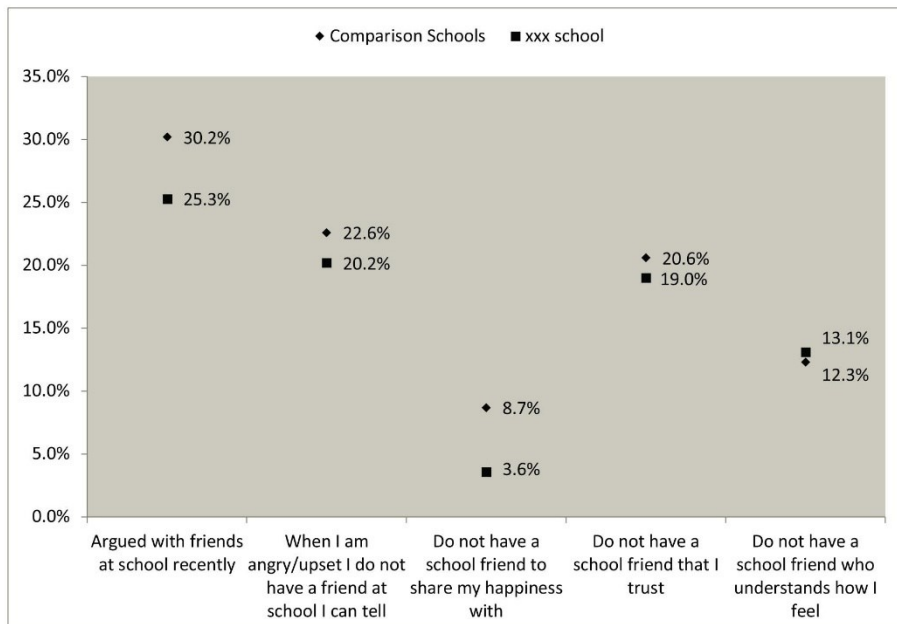


Figure 9a: Response percentage of Year 9 male students to interpersonal support/conflict items

Female students:

- 43.5% of Year 9 female students agreed with the statement that they argued with friends at school recently
- 12.1% of Year 9 female students agreed with the statement that when they are angry/upset they do not have a friend at school they can tell
- 2.2% of Year 9 female students agreed with the statement that they do not have a friend at school they can share their happiness with
- 18.5% of Year 9 female students agreed with the statement that they do not have a school friend they can trust
- 12.1% of Year 9 female students agreed with the statement that they have a school friend who understand how they feel

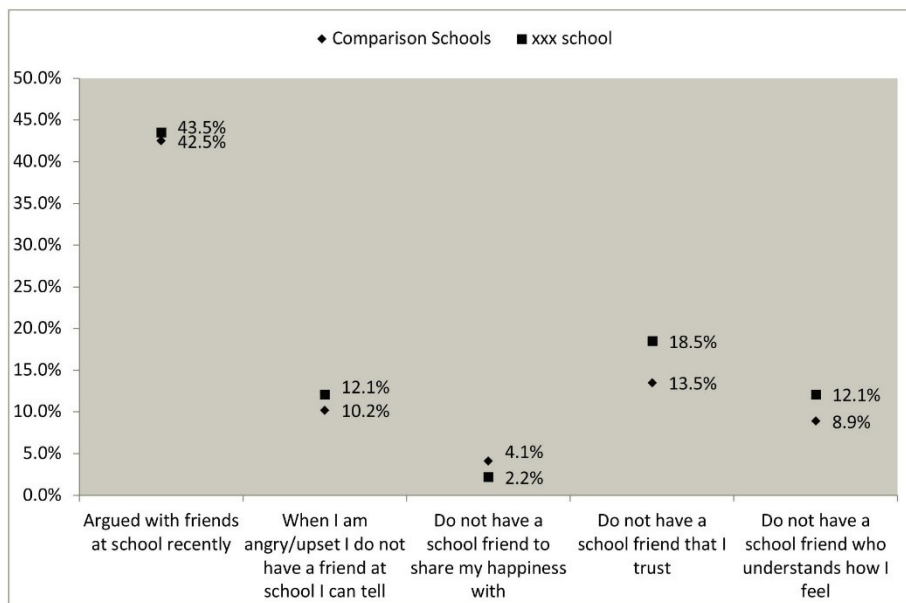


Figure 9b: Response percentage of Year 9 female students to interpersonal support/conflict items

b) Relationship with teachers

Male students:

- 27.6% of Year 9 male students agreed with the statement that they do not like most of their teachers
- 54.5% of Year 9 male students agreed with the statement that they feel they cannot not approach teachers with things that are on their minds

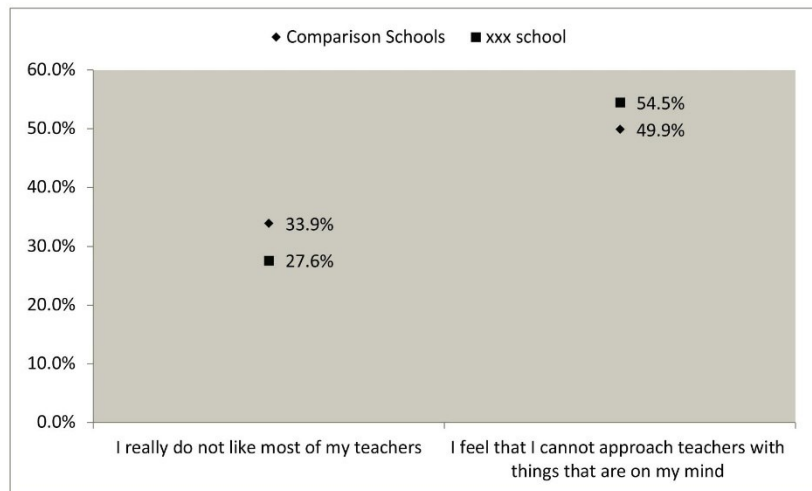


Figure 10a: Percentage of Year 9 male students who reported that they did not like most of their teachers, or could not approach them with things that are on their mind.

Female students:

- 30.4% of Year 9 female students agreed with the statement that they do not like most of their teachers
- 45.7% of Year 9 female students agreed with the statement that they feel they cannot approach teachers with things that are on their minds

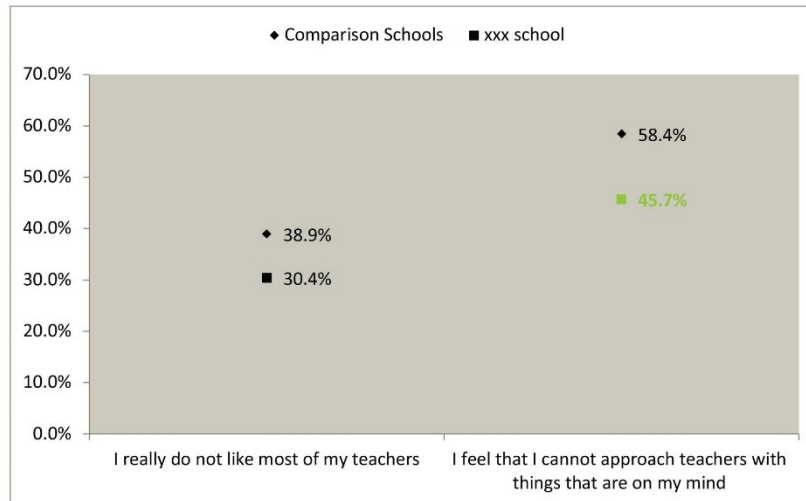


Figure 10b: Percentage of Year 9 female students who reported that they did not like most of their teachers, or could not approach them with things that are on their mind.

4. Strengths and difficulties

a) Students' capacity for relationships

Male students:

- 8% of Year 9 male students agreed with the statement that they do not usually share with others
- 2.3% of Year 9 male students agreed with the statement that they are not helpful if someone is hurt, upset or feeling ill
- 0% of Year 9 male students agreed with the statement that they are not kind to younger children
- 20.7% of Year 9 male students agreed with the statement that they do not regularly volunteer to help others
- 8.1% of Year 9 male students agreed with the statement that they fight a lot and that they can make people do what they want

	XXX School		Comparison Schools	
	<i>n</i>	(%)	<i>n</i>	(%)
<i>I don't usually share with others</i>	7	(8)	112	(9.2)
<i>I am not helpful if someone is hurt, upset or feeling ill</i>	2 (2.3)		89	(7.3)
<i>I am not kind to younger children</i>	0	(0)	74	(6.1)
<i>I do not regularly volunteer to help others (parents, children, teachers)</i>	18	(20.7)	290	(24)
<i>I fight a lot. I can make other people do what I want</i>	7	(8.1)	184	(15.2)

Table 2a: The proportions of XXX School Year 9 male students versus Year 9 male students in other intervention schools who said that they did not share with others, were not kind to younger students, did not regularly volunteer and who fought a lot.

Female students:

- 4.3% of Year 9 female students agreed with the statement that they do not usually share with others
- 1.1% of Year 9 female students agreed with the statement that they are not helpful if someone is hurt, upset or feeling ill
- 1.1% of Year 9 female students agreed with the statement that they are not kind to younger children
- 13% of Year 9 female students agreed with the statement that they do not regularly volunteer to help others
- 7.6% of Year 9 female students agreed with the statement that they fight a lot and that they can make people do what they want

	XXX School		Comparison Schools	
	<i>n</i>	(%)	<i>n</i>	(%)
<i>I don't usually share with others</i>	4	(4.3)	115	(7.1)
<i>I am not helpful if someone is hurt, upset or feeling ill</i>	1	(1.1)	53	(3.3)
<i>I am not kind to younger children</i>	1	(1.1)	42	(2.6)
<i>I do not regularly volunteer to help others (parents, children, teachers)</i>	12	(13)	252	(15.8)
<i>I fight a lot. I can make other people do what I want</i>	7	(7.6)	212	(13.2)

Table 2b: The proportions of XXX School Year 9 female students versus Year 9 female students in other intervention schools who said that they did not share with others, were not kind to younger students, did not regularly volunteer and who fought a lot.

In XXX School 25% of Year 9 male students disagreed with the following statement “There is at least one teacher or adult at this school I can talk to about a problem”



Figure 11a: Percentage of Year 9 male students who said that there was an adult or teacher at the school they could talk to about a problem.

In XXX School 18.5% of Year 9 female students disagreed with the following statement “There is at least one teacher or adult at this school I can talk to about a problem”

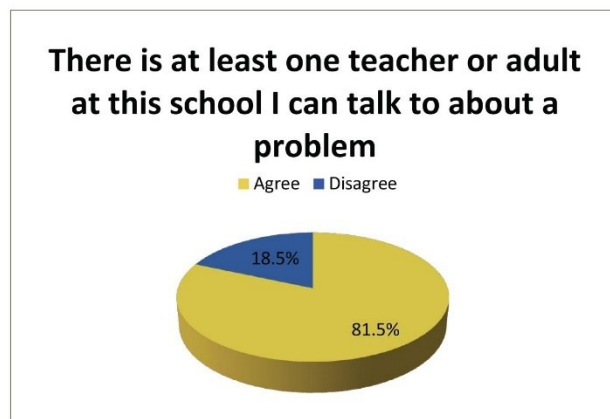


Figure 11b: Percentage of Year 9 female students who said that there was an adult or teacher at the school they could talk to about a problem.

b) Student's capacity for practical reasoning

Male students:

- 96.5% of Year 9 male students agreed with the statement that they are able to make up their minds about things
- 86% of Year 9 male students agreed with the statement that they have been thinking clearly
- 81.2% of Year 9 male students agreed with the statement that they have been dealing with problems well

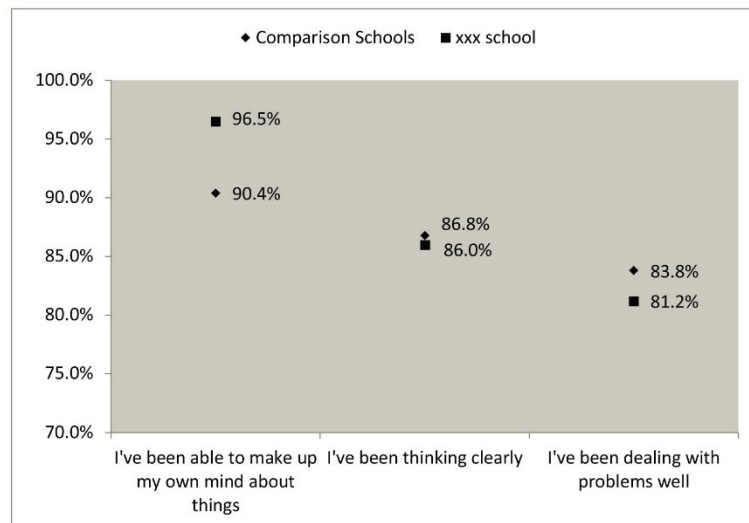


Figure 12a: Percentage of Year 9 male students who reported that they dealt with problems well and were able to make up their own mind about things.

Female students:

- 87% of Year 9 female students agreed with the statement that they are able to make up their minds about things
- 82.6% of Year 9 female students agreed with the statement that they have been thinking clearly
- 83.7% of Year 9 female students agreed with the statement that they have been dealing with problems well

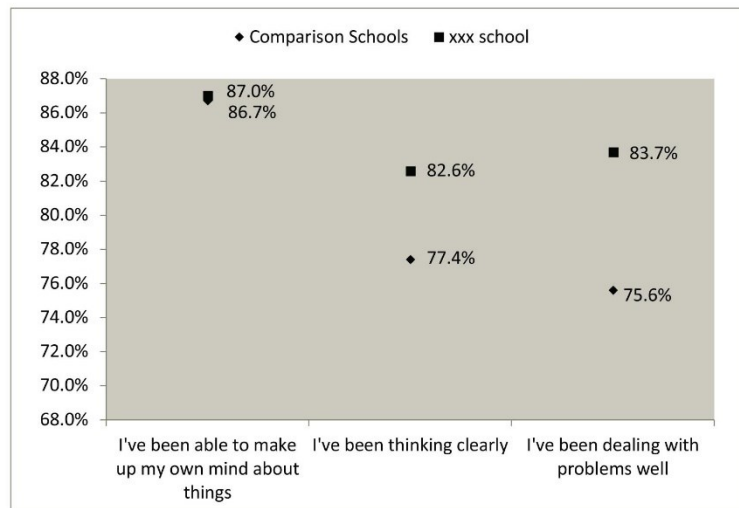


Figure 12b: Percentage of Year 9 female students who reported that they dealt with problems well and were able to make up their own mind about things.

In XXX School 44.4% of Year 9 male students disagreed with the following statement “I get very angry and often lose my temper”



Figure 13a: Percentage of Year 9 male students who agreed/disagreed with the statement “I get very angry and often lose my temper”

In XXX School 59.8% of Year 9 female students disagreed with the following statement “I get very angry and often lose my temper”



Figure 13b: Percentage of Year 9 female students who agreed/disagreed with the statement “I get very angry and often lose my temper”

c) Future Aspirations

Male students:

- 1.1% of Year 9 male students agreed with the statement that they do not try hard in school
- 1.1% of Year 9 male students agreed with the statement that doing well in school is not very important for them
- 4.6% of Year 9 male students agreed with the statement that continuing or completing education is not important to them
- 16.3% of Year 9 male students agreed with the statement that they do not feel like they are successful in this school

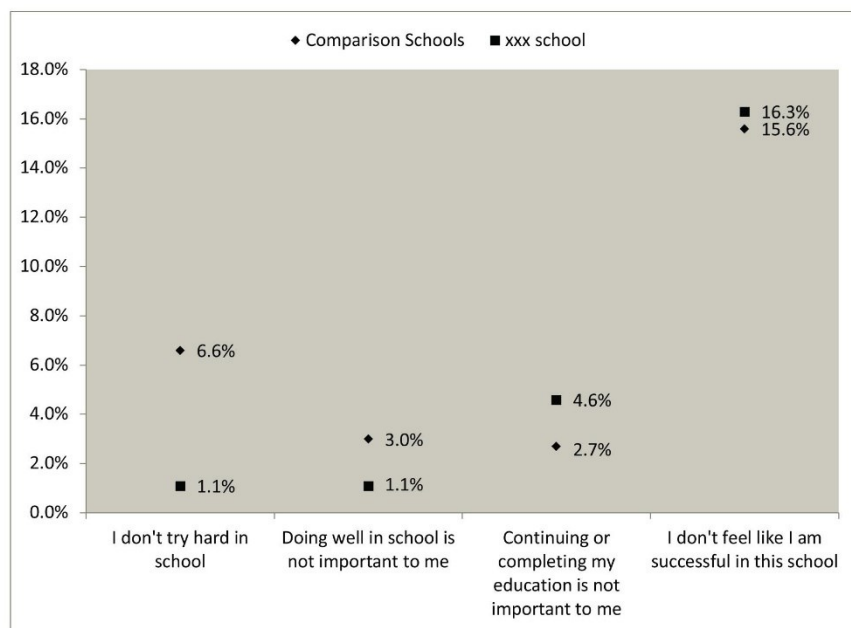


Figure 14a: Percentage of Year 9 male students who do not feel successful at school or do not feel that school is important

Female students:

- 1.1% of Year 9 female students agreed with the statement that they do not try hard in school
- 1.1% of Year 9 female students agreed with the statement that doing well in school is not very important for them
- 0% of Year 9 female students agreed with the statement that continuing or completing education is not important to them
- 8.7% of Year 9 female students agreed with the statement that they do not feel like they are successful in this school

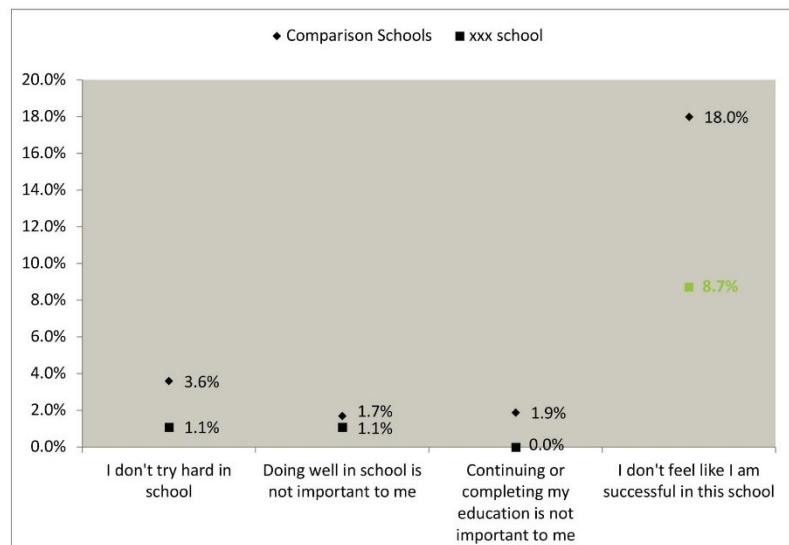


Figure 14b: Percentage of Year 9 female students who do not feel successful at school or do not feel that school is important

When asked if they have been feeling optimistic about the future, 18.8% of Year 9 male students at XXX School reported that they rarely/never felt optimistic about their future

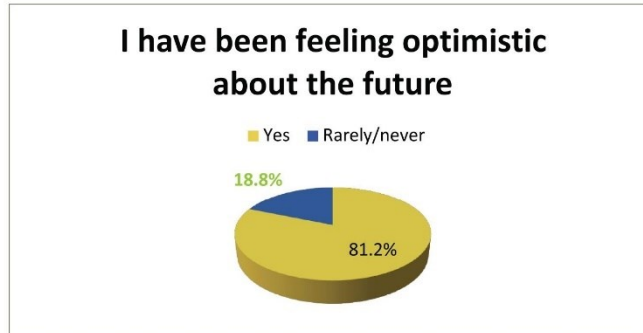


Figure 15a: Percentage of Year 9 male students who do not feel optimistic about their future.

When asked if they had been feeling optimistic about the future, 20.9% of Year 9 female students at XXX School reported that they rarely/never felt optimistic about their future

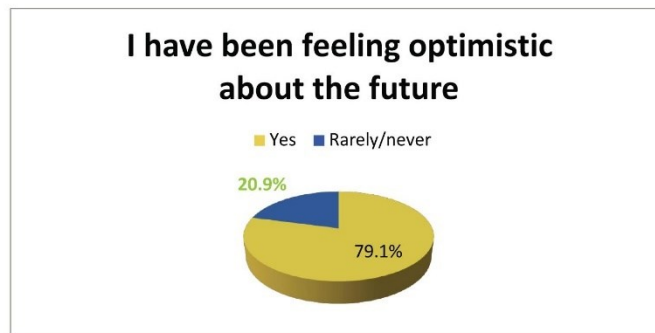


Figure 15b: Percentage of Year 9 female students who do not feel optimistic about their future.

5. Teaching and Learning

Male students:

- 14.8% of Year 9 male students agreed with the statement that teachers at this school do not believe all students can learn
- 19.5% of Year 9 male students agreed with the statement that students' ideas are not listened to or valued
- 20.7% of Year 9 male students agreed with the statement that this school does not really care about students as individuals
- 22.7% of Year 9 male students agreed with the statement that most teachers do not listen to what they have to say

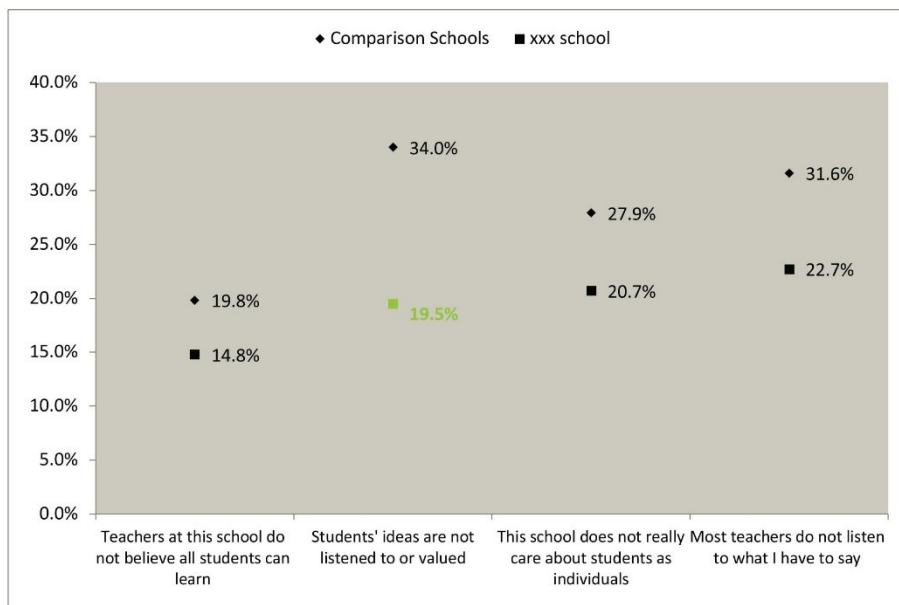


Figure 16a: Year 9 male students' responses to items regarding teacher support.

Female students:

- 18.5% of Year 9 female students agreed with the statement that teachers at this school do not believe all students can learn
- 31.5% of Year 9 female students agreed with the statement that students' ideas are not listened to or valued
- 28.6% of Year 9 female students agreed with the statement that this school does not really care about students as individuals
- 28.3% of Year 9 female students agreed with the statement that most teachers do not listen to what they have to say

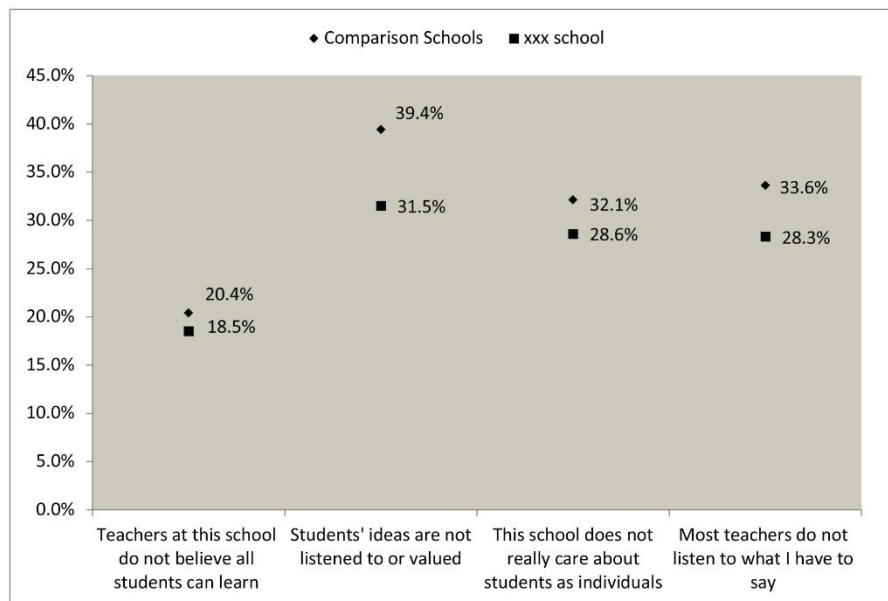


Figure 16b: Year 9 female student's responses to items regarding teacher support.

THANK YOU!

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 and Social Care*

Published by the NIHR Journals Library

Appendix 9: Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention

Research report



Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention

Chris Bonell,¹ Elizabeth Allen,² Charles Opondo,² Emily Warren,¹ Diana Ruth Elbourne,² Joanna Sturgess,² Leonardo Bevilacqua,³ Jennifer McGowan,³ Anne Mathiot,³ Russell M Viner³

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jech-2018-211443>).

¹Department of Public Health, Environments and Society, London School of Hygiene and Tropical Medicine, London, UK
²Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK
³Department of Social Science, UCL Institute of Child Health, London, UK

Correspondence to

Professor Chris Bonell, Department of Public Health, Environments and Society, London School of Hygiene and Tropical Medicine, London WC1H 9SH, UK; chris.bonell@lshtm.ac.uk

Received 8 August 2018
Revised 19 December 2018
Accepted 10 January 2019
Published Online First
5 February 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY. Published by BMJ.

To cite: Bonell C, Allen E, Opondo C, et al. *J Epidemiol Community Health* 2019;**73**:455–464.

ABSTRACT

Background Interventions to modify school environments are effective in promoting young people's health across outcomes, but mechanisms are poorly understood. We assessed mediation in a trial of the Learning Together intervention, building on the recent publication of results of effectiveness for reducing bullying and benefits across secondary outcomes and generally good implementation fidelity.

Methods Within a cluster-randomised trial involving 40 English schools, we examined student-reported and staff-reported school climate and student-reported involvement with delinquent peers at 24-month and 36-month follow-up, assessing the reliability of measures and whether these mediated health outcomes at a final follow-up.

Results Response rates and reliability were good for student-reported but not staff-reported measures. The intervention increased student-reported but not staff-reported-positive school climate but, like effects on student health outcomes, these manifested only at a final follow-up. The intervention reduced student-reported contact with delinquent peers at an interim follow-up. Student-reported potential mediators measured at the interim follow-up were associated with most health outcomes at the final follow-up. Adjustment for student-reported school climate and contact with delinquent peers at the interim follow-up did not reduce the associations between trial arm and our health outcomes.

Conclusion Despite being constrained by imperfect measures and by the late manifestation of impacts on student-reported school climate undermining ability to assess mediation, our study for the first time provides tentative evidence that mediation of intervention effects via improved climate and disengagement from delinquent peers is plausible. Our study provides the first evidence from a trial that whole-school interventions may work by modifying school environments and student relationships.

Trial registration number ISRCTN10751359.

INTRODUCTION

There is increasing interest in interventions aiming to promote young people's health by making overall school environments more health-promoting.¹ A Cochrane review of 'health promoting schools' interventions (with environment, community and curriculum components) reported various benefits,

including reducing bullying victimisation, smoking and body mass index, and increasing physical activity.² Another review focused on interventions with environmental and not curriculum components also reported multiple health benefits.³

But how do such interventions work? A systematic review of theories of how school environments influence health⁴ concluded that the theory of human functioning and school organisation is the most comprehensive theory of change for such interventions.⁵ This postulates that for young people to choose healthier over riskier behaviours, they must possess the autonomy and ability to reason and form relationships, to make informed, healthy decisions. These capacities are facilitated by student engagement with school: good relationships with teachers; commitment to learning; and sense of belonging and participation in the school community. A refinement of the theory suggests that students lacking such school commitments may engage with delinquent peers and risky behaviours as alternative markers of belonging and identity.⁶ The theory also suggests that schools can increase student commitment by modifying school organisation: distributing authority between staff; promoting good staff-student relationships; integrating academic education and broader student development; and ensuring school culture reflects that of the local community.

While there is some evidence in support of this theory from observational studies of school-level determinants of student health,⁷ such mechanisms remain largely unexamined in intervention studies. The only study that has examined whether the health effects of whole-school interventions are mediated by student commitments to school was the Gatehouse Project. This involved a randomised controlled trial (RCT) of a whole-school intervention delivered in Australian secondary schools. Despite reporting effects on various measures of adolescent health-related risk behaviours, the study found no evidence of effects on student attachment to school, suggesting that attachment was not a mediator of health effects or that the measure failed to assess attachment.⁸

We explore these questions in relation to our own recent RCT of the Learning Together intervention.⁹ This whole-school intervention aimed to support schools to implement the restorative practice, staff/student action groups, and a student

BMJ

Bonell C, et al. *J Epidemiol Community Health* 2019;**73**:455–464. doi:10.1136/jech-2018-211443

455

J Epidemiol Community Health: first published as 10.1136/jech-2018-211443 on 5 February 2019. Downloaded from <http://jech.bmj.com/> on January 18, 2022 at LSHTM Consortium. Protected by copyright.

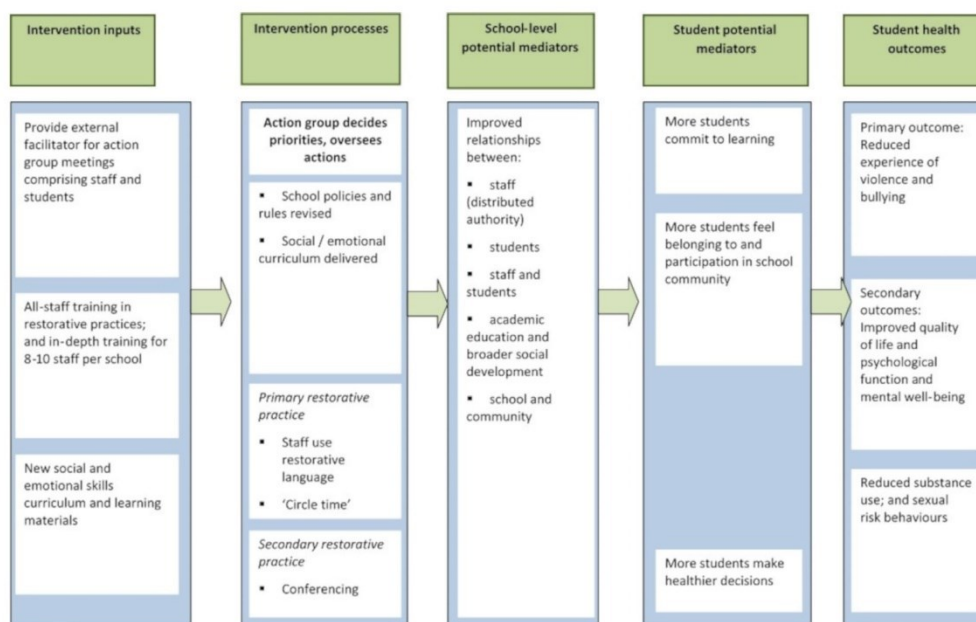


Figure 1 Logic model

social and emotional skills curriculum to reduce bullying and aggression, and promote student health across various secondary outcomes. The restorative practice aims to improve relationships to prevent and/or resolve conflicts between students or between staff and students.¹⁰ It aims to prevent incidents via methods such as 'circle-time' (bringing staff and students together to build relationships) and/or resolve incidents via methods such as 'conferencing' (bringing together conflicting parties to find ways to avoid further harms). Action groups are school meetings involving diverse students, and senior and junior staff. They coordinate whole-school intervention delivery and review school rules and policies to ensure that these support whole-school change. They aim to build better relationships between the staff and students sitting on the group and signal to the wider student body that the school cares about and intends to act on the views of staff and students to build a supportive school climate.³ Social and emotional education aims to ensure that schools teach not only academic knowledge but also attend to students' broader social development.¹¹ Informed by the theory of human functioning and school organisation,⁵ these intervention components were theorised to work synergistically within Learning Together to: distribute decision-making authority across the school; strengthen relationships between and among staff and students; and integrate students' academic learning and broader development. The intervention did not aim to improve relationships between schools and their local communities, although this might occur as a by-product. These impacts were theorised to transform the whole-school climate and improve staff-student relationships, student commitment to learning and sense of belonging and participation in the school community, thereby reducing student engagement with delinquent peers and risk behaviours, and promoting student health (figure 1).

An RCT of the intervention reported a range of benefits including reducing bullying victimisation, use of tobacco, alcohol and drugs, and contact with the police, and promoting mental well-being, quality of life and psychological functioning. The intervention was found to be implemented with good fidelity although this was much lower for the curriculum component.¹² This paper aims to examine potential mediators of such effects. We first assess the reliability of our potential mediators. These include existing student-reported measures of views of school climate and engagement with delinquent peers. We also examine a new measure of staff-reported school organisational climate to explore whether intervention effects on the school organisational environment might explain student health outcomes. Our second aim is to assess whether these measures might be mediators of intervention effects on our primary and secondary health outcomes.

METHODS

Here we provide a summary of methods for the trial. For full details, including sample size calculation, see published protocol and trial report.^{9 12} We undertook a two-arm parallel repeat cross-sectional cluster RCT of the Learning Together intervention in 40 secondary schools eligible to take part as state schools in south-eastern England with government inspections rated as requires improvement or above, recruited by the trial team via emails. Our eligible study population consisted of all students: at baseline at the end of year 7 in 2014 (11–12 years); at interim 24-month follow-up; and final 36-month follow-up in 2017. Student data were collected using paper questionnaires in classrooms under examination conditions by trained fieldworkers blind to allocation. Using computer-generated random numbers, schools were then allocated 1:1 to intervention or control

stratified by schools: single sex versus mixed sex; school-level student free-school-meal eligibility (0%–23%; >23%) indicating poverty; and General Certificate of Secondary Education results accounting for student baseline attainment (above/below the median score for England of 1000).

All staff in intervention schools received training to implement restorative practices. Around 5–10 staff per school received in-depth training to deliver restorative conferences. All schools were provided with a manual to guide the action group comprising at least six staff and six students, led by a member of the school's senior leadership team. Groups were supported by external facilitators in the first 2 years of intervention but in the third year, the group was facilitated by school staff only. Action groups aimed to revise rules and policies so that these supported deliveries of restorative practice and coordinate implementation. Schools were provided with lesson plans and slides to guide the delivery of a social and emotional skills curriculum targeting students in years 8–10 who received 5–10 hours of teaching per year. Schools randomised to the control group continued with the normal practice.

Primary outcomes were bullying victimisation measured by the Gatehouse Bullying Scale⁸ and perpetration of aggression measured by the Edinburgh Study of Youth Transitions and Crime Scale¹³ at 36 months. Secondary outcome included use of tobacco, alcohol and drugs, mental well-being, psychological functioning and quality of life and contacts with the police and National Health Service (NHS), as described in our protocol and the main trial paper.^{9,12} The focus of the present paper is on three potential mediator measures, informed by the intervention logic model (figure 1).

The first focused on student reports of school climate, assessed using an established measure, the Beyond Blue School Climate Questionnaire (BBSCQ) scale, which includes 28 student-reported items arranged in four subscales covering: staff–student relationships, student sense of belonging in the school community; student commitment to learning; and student participation at school (online supplementary table S1).¹⁴ Students were asked to rate their level of agreement with items and their responses scored between 1 (complete disagreement) and 4 (complete agreement). Scores were then averaged within the subscales to obtain the subscale scores, and across all items to obtain the overall BBSCQ score. This measure aligns closely with the key theoretical constructs from the theory of human functioning and school organisation concerning staff–student relationships, student commitment to learning, sense of belonging and participation.⁵

The second potential mediator focused on student reports of engagement with delinquent peers, assessed by the Young People's Development Programme single-item measure asking students whether their friends who are the same age as them have been told off, stopped or picked up by the police in the last 12 months.¹⁵ Data for these two measures were collected via baseline, interim and final follow-up student surveys.

The third potential mediator focused on staff reports of their perception of school organisational climate, using a new scale (online supplementary table S1), which was assessed for the reliability at baseline and amended to include the most reliable items.¹⁶ This was a 26-item scale with subscales measuring: whether the authority is distributed among staff; staff–student relationships; integration of students' academic education and broader development; and school–community relationships. Staff were asked to rate their level of agreement with items, with responses scored between 1 (strongly agree) and 4 (strongly disagree). Items were recoded so that a higher score indicated

what, from the perspective of our theory of change, would be regarded as a less-healthy school organisational climate. Factor scores were derived within the subscales to obtain the subscale scores and across the subscale scores to obtain the overall score. Data for this measure were collected via structured telephone interviews from staff in intervention and control schools. 'Baseline' data were collected in September–November 2014 from one member of each school's senior leadership team and two other members of staff identified by this individual. Interim follow-up data were then collected in September–November 2017 from one member of each school's senior leadership team or the staff member leading the action group.

The trial was approved by the UCL (ref 5248/001) and IoE (ref. FCL 566) Research Ethics Committees. Written, informed consent was obtained at school level (head-teacher) for random allocation and intervention, and at the individual-participant level for data collection. Information sheets and consent forms for surveys were identical in intervention and control schools. Parents of students were informed about the study and could withdraw their children from research activities. The trial was prospectively registered as ISRCTN10751359 with the ISRCTN Registry on 30 January 2014.⁹

For potential mediators, completion was assessed in terms of the proportion of items completed by participants and the proportion of participants who completed at least half of the items in a subscale or scale. We then assessed interitem scale reliability using Cronbach's alpha. We used this rather than ordinal alphas for ordinal scales for simplicity and because Cronbach's alpha provides suitably conservative estimates for such measures.¹⁷

We then used a causal steps approach to assess whether our potential mediators might have mediated intervention effects on our primary and secondary outcomes.¹⁸ Four criteria needed to be met. The first, that the intervention was associated with the outcomes, was examined in our main trial analysis where, as previously stated, we observed a range of statistically significant positive effects. The second criterion required an association between the intervention and the potential mediators. We therefore assessed associations between trial arm and: our student-reported potential mediators measured at interim and final follow-ups; and our staff-reported potential mediator measured at the interim follow-up. These analyses, like those reported on our primary and secondary outcomes, used the intention-to-treat principle that is including all schools and participants in their groups as allocated. Each measure was analysed using a separate mixed model with the outcomes from each time-point treated as a repeated measure outcome. Fixed effects of the arm, follow-up time and the interaction between arm and time were specified, and the estimated baseline measures were constrained to be identical in the two arms of the trial, equivalent to adjusting for baselines. Random effects for schools and participants were specified to allow for correlations within schools and repeated measures within participants. Explicit consideration of potential confounders and how they are controlled for is an important part of statistical mediation analysis. We, therefore, report unadjusted analyses as well as analyses adjusted for baseline measures of the outcomes, sex, ethnicity, socioeconomic status (measured using the Family Affluence Scale)¹⁹ as well as for the school-level stratifying factors (single-sex versus mixed-sex school; school level deprivation; academic attainment strata). For continuous outcomes, we report unadjusted and adjusted mean differences with 95% CIs and adjusted effect sizes (standardised mean difference). For binary and ordinal outcomes, we report unadjusted and adjusted ORs.

Research report

The third criterion required that the potential mediator was associated with the outcomes. We, therefore, examined associations between our potential mediators measured at the interim follow-up, and our primary and secondary outcomes measured at the final follow-up so that we were assured of the temporality of the associations. Adjustment for potential confounders was as per previous analyses. The final criterion required that adjusting for potential mediators reduced the association between the intervention and the outcome measures. Therefore, we assessed the effect of adjusting for our potential mediators measured at the interim follow-up on the associations between trial arm and our primary and secondary outcomes measured at the final follow-up in the fully adjusted analysis. As differences between students who completed both baseline and follow-up surveys and those who only completed baseline were observed, we conducted a sensitivity analysis using multiple imputation by chained equations to impute missing data for participants with incomplete outcome data. All analyses were completed by staff blind to allocation.

RESULTS

For trial participation and follow-up rates, participant characteristics and overall outcomes by allocation, see the main trial report.¹² Student data were available from all schools. Student-reported measures of potential mediators had good

response rates and reliability with over 85% of all participants completing all items and with multi-item scales and subscales having Cronbach's alphas of >90% for the overall scale and 70% for the subscales at the baseline and interim follow-up (table 1). Staff data were available from 40 schools at the baseline and 31 schools at the interim follow-up. The staff-reported potential mediator had slightly lower item-response rates, but still >70%. Cronbach's alpha for the overall staff-reported measure was >80% at the baseline and over 60% at the interim follow-up, with Cronbach's alphas for the subscales being somewhat lower. There were small differences between students completing all surveys and those completing only baselines, with attrition higher among those from smaller schools or more deprived neighbourhoods, of white British ethnicity, with non-working parents or not living with two biological parents (table 2).

The intervention had statistically significant effects on potential mediators, being associated with increased student-reported-positive school climate as well as its constituent subscales at the final but not the interim follow-up. The intervention was also significantly associated with a decrease in student-reported measure of involvement with delinquent peers at the interim follow-up but one slightly short of statistical significance at the final follow-up (table 3). The intervention had no statistically significant effects on the overall

Table 1 Mediator measures response and reliability

Measure	Response rates				Internal consistency—Cronbach's alpha (standardised)	
	Baseline		Interim follow-up		Baseline	Interim
	Completed all items (n)	Completed half or more of items (n)	Completed all items (n)	Completed half or more of items (n)		
Student view on school climate						
Overall	5733 85.99%	6635 99.52%	5549 88.22%	6265 99.60%	0.9137	0.9170
Student sense of belonging subscale	6293 94.39%	6613 99.19%	5965 94.83%	6240 99.21%	0.7952	0.8225
Student commitment to academic values subscale	6519 97.78%	6581 98.71%	6190 98.41%	6231 99.06%	0.7394	0.7732
Student perception of supportive teacher relationships subscale	6221 93.31%	6631 99.46%	5935 94.36%	6247 99.32%	0.8804	0.8938
Student perception of participative school environment subscale	6396 95.94%	6600 99.00%	6071 96.52%	6231 99.06%	0.8005	0.8313
Student report of friends' contact with police in the last year	6494 97.41%	NA	6167 98.04%	NA	NA	NA
Staff view on school or organisational climate						
Overall	93 77.50%	99 82.50%	31 77.50%	31 77.50%	0.8357	0.6266
Authority distributed among staff subscale	98 81.67%	99 82.50%	31 77.50%	31 77.50%	0.6340	0.5952
Staff relationships with students subscale	98 81.67%	99 82.50%	31 77.50%	31 77.50%	0.7349	0.7633
Integration of students' academic education and broader social development subscale	98 81.67%	99 82.50%	31 77.50%	31 77.50%	0.6984	0.4839
School-community relationships subscale	96 80.00%	99 82.50%	31 77.50%	31 77.50%	0.7363	0.7536

Table 2 Differences in characteristics between students completing all surveys and those completing baseline only

Covariate	Category	Total	Baseline data only
		N (%) n=3337	N (%) n=1040
School size	Small	1574 (47.17)	542 (52.12)
	Large	1763 (52.83)	498 (47.88)
School neighbourhood deprivation	Low score	1696 (47.83)	525 (50.48)
	High score	1741 (52.17)	515 (49.52)
Free school meal eligibility	Low score	1663 (49.84)	527 (50.67)
	High score	1674 (50.16)	513 (49.33)
Student gender	Female	1644 (50.15)	515 (50.54)
	Male	1634 (49.85)	504 (49.46)
Student ethnicity	White British	1383 (41.46)	446 (43.34)
	Other	1921 (58.14)	583 (56.66)
Family structure	Two parents	2388 (71.97)	647 (62.75)
	Other	930 (28.03)	384 (37.25)
Parental working	Not in work	287 (10.57)	115 (14.32)
	In work	2429 (89.43)	688 (85.68)
Family affluence	Low affluence	1194 (36.81)	371 (36.88)
	High affluence	2050 (63.19)	365 (63.12)

staff-reported measure of school organisational climate or its constituent subscales measured at the interim follow-up.

Student reports of school climate and of friends having contact with police measured at the interim follow-up were associated with primary health outcome measures as well as with secondary health outcome measures (other than age of sexual debut and use of contraception at last sexual intercourse) at the final follow-up in unadjusted and adjusted analyses (table 4). Staff reports of school organisational climate measured at the interim follow-up were not associated with any primary or secondary outcomes at the final follow-up in unadjusted or adjusted analyses, other than for one unadjusted association between school organisational climate and students' use of NHS in the past 12 months, which disappeared on adjustment.

Our analysis of the effect of additionally adjusting for potential mediators measured at the interim follow-up on the associations previously found between the intervention and primary and secondary health outcomes measured at the final follow-up found no evidence that this adjustment made any difference except marginally in the case of the intervention effect on well-being, where adjustment for both student-reported school climate and friends' contact with the police removed the previously statistically marginal intervention effect at the final follow-up (table 5).

The multiple imputation analysis produced results, available on request, which did not differ from the main analysis in the pattern, size or statistical significance of the associations found.

CONCLUSION

Summary of key findings

The student-reported measures of potential mediators had good response rates and reliability. The staff-reported measure had somewhat lower response rates and interitem reliability. The intervention appeared to impact on: student perceptions

of school climate but these (like intervention effects on primary and secondary student health outcomes) did not manifest until the final follow-up; and student contact with delinquent peers at the interim follow-up. The student-reported potential mediators measured at the interim follow-up were associated with most student health outcomes. Adjustment for student-reported school climate and contact with delinquent peers at the interim follow-up did not affect associations between intervention and health outcomes.

Study limitations

The study used a well-established, multi-item measure of student views of school climate drawing on data from all students completing surveys. However, the study used a weaker, single-item measure of student involvement with delinquent peers and used a new and less reliable staff-reported measure of school organisational climate which drew on very small samples of staff (three individuals at baseline and one individual at the interim follow-up). Therefore, our ability to assess whether student contact with delinquent peers and staff-reported organisational climate acted as potential mediators of intervention effects on student health may have been limited. The RCT included the interim follow-up at 24 months and the final follow-up at 36 months. Intervention effects on primary and secondary health outcomes as well as effects on student views of school climate manifested only at the latter time-point. This meant that our ability to determine definitely whether intervention effects on school climate might mediate effects on student health outcomes was limited. The number of statistical tests might have introduced some false positive results, but we hope that being focused on prior hypotheses limited bias.

Implications for research and policy

Our study provides the first evidence from an RCT that well-implemented whole-school interventions may be effective both in promoting student health and in improving student relationships with teachers, and sense of commitment, belonging and participation at school. Despite being constrained in its ability to assess mediation by the late manifestation of impacts on student views of school, our study suggests that mediation of intervention effects via these factors is at least plausible, in that the intervention had effects on student-reported potential mediators, which were associated with student health outcomes. The lack of evidence for mediation for the staff-reported measure of school organisational climate may have reflected the poor reliability of this measure. The fact that implementation fidelity was stronger for restorative practice and action groups rather than the curriculum further suggests that any impacts on health outcomes were likely to be achieved through changes to school environment rather than through changes to individual student health-related knowledge and skills. Further research is therefore required to investigate these matters with a better measure of school organisation and longer time scales between measurements. Nonetheless, our study offers tentative evidence that whole-school interventions might work by modifying the school climate as predicted by the theory of human functioning and school organisation.⁵ Regarding generalisability, our trial was carried out in a representative sample of schools in and around London. Our process evaluation identified no factors

Table 3 Intervention effects on mediators at the interim and final follow-ups

Measure	Interim follow-up				Final follow-up			
	Arm		Unadjusted effect		Arm		Unadjusted effect	
	Control Mean (SE)	Intervention Mean (SE)	Difference (95% CI)	P value	Control Mean (SE)	Intervention Mean (SE)	Difference (95% CI)	P value
Student view of school climate								
Overall	2.92 (0.03)	2.90 (0.03)	-0.00 (-0.02 to 0.02)	0.915	2.82 (0.03)	2.85 (0.03)	0.05 (0.03 to 0.08)	<0.001
Student perception of supportive teacher relationships subscale	2.76 (0.04)	2.70 (0.04)	-0.03 (-0.06 to 0.00)		2.64 (0.03)	2.66 (0.03)	0.05 (0.02 to 0.08)	0.06
Student sense of belonging	2.84 (0.03)	2.88 (0.03)	0.04 (0.01 to 0.07)		2.78 (0.02)	2.84 (0.02)	0.06 (0.03 to 0.09)	0.06
Student perception of participative school environment subscale	2.96 (0.03)	2.91 (0.03)	-0.02 (-0.05 to 0.01)		2.81 (0.04)	2.82 (0.04)	0.05 (0.02 to 0.08)	0.05
Student commitment to academic values subscale	3.51 (0.01)	3.53 (0.01)	0.00 (-0.02 to 0.03)		3.42 (0.01)	3.46 (0.01)	0.03 (0.01 to 0.06)	0.03
Staff view of school organisational climate								
Overall*	-0.13 (0.14)	-0.05 (0.29)	0.08 (-0.46 to 0.63)	0.766				0.499
Authority distributed among staff subscale	2.04 (0.11)	2.49 (0.20)	0.45 (0.05 to 0.85)					
Staff relationships with students subscale	3.04 (0.14)	2.94 (0.23)	-0.10 (-0.58 to 0.38)					
Integration of students' academic education and broader social development subscale	2.07 (0.10)	1.65 (0.09)	-0.43 (-0.70 to -0.15)					
School-community relationships subscale	1.98 (0.13)	2.07 (0.20)	0.08 (-0.35 to 0.52)					
Student report of friends' contact with police in the last year								
No	2203 (70.05)	2271 (75.15)	0.80 (0.66 to 0.96)	0.024	2073 (68.64)	2044 (73.21)	0.84 (0.69 to 1.02)	0.074
Yes	942 (29.95)	751 (24.85)			947 (31.36)	748 (26.79)		0.83 (0.69 to 1.01)

*The overall score is a weighted sum of subscales; its range is therefore expected to differ from those of the subscales on which it is based.

Table 4 Effects of potential mediators at interim follow-up on outcomes at the final follow-up

Continuous student outcomes	Student view of school climate			Friends' contact with police in last year			Staff view of school organisational climate		
	Unadjusted effect	Adjusted effect		Unadjusted effect	Adjusted effect		Unadjusted effect	Adjusted effect	
	Difference (95% CI)	Difference (95% CI)	P value	Difference (95% CI)	Difference (95% CI)	P value	Difference (95% CI)	Difference (95% CI)	
Bullying victimisation (GBS overall score)	-0.20 (-0.24 to -0.17)	-0.22 (-0.25 to -0.18)	<0.001	0.12 (0.08 to 0.15)	0.12 (0.09 to 0.16)	<0.001	-0.01 (-0.04 to 0.04)	0.00 (-0.03 to 0.04)	
Aggression perpetration (ESYTC overall score)	-2.49 (-2.87 to -2.11)	-2.57 (-2.96 to -2.18)	<0.001	2.73 (2.34 to 3.11)	2.73 (2.32 to 3.11)	<0.001	-0.20 (-0.59 to 0.19)	0.07 (-0.36 to 0.50)	
Quality of life (PedsQL overall score)	5.81 (4.84 to 6.77)	5.84 (4.87 to 6.81)	<0.001	-2.03 (-3.00 to -1.07)	-2.40 (-3.36 to -1.44)	<0.001	-0.50 (-1.58 to 0.59)	0.29 (-0.70 to 1.27)	
Psychological functioning (SDQ total difficulties score)	-3.17 (-3.54 to -2.79)	-3.21 (-3.59 to -2.83)	<0.001	1.17 (0.80 to 1.55)	1.31 (0.93 to 1.69)	<0.001	0.01 (-0.39 to 0.41)	0.06 (-0.33 to 0.45)	
SWEMWBS total well-being index	3.83 (3.44 to 4.22)	3.91 (3.51 to 4.31)	<0.001	-0.71 (-1.12 to -0.31)	-0.81 (-1.22 to -0.40)	<0.001	0.12 (-0.28 to 0.53)	0.23 (-0.19 to 0.64)	
Health utility (CHUBD overall score)	0.06 (0.05 to 0.06)	0.06 (0.05 to 0.07)	<0.001	-0.02 (-0.03 to -0.01)	-0.02 (-0.03 to -0.01)	<0.001	-0.00 (-0.01 to 0.00)	0.00 (-0.01 to 0.01)	
Age of sexual debut	-0.16 (-1.04 to 0.72)	0.09 (-0.81 to 1.00)	0.724	0.61 (-0.22 to -1.43)	0.46 (-0.37 to 1.29)	0.281	-0.46 (-1.41 to 0.50)	-0.95 (-2.02 to 0.12)	
Bullying perpetration (MAS bullying subscale score)	-1.54 (-1.75 to -1.33)	-1.53 (-1.74 to -1.33)	<0.001	1.55 (1.34 to 1.77)	1.46 (1.25 to 1.67)	<0.001	-0.13 (-0.58 to 0.31)	0.10 (-0.20 to 0.39)	
Categorical student outcomes	OR (95% CI)	OR (95% CI)	P value	OR (95% CI)	OR (95% CI)	P value	OR (95% CI)	OR (95% CI)	
Ever smoked	0.42 (0.35 to 0.51)	0.40 (0.33 to 0.49)	<0.001	3.89 (3.24 to 4.66)	4.12 (3.41 to 4.97)	<0.001	0.92 (0.67 to 1.26)	0.96 (0.71 to 1.30)	
Ever drunk alcohol	0.46 (0.39 to 0.55)	0.44 (0.37 to 0.53)	<0.001	2.63 (2.22 to 3.11)	2.69 (2.25 to 3.22)	<0.001	0.94 (0.60 to 1.48)	1.08 (0.82 to 1.41)	
Ever been really drunk	0.62 (0.50 to 0.76)	0.58 (0.47 to 0.71)	<0.001	1.86 (1.54 to 2.24)	2.03 (1.67 to 2.47)	<0.001	0.91 (0.70 to 1.18)	0.95 (0.73 to 1.25)	
Ever been offered illicit drugs	0.36 (0.31 to 0.42)	0.33 (0.29 to 0.39)	<0.001	3.87 (3.27 to 4.44)	3.91 (3.29 to 4.51)	>0.001	1.05 (0.81 to 1.35)	1.08 (0.85 to 1.38)	
Used contraception at last sexual intercourse	1.69 (1.00 to 2.86)	2.22 (1.21 to 4.07)	0.049	1.38 (0.83 to 2.31)	1.25 (0.72 to 2.18)	0.435	0.70 (0.48 to 1.02)	0.76 (0.44 to 1.29)	
Use of NHS in past 12 months	0.80 (0.70 to 0.92)	0.79 (0.68 to 0.92)	0.002	1.45 (1.26 to 1.66)	1.42 (1.22 to 1.64)	>0.001	0.87 (0.78 to 0.97)	0.89 (0.79 to 0.99)	
Contact with police	0.39 (0.32 to 0.49)	0.37 (0.29 to 0.46)	<0.001	5.08 (4.15 to 6.22)	5.09 (4.13 to 6.28)	<0.001	0.80 (0.64 to 0.99)	0.86 (0.64 to 1.03)	

CHUBD, Child Health Utility 9D; ESYTC, Edinburgh Study of Youth Transitions and Crime; GBS, Gatehouse Bullying Scale; MAS, Modified Aggression Scale; NHS, National Health Service; PedsQL, Pediatric Quality of Life Inventory; SDQ, Strengths and Difficulties Questionnaire; SWEMWBS, Short Warwick Edinburgh Mental Wellbeing Scale.

Table 5 Intervention effects on student primary and secondary outcomes at the final follow-up before and after adjusting for potential mediators at the interim follow-up

Continuous student outcomes	Arm		Unadjusted effect		Adjusted effect (baseline covariates)		Adjusted effect (baseline covariates plus student view of school climate)		Adjusted effect (baseline covariates plus friends' contact with police in last year)		Adjusted effect (baseline covariates plus staff view of school organisational climate)	
	Control Mean (SE)	Intervention Mean (SE)	Difference (95% CI)	P value	Difference (95% CI)	P value	Difference (95% CI)	P value	Difference (95% CI)	P value	Difference (95% CI)	P value
Bullying victimisation (GBS overall score)	0.34 (0.02)	0.23 (0.02)	-0.03 (-0.06 to -0.00)	0.039	-0.03 (-0.06 to -0.00)	0.044	-0.04 (-0.08 to -0.01)	0.007	-0.04 (-0.07 to -0.01)	0.016	-0.02 (-0.07 to 0.03)	0.380
Aggression perpetration (ESTYC overall score)	4.33 (0.20)	4.04 (0.21)	-0.07 (-0.38 to 0.25)	0.684	-0.13 (-0.43 to 0.18)	0.421	-0.20 (-0.52 to 0.13)	0.229	-0.12 (-0.44 to 0.20)	0.469	0.03 (-0.58 to 0.64)	0.917
Quality of life (PedsQL overall score)	78.82 (0.54)	80.65 (0.55)	1.16 (0.41 to 1.90)	0.002	1.44 (0.70 to 2.17)	<0.001	1.42 (0.64 to 2.21)	<0.001	1.26 (0.49 to 2.03)	0.001	1.43 (0.09 to 2.77)	0.036
Psychological functioning (SDQ total difficulties score)	12.20 (0.18)	11.51 (0.19)	-0.51 (-0.80 to -0.22)	<0.001	-0.54 (-0.83 to -0.25)	<0.001	-0.62 (-0.92 to -0.31)	<0.001	-0.54 (-0.85 to -0.24)	<0.001	-0.38 (-0.92 to 0.17)	0.170
SFEMWBS total well-being index)	22.88 (0.19)	23.32 (0.19)	0.27 (-0.06 to 0.60)	0.115	0.33 (0.00 to 0.66)	0.048	0.27 (-0.08 to 0.62)	0.141	0.23 (-0.13 to 0.58)	0.210	0.21 (-0.36 to 0.80)	0.463
Health utility (CHU9D overall score)	0.85 (0.00)	0.86 (0.01)	0.01 (-0.00 to 0.01)	0.244	0.01 (0.00 to 0.01)	0.080	0.00 (-0.00 to 0.01)	0.153	0.00 (-0.01 to 0.01)	0.217	0.00 (-0.01 to 0.01)	0.570
Age of sexual debut	13.11 (0.43)	12.54 (0.49)	-0.58 (-1.97 to 0.81)	0.416	-0.35 (-1.48 to 0.78)	0.541	-0.24 (-1.46 to 1.00)	0.703	-0.40 (-1.65 to 0.84)	0.525	-0.51 (-2.03 to 1.01)	0.513
Bullying perpetration (MAS bullying subscale score)	2.75 (0.21)	2.33 (0.21)	-0.28 (-0.84 to 0.29)	0.334	-0.26 (-0.57 to 0.05)	0.097	-0.29 (-0.57 to -0.00)	0.047	-0.16 (-0.45 to 0.13)	0.279	-0.27 (-0.69 to 0.14)	0.193
Categorical student outcomes	Control n (%)	Intervention n (%)	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
Ever smoked												
No	2293 (77.70)	2318 (64.17)	0.59 (0.43 to 0.81)	0.001	0.56 (0.43 to 0.78)	0.001	0.55 (0.39 to 0.78)	0.001	0.57 (0.40 to 0.81)	0.002	0.65 (0.44 to 0.97)	0.037
Yes	658 (22.30)	436 (15.83)										
Ever drunk alcohol												
No	1677 (56.43)	1735 (62.43)	0.75 (0.58 to 0.79)	0.029	0.72 (0.56 to 0.92)	0.009	0.72 (0.55 to 0.95)	0.019	0.74 (0.56 to 0.97)	0.028	0.76 (0.53 to 1.09)	0.134
Yes	1295 (43.57)	1044 (37.57)										
Ever been really drunk												
No	788 (53.14)	721 (61.21)	0.50 (0.29 to 0.87)	0.014	0.47 (0.31 to 0.71)	<0.001	0.42 (0.27 to 0.65)	<0.001	0.47 (0.31 to 0.71)	<0.001	0.66 (0.47 to 0.93)	0.019
Yes	695 (46.86)	457 (38.79)										
Ever been offered illicit drugs												
No	1913 (64.41)	1997 (72.54)	0.52 (0.34 to 0.79)	0.002	0.51 (0.36 to 0.73)	<0.001	0.44 (0.30 to 0.64)	<0.001	0.51 (0.36 to 0.73)	<0.001	0.70 (0.50 to 0.96)	0.028
Yes, but did not try	744 (25.05)	567 (20.60)										
Yes, and tried them	313 (10.54)	189 (6.87)										
Used contraception at last sexual intercourse												
No	64 (23.10)	36 (21.95)	1.18 (0.56 to 2.48)	0.658	1.08 (0.50 to 2.35)	0.841	1.60 (0.65 to 3.91)	0.305	1.34 (0.56 to 3.22)	0.511	1.40 (0.64 to 3.06)	0.405
Yes	213 (76.90)	128 (78.05)										
Use of NHS in past 12 months												
No	1605 (53.22)	1472 (52.59)	0.96 (0.83 to 1.12)	0.639	0.96 (0.82 to 1.11)	0.565	0.96 (0.81 to 1.13)	0.600	0.99 (0.84 to 1.17)	0.916	0.96 (0.82 to 1.13)	0.613
Yes	1411 (46.78)	1327 (47.41)										

Continued

Table 5 Continued

Categorical student outcomes	Control n (%)	Intervention n (%)	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
Contact with police										
No	2626 (86.52)	2485 (88.43)	0.75 (0.57 to 0.99)	0.041	0.74 (0.56 to 0.97)	0.027	0.66 (0.49 to 0.88)	0.004	0.72 (0.54 to 0.95)	0.022
Yes	409 (13.48)	325 (11.57)								0.641

CHUSQ, Child Health Utility 9D NHS National Health Service; ESYTC, Edinburgh Study of Youth Transitions and Crime; GBS, Gatehouse Bullying Scale; PeaSCQ, Pediatric Quality of Life Inventory; SDO, Strengths and Difficulties Questionnaire; SWEMWBS, Short Warwick Edinburgh Mental Wellbeing Scale; MAS, Modified Aggression Scale;

that might suggest that implementation or effects would be different in other English schools.

What is already known on this subject

- ▶ Young people's health can be improved by 'whole-school' interventions that aim to render schools more engaging, participative and inclusive.
- ▶ But previous evaluations have not examined mediators to assess how such interventions might work.

What this study adds

- ▶ We found that as well as improving a wide range of health outcomes, our intervention also improved student reports of an engaging, participative and inclusive school climate, and reduced reports of students having delinquent friends. These factors were themselves associated with health outcomes.
- ▶ This suggests whole-school interventions might work by engaging young people with school and reducing engagement with pro-risk peers.

Acknowledgements Miranda Perry played a key role in developing the intervention and coordinated its implementation in the pilot and phase III trials. The authors are grateful to the staff and students of participating schools for their dedication to the intervention and completion of the outcome surveys and process evaluation surveys and interviews. They are very grateful for the advice and support of our Trial Steering Committee and Data Monitoring Committee. They acknowledge the work of and mourn the loss of Dr Farah Jamal during the trial, whose death at the age of 30 was a tragic loss for public health research.

Contributors CB and RVM: directed the trial from which the data are drawn. CB and EA: designed the analysis for this paper. CO: implemented this design and undertook the analysis. DRE: provided additional statistical expertise. CB: drafted the paper with inputs from EA, CO, DRE and RMV. EW, LB, JM, AM and JS: worked on the trial, contributed to the design of data collection instruments and contributed to the drafting of the paper.

Funding This paper presents independent research funded by the National Institute for Health Research (NIHR) in England under its Public Health Research Board (12/153/60) and the Education Endowment Foundation.

Disclaimer The views expressed in this publication are those of the authors and do not necessarily reflect those of the National Health Service (NHS), the NIHR or the Department of Health for England.

Competing interests RV is President of the Royal College of Paediatrics & Child Health.

Patient consent Obtained.

Ethics approval UCL (reference number 5248/001).

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- 1 Patton G, Sawyer SM, Santelli JS, *et al*. Realising the Future of Global Health: A Lancet Commission on Adolescent Health and Wellbeing. *The Lancet* 2016;387:2423–78.
- 2 Langford R, Bonell CP, Jones HE, *et al*. The WHO Health Promoting School framework for improving the health and well-being of students and their academic achievement. *Cochrane Database Syst Rev* 2014;CD008958.
- 3 Bonell C, Jamal F, Harden A, *et al*. Systematic review of the effects of schools and school environment interventions on health: evidence mapping and synthesis. *Public Health Res* 2013;1:1–320.
- 4 Bonell CP, Fletcher A, Jamal F, *et al*. Theories of how the school environment impacts on student health: systematic review and synthesis. *Health Place* 2013;24:242–9.

Research report

- 5 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.
- 6 Jamal F, Fletcher A, Harden A, et al. The school environment and student health: a systematic review and meta-ethnography of qualitative research. *BMC Public Health* 2013;13:798.
- 7 Bonell C, Pary W, Wells H, et al. The effects of the school environment on student health: a systematic review of multi-level studies. *Health Place* 2013;21:180–91.
- 8 Bond L, Patton G, Glover S, 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.
- 9 Bonell C, Mathiot A, Allen E, et al. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE) trial: update to cluster randomised controlled trial protocol. *Trials* 2017;18:238.
- 10 Morrison B. Restorative justice in schools. In: Elliot E, Gordon RM, eds. *New directions in restorative justice: Issues, practices, evaluation*. Cullompton, UK: Willan Publishing, 2005:26–52.
- 11 Durlak JA, Weissberg RP, Dymnicki AB, et al. The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Dev* 2011;82:405–32.
- 12 Bonell C, Allen E, Warren E, et al. Initiating change in the school environment to reduce bullying and aggression: a cluster randomised controlled trial of the Learning Together (LT) intervention in English secondary schools. *The Lancet* 2018;392:2452–64.
- 13 Smith DJ. *School experience and delinquency at ages 13 to 16*. Edinburgh: Centre for Law and Society, University of Edinburgh, 2006.
- 14 Sawyer MG, Pfeiffer S, Spence SH, et al. School-based prevention of depression: a randomised controlled study of the beyondblue schools research initiative. *J Child Psychol Psychiatry* 2010;51:199–209.
- 15 Wiggins M, Bonell C, Sawtell M, et al. Health outcomes of youth development programme in England: prospective matched comparison study. *BMJ* 2009;339:b2534.
- 16 Shackleton N, Fletcher A, Jamal F, et al. A new measure of unhealthy school environments and its implications for critical assessments of health promotion in schools. *Crit Public Health* 2017;27:248–62.
- 17 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
- 18 Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986;51:1173–82.
- 19 Currie CE, Elton RA, Todd J, et al. Indicators of socioeconomic status for adolescents: the WHO Health Behaviour in School-aged Children Survey. *Health Educ Res* 1997;12:385–97.

**Appendix 10: Moderated mediation analyses to assess intervention mechanisms
for impacts on victimisation, psycho-social problems and mental wellbeing:
Evidence from the INCL USIVE realist randomized trial**



Moderated mediation analyses to assess intervention mechanisms for impacts on victimisation, psycho-social problems and mental wellbeing: Evidence from the INCLUSIVE realist randomized trial

G.J. Melendez-Torres^{a,*}, Emily Warren^b, Russell Viner^c, Elizabeth Allen^d, Chris Bonell^b

^a Peninsula Technology Assessment Group, College of Medicine and Health, University of Exeter, Exeter, UK

^b Department of Public Health, Environments and Society, London School of Hygiene and Tropical Medicine, London, UK

^c Department of Social Science, UCL Institute of Child Health, London, UK

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

ARTICLE INFO

Keywords:

Realist trials
School health
Mental health
Bullying
School climate

ABSTRACT

Realist evaluations aim to evaluate interventions by understanding the mechanisms they trigger, assessing not merely what works but what works for whom, under what conditions, and how. Significant disagreement in the literature exists as to whether randomized trials can be used as a tool for realist evaluation. INCLUSIVE, which was the first realist randomized trial explicitly designed as such, evaluated the impact of Learning Together, a school-based intervention for students aged 12–15 that included restorative practice, on bullying victimisation, mental wellbeing and psychological problems. Drawing on hypotheses generated through qualitative research, this analysis tested if school belonging was a mediator of intervention effects, and in which contexts. We estimated a series of fully longitudinal multilevel moderated mediation models including intervention allocation, student reports of school belonging at 24 months and victimisation and wellbeing outcomes at 36 months, and stratified on the basis of whether, at baseline, schools were: a) rated 'outstanding' for leadership, b) below the median for average levels of victimisation, and c) above the median on school inclusivity. Findings suggested that in unstratified models, belonging was not a mediator for any outcome; but in each of the strata defined above, belonging was a significant mediator at the student level. However, in the strata where belonging was not a mediator, the intervention still had a significant effect on each outcome. Analyses point to a strong but conditional role for belonging as a mediator of intervention pathways; in schools where belonging was not a mediator (e.g. above-median victimisation levels), other mechanisms may have been activated. This is consistent with a realist understanding of context-mechanism linkages generating outcomes. Our analyses suggest that realist evaluations can be pursued within randomized trials and that such analyses can offer more nuanced evidence regarding in which contexts interventions might effectively be implemented.

1. Introduction

Realist evaluations aim to evaluate interventions by understanding the mechanisms they trigger, assessing not merely what works but what works for whom, under what conditions, and how (Pawson and Tilley, 1997). They do so by formulating and assessing hypotheses in the form of what mechanisms operate in what contexts to generate what outcomes. Such analyses are potentially valuable in offering more nuanced suggestions as to the range of contexts that interventions might effectively be implemented within post-evaluation, and whether

interventions should be tailored to potentiate the mechanism most likely to occur in particular contexts.

While realist evaluation has been described as being 'methods neutral', there is disagreement and inconsistency within the literature on realist evaluations about whether or not randomized controlled trials (RCT) may be used as a tool of realist evaluation, with some describing realist trials as an oxymoron (Pawson and Tilley, 1997; Marchal et al., 2013). The latter position is argued in terms of trials in practice failing to include sufficiently heterogeneous contexts to enable hypotheses about contextual variation to be assessed, or more fundamentally, trials being

* Corresponding author.

E-mail addresses: g.j.melendez-torres@exeter.ac.uk (G.J. Melendez-Torres), emily.warren@lshtm.ac.uk (E. Warren), r.viner@ucl.ac.uk (R. Viner), elizabeth.allen@lshtm.ac.uk (E. Allen), chris.bonell@lshtm.ac.uk (C. Bonell).

<https://doi.org/10.1016/j.socscimed.2021.113984>

Received in revised form 5 April 2021; Accepted 29 April 2021

Available online 3 May 2021

0277-9536/© 2021 Elsevier Ltd. All rights reserved.

epistemologically positivist and thus inimical to realist enquiry (Pawson and Tilley, 1997; Marchal et al., 2013; Van Belle et al., 2016). As proponents of realist RCTs, we have previously challenged these arguments on methodological and theoretical grounds. We have argued that trials are not, of necessity, positivist, since they: embrace a hypothetico-deductivist, not an empiricist epistemology; need not imply a unity of methods with natural science research, for example, because they can include interpretive alongside correlational research; and do not presume non-contingent generalisability of findings (Bonell et al., 2012, 2016a, 2018a).

Despite these theoretical justifications for realist trials, to date no empirical analyses have demonstrated that trial data can be used to inform realist ends of developing and testing hypotheses about what mechanisms work in what contexts to generate what outcomes. The present paper draws on empirical data from the avowedly realist Initiating Change Locally in bullying and aggression through the School Environment (INCLUSIVE) RCT. It aims to explore whether trials might contribute to realist evaluation, in the manner we have previously proposed (Jamal et al., 2015), by drawing on novel qualitative and quantitative analyses, first to develop and then to test hypotheses about what mechanisms operate in what contexts to generate what outcomes.

The trial evaluated the Learning Together intervention, a whole-school intervention involving action groups and restorative practice aiming to reduce bullying and aggression, and promote student health and wellbeing in English secondary schools. The primary outcomes of this trial were a reduction in bullying victimisation and aggression perpetration at 36 months; however, a number of secondary outcomes related to mental wellbeing, psychological problems, risk behaviours and school commitment were collected as well. There is good evidence from previous studies that whole-school interventions involving student contribution to school policy via such groups are a promising means of preventing bullying and aggression, and promoting students' health (Smith et al., 2004; Vreeman and Carroll, 2007). There is also increasing evidence, though previously from quasi-experimental studies, that restorative practice interventions (primary restorative practice bringing together students to discuss their feelings and relationships to prevent conflict, and secondary restorative practice bringing together parties to conflict so that relationships may be healed and perpetrators appreciate the harms caused) can prevent violence and aggression in schools (Buckley and Maxwell, 2007; Kane et al., 2007; Kokotsaki et al., 2014).

The Learning Together intervention provided schools with a number of resources: intervention manual; report on student needs (involvement in bullying, substance use and other health-related behaviours, mental health and school experiences) based on baseline survey results; staff training in restorative practice (introductory 2-h training for all staff on use of restorative language and primary restorative practice to maintain good relationships, in-depth 3-day training for selected school staff with responsibility for behaviour management to support implementation of restorative conferences to address incidents addressing language, skills and delivery); external facilitator with experience of school management to support action groups and facilitate student contributions; and curriculum materials for social and emotional skills lessons. School staff and students drew on these resources to implement various activities. Action groups comprised of staff and students met twice-termly to review data, revise school policies and coordinate the intervention, tailoring this to local needs. Primary restorative practice (e.g. 'circle time') was used by teachers in classrooms to prevent misbehaviour, and secondary restorative practice (e.g. 'restorative conferences') was used by selected staff to address serious misbehaviour. A social and emotional skills curriculum was delivered by teachers to students in years 8–10 (age 12–15) for 5–10 h per year.

The intervention theory of change was informed by the theory of human functioning and school organisation. This proposes that schools can promote students' health by increasing students' commitment to learning and sense of belonging in the school community, which, particularly for socio-economically disadvantaged students, requires

eroding staff-student boundaries (increasing affective relationships) and reframing school provision on students' expressed needs. Increasing students' commitment and belonging in turn helps them develop their 'practical reasoning' capacity and peer affiliations supportive of healthier decisions (Markham and Aveyard, 2003).

Informed by this theory, the intervention was originally theorised to increase student sense of belonging in and commitment to school via: action groups re-focusing school policies and activities (e.g. policies and school systems addressing behaviour management, pastoral care, inclusion) on student needs as expressed in the need survey and action group meetings; action groups and restorative practices (circle time, restorative conferences) enhancing staff-student affecting relationships; and the curriculum building student practical reasoning capacity by promoting social and emotional skills.

Previous analyses from the INCLUSIVE RCT shed some light on the intervention's mechanisms but have not yet assessed whether there is evidence for different mechanisms generating different outcomes in different contexts. The main trial publication (Bonell et al., 2018b) focused on overall impacts, reporting a reduction in one primary outcome, self-reported bullying victimisation, but not the other, perpetration of aggression, at 36 months. It also reported effects on various secondary outcomes at 36 months: improved health-related quality of life and mental wellbeing; and reduced psycho-social problems, alcohol consumption, drunkenness, smoking, drug use and contact with the police; and reported larger effects on some outcomes for boys and for those reporting bullying victimisation and perpetration of aggression at baseline. However, despite the theory of change suggesting effects might be larger for socio-economically disadvantaged students, there was no evidence of greater benefits for students of low socio-economic status. The main trial report also reported that the curriculum element was delivered with poor fidelity and therefore was unlikely to explain outcomes. A subsequent paper analysed mediating variables across the sample and reported evidence of effects on student-reported belonging and commitment to school (measured using established scales (Sawyer et al., 2010)) but these manifested only at 36 months not 24 months, and there was no evidence that intermediate impacts on belonging or commitment at 24 months mediated the effects of allocation to the intervention on primary and secondary outcomes at 36 months (Bonell et al., 2019). Subsequently, and informed by realist evaluation, we sought to use novel qualitative and quantitative analyses to examine whether different intervention mechanisms might generate outcomes in different contexts. We have previously reported our novel use of qualitative data from interviews and focus groups with students and staff to explore this question (Warren et al., 2019, 2020). In this paper, we draw on these qualitative analyses to define hypotheses which are then tested using novel analyses of moderated mediation.

These qualitative analyses suggested that building students' sense of belonging was a more important intervention mechanism for reducing bullying and improving mental wellbeing than building commitment to learning. Qualitative data suggested that the action group's work could increase student belonging, both among students sitting on the group but also among students across the school, via: student awareness and approval of the action group's work; students becoming involved in activities spinning off from the group, such as rewriting school rules; and/or action groups implementing actions that benefited all students' sense of belonging (e.g. changes to school behaviour management, pastoral and/or inclusion policies). The qualitative research suggested that this belonging-based mechanism would only occur in schools with sufficient management capacity and a pre-existing inclusive ethos to ensure that the action groups functioned well enough to deliver these benefits.

However, the qualitative data also suggested that other intervention mechanisms not involving student belonging could still bring about impacts on student bullying and mental wellbeing, the key one being a mechanism whereby restorative practice directly reduced bullying and promoted student mental wellbeing by curtailing bullying and conflict

rather than via increasing belonging. This mechanism was most likely to predominate in schools where the belonging mechanism was less active and which needed to use restorative practice to address high rates of bullying (Warren et al., 2020).

Informed by this qualitative research, we hypothesised that, in high capacity/inclusive ethos schools not facing high rates of bullying, intervention effects reducing bullying and psychological problems and improving mental wellbeing would be mediated by increased student belonging. However, in schools with lower capacity, a less inclusive ethos and faced with high rates of bullying, intervention effects on bullying, psychological problems and mental wellbeing would occur directly via restorative practice and not be mediated by student belonging. These hypotheses are summarised in Fig. 1.

The present paper aims to test these hypotheses about which mechanisms generated what outcomes in which school contexts using novel analyses of moderated mediation. We examine the following questions:

1. Is school belonging a mediator, at the level of the school and the student, of intervention effects reducing bullying victimisation and psychological problems and improving mental wellbeing?
2. Do school contextual characteristics moderate the role of belonging as a mediator of intervention effects?

2. Methods

Analysis drew on data from the INCLUSIVE cluster RCT, which tested the effectiveness of the Learning Together intervention described above in secondary schools in England. Methods for the trial have been published elsewhere (Bonell et al., 2014, 2018b). In short, 40 broadly representative schools were randomly allocated after baseline surveys in a 1:1 ratio to either intervention or usual treatment. The intervention and trial period ran for 36 months, with student surveys at baseline (age 11–12), 24 months (age 13–14) and 36 months (age 14–15). Students consenting to participate and not withdrawn from the research by parents completed paper questionnaires in classrooms under examination conditions supported by trained researchers blinded to schools' allocation status, with teachers present but unable to read responses. The trial was approved by the University College London ethics committee (ref.

5248/001). Written, informed consent was sought from head teachers for allocation and intervention, and from individual students for survey participation.

This analysis aims to examine whether associations between an 'exposure' (in this case allocation to the intervention group) and outcomes (bullying victimisation, psychological problems and mental wellbeing) are mediated by another factor hypothesised to lie on the causal pathway from exposure to outcomes. The mediator used in this analysis was school belonging, a subscale of the Beyond Blue School Climate Questionnaire (Sawyer et al., 2010). School belonging is measured using an eight-item subscale of this questionnaire; items have a four-step Likert scale averaged to construct a score (see Supplementary File 1 for items and scoring). This measure was developed in Australia (Sawyer et al., 2010) using questions from the Gatehouse (Bond et al., 2004), Quality of School Life (Epstein and McPartland, 1976), Patterns of Adaptive Learning (Roeser et al., 1996), Manitoba School Improvement Survey (Earl and Lee, 1998) and Psychological Sense of School Membership (Goodenow, 1993) instruments. Cronbach's alpha for the subscale of 0.85 was reported for the original Australian adolescent sample (personal communication, Lyndal Bond, July 21, 2011) and of 0.80 for the present study sample (Bonell et al., 2016b).

The outcomes used in the analysis were the Gatehouse Bullying Scale, a six-item score of the frequency and impact of experience of different forms of bullying victimisation (for example, I have been deliberately left out), with range 0–12 (Bond et al., 2007); the Strengths and Difficulties Questionnaire, a standard measure of child psychological problems (for example, I fight a lot, I can make other people do what I want; I worry a lot), with range 0–35 (Goodman, 2006); and the Short Warwick-Edinburgh Mental Wellbeing Scale, which captures both subjective and functional psychological wellbeing (for example, I've been feeling optimistic about the future and I've been feeling close to other people), with range 7–35 (Clarke et al., 2011). In the primary trial analysis, the intervention reduced victimisation and psychological problems and improved mental wellbeing at 36 months; that is, there was a significant first-order effect of the intervention on each of the outcomes to be considered in these mediation models. Mediator and outcome variables were modelled using normal distributions.

We also considered three stratifying variables: whether schools had 'outstanding' Ofsted ratings for leadership and administration at

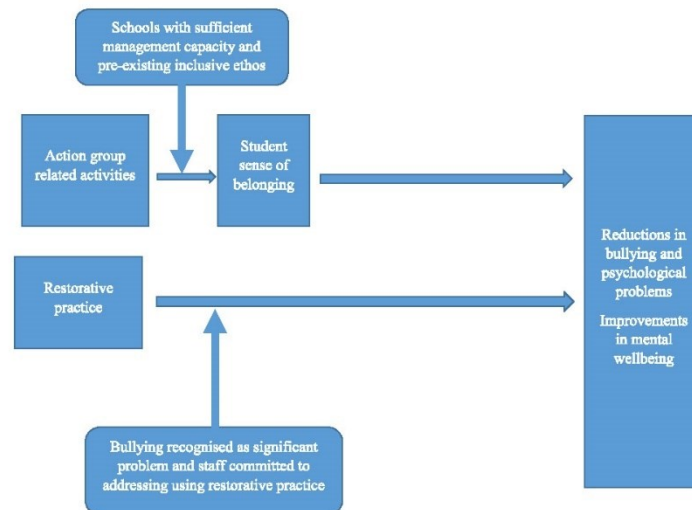


Fig. 1. Key hypotheses suggested by qualitative research.

baseline, to reflect hypotheses about school management capacity as an important contextual factor shaping intervention functioning; whether schools had above-median levels of student reports of inclusivity at baseline (measured via total summed scale scores for the Beyond Blue School Climate Questionnaire (Sawyer et al., 2010)) to reflect baseline school inclusive ethos; and whether schools had above-median levels of bullying victimisation at baseline, to reflect the salience of the behaviour the intervention sought to address.

To construct a fully longitudinal mediation model, we used the 24-month measurement wave of school belonging alongside the 36-month measurement wave of bullying victimisation, psychological problems and mental wellbeing. We used the 2-1-1 multilevel mediation model described by Pituch and Stapleton (2012), so named because it includes an intervention at level 2, or school level, and mediator and outcomes measured at level 1, or student level. An important feature of Pituch and Stapleton's model is that it disaggregates the impact of the mediator on the outcome into student-level and school-level contextual effects, in contrast to 'standard' 2-1-1 mediation models which only consider cluster-level pathways (that is, do not consider student-level relationships between mediator and outcome). This distinction is important because, when analysing mediation in multilevel contexts, involving student-level relationships and school-level relationships can lead to misleading and underpowered conclusions, and disaggregating student-level and school-level relationships can provide additional insights into how mediational pathways function. As defined by Raudenbush and Bryk (2002), contextual effects refer to the impact of a variable on an outcome modelled across multiple levels that arises above and beyond the individual-level relationship: for example, the relationship between average school socio-economic position and academic attainment that goes beyond the individual-level relationship between socio-economic position and attainment. The 2-1-1 multilevel mediation model used here models two separate mediational pathways corresponding to student-level mediation and school-level mediation, both of which share the same estimate of the intervention's impact on the mediator (see Fig. 2). Thus, each mediation model is composed of three regression equations estimated simultaneously: at school level, a) the school-level average of the mediator regressed on intervention allocation and b) the outcome regressed on both school-level average of the mediator and intervention allocation; and at student level, c) the school-centred value of the outcome regressed on the student-level value of the mediator.

Our analysis strategy unfolded in four steps, undertaken for each outcome separately. We undertook a separate analysis model for each outcome as the number of parameters in a simultaneous outcomes model would have exceeded the number of clusters, leading to unstable estimation and untrustworthy parameter estimates. First, we developed a mediation model using the regression specification as defined above. Second, we estimated mediation models stratified on each of the three grouping variables described above. Third, we examined the results of a multi-parameter Wald test comparing the magnitude of paths for the

same model between the two levels of the grouping variable. We used this to infer the presence of moderated mediation. Fourth, we quantified the indirect effect where it was appropriate to estimate this, using a Monte Carlo bootstrapping algorithm with 1 million draws. Student-level indirect effects thus refer to the product of the coefficient linking mediator to intervention status with the coefficient linking the mediator to the student-level mediator, while school-level indirect effects refer to the product of the coefficient linking mediator to intervention status with the coefficient linking the outcome to the school-level mediator.

All analyses were undertaken in Mplus v8.2 and used full information maximum likelihood for missing data.

3. Results

Of 7121 students registered in trial-participating schools at baseline, 6667 (93.6%) provided data at baseline: 3320 (94.4%) of 3516 in the intervention group and 3347 (92.8%) of 3605 in the control group. All schools participated in the follow-up surveys at 24 months and 36 months; the numbers of students who completed the questionnaires at baseline, 24 months (3074 in the intervention group, 3166 in the control group), and 36 months (2836 in the intervention group, 3054 in the control group) were similar in each group. Student and school characteristics and outcomes at baseline were well balanced across arms. The analysis sample for this study comprised 8179 students, of whom 4082 were in control schools and 4097 were in intervention arms. Descriptive statistics of variables used in this analysis and tables of relationships between stratification variables are presented in Supplementary File 1.

Victimisation. The unstratified model (see Table 1) did not suggest that belonging mediated the significant relationship between school allocation to the intervention and reductions in victimisation. While reports of lower victimisation at 36 months were linked to higher levels of belonging at 24 months at the student level, belonging was not linked to intervention.

However, models stratified by Ofsted rating for leadership suggested that in the group rated outstanding, belonging was a significant mediator for reductions in victimisation at the student, but not contextual, level. Specifically, within the outstanding subgroup, the intervention increased levels of belonging (mean difference [MD] = 0.197, standard error [SE] = 0.053). Subsequently, belonging was linked at the student level to reductions in victimisation level of about one point with each one-point improvement in belonging ($\beta = -0.971$, SE = 0.101). This was supported with a significant bootstrapped indirect effect, estimated by 'multiplying' the coefficients for difference in belonging by intervention and differences in victimisation by belonging ($\beta = -0.191$, 95% CI [-0.305, -0.087]). In contrast, there was no evidence of mediation through belonging in the subgroup not rated as outstanding, given no significant link between the intervention and belonging. Between models, paths were significantly different ($\chi^2 = 31.900$, $df = 4$, $p < 0.0001$).

Evidence for a similar pattern was found for schools that were below

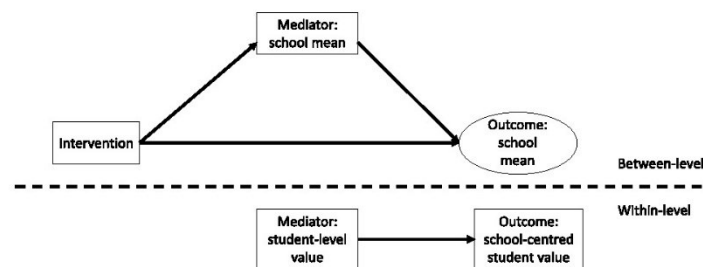


Fig. 2. Moderated mediation model diagram.

Table 1
Moderated mediation models for bullying victimisation.

	Unstratified	Ofsted rating		Victimisation at baseline		School inclusivity at baseline	
		Other	Outstanding	Above median	Below median	Below median	Above median
School-level path estimates							
Intervention → belonging	0.028 (0.035)	-0.025 (0.034)	0.197 (0.053)***	-0.057 (0.040)	0.112 (0.041)**	-0.038 (0.038)	0.094 (0.037)*
Intervention → victimisation	-0.215 (0.090)*	-0.203 (0.089)*	-0.015 (0.418)	-0.293 (0.109)**	-0.078 (0.151)	-0.284 (0.115)*	-0.125 (0.163)
Belonging → victimisation	0.510 (0.400)	0.688 (0.517)	-1.092 (1.402)	1.110 (0.514)*	-0.135 (0.651)	0.763 (0.476)	0.161 (0.723)
Student-level path estimates							
Belonging → victimisation	-1.127 (0.058)***	-1.183 (0.067)***	-0.971 (0.101)***	-1.178 (0.080)***	-1.091 (0.080)***	-1.208 (0.087)***	-1.075 (0.076)***
Indirect effects, student-level (asymmetric 95% CI)							
Wald test (χ^2, <i>df</i>, <i>p</i>-value)		31.900, 4, <0.0001	-0.191 (-0.305, -0.087)	12.486, 4, 0.014	-0.122 (-0.214, -0.034)	8.686, 4, 0.069	-0.380 (-0.676, -0.085)

Note. Estimates are presented as coefficient (standard error) unless otherwise noted. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the median for bullying victimisation at baseline, where belonging was a significant mediator at the student, but not school levels, with significant differences between strata in the path estimates ($\chi^2 = 12.486$, $df = 4$, $p = 0.014$) and a significant indirect ($\beta = -0.122$, 95% CI [-0.214, -0.034]).

While school inclusivity at baseline did not moderate the mediational pathway through belonging to victimisation ($\chi^2 = 8.686$, $df = 4$, $p = 0.069$), there is some evidence that belonging mediated at the student level, but not at the school level, between intervention and reductions in victimisation only in schools that were above the median for inclusivity at baseline, including a significant indirect effect ($\beta = -0.380$, 95% CI [-0.676, -0.085]).

Psychological problems. An unstratified model did not suggest that belonging was an overall mediator for the impact of the intervention on SDQ, as there was no link between belonging and allocation (see Table 2). However, stratified models and bootstrapped indirect effects suggested that belonging was a mediator at student level for reductions in psychological problems in schools that were rated outstanding for leadership, in schools below the median for victimisation at baseline, and for schools above the median for inclusivity at baseline. In each of these stratified models, path estimates were significantly different between strata.

Surprisingly, contextual effects for the mediator-outcome relationship were larger in each of these strata and in the opposite direction of the student-level effect, suggesting that intervention impacts are less strongly felt the greater the school-level improvement in belonging. However, these effects were all imprecisely estimated and non-

significant.

Mental wellbeing. An unstratified model did not suggest that belonging was a significant mediator of intervention impacts on mental wellbeing due to a non-significant link between mediator and intervention allocation (see Table 3). However, stratified models and bootstrapped indirect effects suggested that belonging was a mediator at student level for improvements in mental wellbeing in schools that were rated outstanding for leadership, in schools below the median for victimisation at baseline, and for schools above the median for inclusivity at baseline. In each of these stratified models, path estimates were significantly different between strata.

As was the case for analyses on psychological problems, contextual effects for the mediator-outcome relationship were in the opposite direction of the student-level effect.

4. Discussion

Summary of findings. We were able to draw on our prior qualitative research to develop hypotheses (see Fig. 1) about what mechanisms might generate what outcomes in what settings, which were more focused than those present in our original theory of change. We were then able to use novel analyses of moderated mediation to test these hypotheses in order to determine what mechanisms were most likely to generate outcomes in different settings.

Previous analyses across all schools found no evidence that student sense of belonging mediated the impact of intervention on bullying victimisation, psychological problems and mental wellbeing. However,

Table 2
Moderated mediation models for psychological problems.

	Unstratified	Ofsted rating		Victimisation at baseline		School inclusivity at baseline	
		Other	Outstanding	Above median	Below median	Below median	Above median
School-level path estimates							
Intervention → belonging	0.028 (0.035)	-0.025 (0.034)	0.197 (0.053)***	-0.057 (0.040)	0.112 (0.041)**	-0.038 (0.038)	0.094 (0.037)*
Intervention → SDQ	-0.654 (0.234)**	-0.886 (0.227)***	-0.954 (1.104)	-0.933 (0.305)**	-0.550 (0.295)	-1.259 (0.306)***	-0.362 (0.227)
Belonging → SDQ	2.424 (1.834)	0.457 (1.380)	7.836 (5.969)	0.716 (1.881)	3.122 (2.624)	0.586 (2.058)	4.277 (2.86)
Student-level path estimates							
Belonging → SDQ	-4.000 (0.157)***	-4.038 (0.172)***	-3.884 (0.360)***	-3.933 (0.296)***	-4.053 (0.156)***	-3.937 (0.280)***	-4.047 (0.180)***
Indirect effects, student-level (asymmetric 95% CI)							
Wald test (χ^2, <i>df</i>, <i>p</i>-value)		26.322, 4, <0.0001	-0.765 (-1.213, -0.352)	10.179, 4, 0.038	-0.454 (-0.784, -0.130)	14.611, 4, 0.006	-0.380 (-0.676, -0.085)

Note. Estimates are presented as coefficient (standard error) unless otherwise noted. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. SDQ: Strengths and Difficulties Questionnaire.

Table 3
Moderated mediation models for mental wellbeing.

	Unstratified	Ofsted rating		Victimisation at baseline		School inclusivity at baseline	
		Other	Outstanding	Above median	Below median	Below median	Above median
School-level path estimates							
Intervention → belonging	0.028 (0.035)	-0.025 (0.034)	0.197 (0.053)***	-0.057 (0.040)	0.112 (0.041)**	-0.038 (0.038)	0.094 (0.037)*
Intervention → mental wellbeing	0.417 (0.278)	0.625 (0.261) *	0.386 (1.466)	0.951 (0.328) **	0.042 (0.423)	1.080 (0.270) ***	0.219 (0.427) ***
Belonging → mental wellbeing	-2.479 (1.834)	-0.551 (1.674)	-4.400 (6.287)	-0.800 (2.249)	-3.017 (2.760)	2.094 (1.726)	-5.089 (2.783)
Student-level path estimates							
Belonging → mental wellbeing	3.673 (0.160) ***	3.687 (0.158) ***	3.704 (0.412)***	3.532 (0.206) ***	3.816 (0.227) ***	3.577 (0.244) ***	3.774 (0.207) ***
Indirect effects, student-level (asymmetric 95% CI)							
			0.729 (0.335, 1.172)		0.427 (0.120, 0.744)		0.355 (0.079, 0.634)
Wald test (χ^2, <i>df</i>, <i>p</i>-value)							
		22.522, 4, 0.0002		14.671, 4, 0.005		17.016, 4, 0.002	

Note. Estimates are presented as coefficient (standard error) unless otherwise noted. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

stratifying mediation models by school characteristics uncovered evidence of a mediational effect. This was principally by locating schools for which the relationship between belonging and intervention allocation was meaningful. Schools where belonging was a mediator for outcomes were defined by leadership rated as outstanding, below-median rates of bullying victimisation and above-median student-reported inclusive ethos at baseline, corresponding to mechanisms described in the top half of Fig. 1. Moreover, even in schools where belonging was not a mediator, there was still evidence that the Learning Together intervention was associated with benefits across these outcomes. Given the complex nature of the intervention, we would hypothesise that other mechanisms than improved belonging were activated to lower the rate of bullying in intervention schools. These may have involved the direct effects of restorative practice, such as de-escalating bullying and aggression and modelling prosocial skills, as identified through our qualitative research. These mechanisms are represented in Fig. 1 by the direct, unmediated pathways from the intervention to outcomes.

It is of note that our findings were consistent across the three 'positive' levels of the strata used in our analysis; that is, in schools that at baseline either reported above-median inclusivity, or below median bullying victimisation, or leadership rated as outstanding. Our exploration of these school characteristics identified that these are not all the same schools, suggesting that each stratifying variable captured a meaningfully different split of schools. However, it is possible, and worthy of further consideration, that because schools with one of these characteristics may also be more likely to have another characteristic, that these stratifying variables reflect different facets of an underlying construct.

Finally, in analyses for psychological problems and mental wellbeing but not for victimisation, school-level effects for the relationship between belonging and the outcomes were in the opposite direction to the student-level relationship. This was especially notable in strata where belonging was a significant mediator; as noted above, these were strata with schools that were more often than not already advantaged. These contextual effects, while consistently non-significant and imprecisely estimated, likely reflected a 'bounding effect', that is to say there was limited room for improvement. Another way of expressing this is to consider that, in schools that already had a positive environment before the trial, Learning Together may not have offered any school-level benefits, instead only producing an improvement in individual students' experiences.

Limitations. This paper explores only one mediator, concerning student sense of belonging in schools. Based on the analysis presented here, we can only conjecture that other mechanisms involving restorative practice curtailing bullying might underlie intervention effects in school contexts where effects were not mediated by belonging. We did not have quantitative measures at 24 months suitable for assessing these other mechanisms directly. Our results cannot automatically be

translated to other school settings, though our context-based analyses provide some indication of the types of schools most likely to benefit from an intervention such as Learning Together, and how they might benefit.

Implications for research and policy. These findings largely support the hypotheses informed by our qualitative research and summarised in Fig. 1 (Warren et al., 2019, 2020). The intervention likely triggered multiple mechanisms, the importance of which varied across school contexts. In schools with high baseline prevalence of bullying, the intervention was effective in reducing bullying victimisation, but this did not appear to involve a mechanism involving the building of belonging. In such schools, an alternative mechanism of identifying cases of bullying, and ensuring these were addressed and curtailed via use of restorative practice is possible; however, these mechanisms were not assessed in the present analysis.

The findings from our various analyses, taken together, offer some support for the theory of human functioning and school organisation but also suggest refinements (Markham and Aveyard, 2003). It appeared that taking steps to improve student-teacher relationships and re-centre provision of students' expressed needs did improve a range of health outcomes and, at least in some schools, this occurred through building students' sense of belonging in school. But, in the case of the Learning Together intervention, the intervention mechanism involving increased belonging was stronger in schools that already had strong capacity and a supportive ethos; and intervention mechanisms not involving student sense of belonging lay behind intervention effects in other schools (Warren et al., 2020; Giddens, 1991). The fact that, in some schools, the intervention achieved benefits that were not mediated by increased student belonging does not suggest that the theory of human functioning and school organisation is wrong. Rather, it indicates that the intervention worked via mechanisms more varied than those initially theorised, but which were identified in the qualitative research and subsequently supported in these quantitative analyses.

Our analyses suggest that realist evaluations can be pursued within an RCT design and that such analyses can offer more nuanced evidence as to in which contexts interventions might effectively be implemented and how interventions might be tailored to potentiate the mechanisms that might be important to particular contexts. The Learning Together intervention appears likely to be effective in a range of schools. It may be that, in schools with lower capacity and higher baseline levels of bullying, the intervention might concentrate on delivering restorative practice whereas in schools with more capacity, more inclusive ethos and lower rates of baseline bullying, the intervention might instead concentrate on action groups to build student sense of belonging.

Crucially, the INCLUSIVE trial encompassed sufficient heterogeneity of school contexts to enable realist evaluation and provide data for our stratified analyses. The Learning Together intervention was open to local adaptation so that different mechanisms might ensue in different

schools, and this local adaptability was perfectly consistent with the RCT design as others have discussed, for example, in terms of fidelity of function (Hawe et al., 2004). Our analyses were of quantitative indicators but were not naively positivist (Bonell et al., 2018a); we recognised that these were imperfect empirical markers of underlying causal mechanisms which, though real, were not directly observable. Our use of a randomized design in fact strengthened our ability to undertake realist analyses. Random allocation provided better control of confounders, so that the 'signal' could be separated from the 'noise', which was important in the nuanced and potentially underpowered analyses we undertook.

Finally, our analyses suggest that the Learning Together intervention's focus on local data and local participative decision-making allowed it to promote health via a variety of mechanisms, with different schools benefiting from some mechanisms more than others. These realist analyses offer further evidence that whole-school interventions, such as Learning Together, offer a potent and flexible means of promoting young people's health.

CREDIT statement

Melendez-Torres: Conceptualisation; Methodology; Formal analysis; Writing (original draft). Warren: Conceptualisation; Writing (review & editing). Viner: Writing (review & editing); Funding acquisition. Allen: Methodology; Data curation; Writing (review & editing). Bonell: Conceptualisation; Methodology; Writing (original draft); Funding acquisition.

Acknowledgements

This study was funded by the National Institute for Health Research in England under its Public Health Research Board (12/153/60) and the Education Endowment Foundation. The views expressed in this publication are those of the authors and do not necessarily reflect those of the UK NHS, the National Institute for Health Research, or the Department of Health for England.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.socscimed.2021.113984>.

References

- Bond, L., Thomas, L., Coffey, C., Glover, S., Butler, H., Carlin, J.B., et al., 2004. Long-term impact of the Gatehouse project on cannabis use of 16-year-olds in Australia. *J. Sch. Health* 74 (1), 23–29.
- Bond, L., Wolfe, S., Tollit, M., Butler, H., Patton, G., 2007. A comparison of the Gatehouse Bullying Scale and the peer relations questionnaire for students in secondary school. *J. Sch. Health* 77 (2), 75–79.
- Bonell, C., Fletcher, A., Morton, M., Lorenc, T., 2012. Realist randomised controlled trials: a new approach to evaluating complex public health interventions. *Soc. Sci. Med.* 75 (12), 2299–2306.
- Bonell, C., Allen, E., Christie, D., Bourne, D., Fletcher, A., Grieve, R., et al., 2014. Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial. *Trials* 15, 381.
- Bonell, C., Warren, E., Fletcher, A., Viner, R., 2016a. Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al. *Trials* 17 (1), 478.
- Bonell, C., Shackleton, N., Fletcher, A., Jamal, F., Allen, E., Mathiot, A., et al., 2016b. Student- and school-level belonging and commitment and student smoking, drinking and misbehaviour. *Health Educ. J.* 76 (2), 206–220.
- Bonell, C., Moore, G., Warren, E., Moore, L., 2018a. Are randomized controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services. *Trials* 19, 238.
- Bonell, C., Allen, E., Opondo, C., Warren, E., Bourne, D., Sturgess, J., et al., 2019. Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention. *J. Epidemiol. Community Health* 73, 455–464. <https://doi.org/10.1136/jech-2018-211443>.
- Bonell, C., Allen, E., Warren, E., McGowan, J., Bevilacqua, L., Jamal, F., et al., 2018b. Initiating change in the school environment to reduce bullying and aggression: a cluster randomised controlled trial of the Learning Together (LT) intervention in English secondary schools. *Lancet* 392 (10163), 2452–2464.
- Buckley, S., Maxwell, G.M., 2007. Respectful Schools: Restorative Practices in Education: A Summary Report. Office of the Children's Commissioner and the Institute of Policy Studies, School of Government, Victoria University, Wellington.
- Clarke, A., Friede, T., Putz, R., Ashdown, J., Martin, S., Blake, A., et al., 2011. Warwick-edinburgh mental well-being scale (WEMWBS): validated for teenage school students in England and Scotland. A mixed methods assessment. *BMC Publ. Health* 11 (1), 487.
- Eart, L.M., Lee, L.E., 1998. Evaluation of the Manitoba School Improvement Program. University of Toronto, Toronto, ON.
- Epstein, J.L., McPartland, J.M., 1976. The concept and measurement of the quality of school life. *Am. Educ. Res. J.* 13 (1), 15–30.
- Giddens, A., 1991. *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Polity, Cambridge.
- Goodenow, C., 1993. Classroom belonging among early adolescent students: relationships to motivation and achievement. *J. Early Adolesc.* 13 (1), 21–43.
- Goodman, R., 2006. The Strengths and Difficulties questionnaire: a research note. *JCPP (J. Child Psychol. Psychiatry)* 38 (5), 581–586.
- Hawe, P., Shiell, A., Riley, T., 2004. Complex interventions: how 'out of control' can a randomised controlled trial be? *Br. Med. J.* 328, 1561–1563.
- Jamal, F., Fletcher, A., Shackleton, N., Bourne, D., Viner, R., Bonell, C., 2015. The three stages of building and testing mid-level theories in a realist RCT: a theoretical and methodological case-example. *Trials* 16 (1), 466.
- Kane, J.G., McCluskey, L.G., Riddell, S., Stead, J., Weedon, E., 2007. Restorative Practices in Scottish Schools. Scottish Executive, Edinburgh.
- Kokotsaki, D., White, C., Hopkins, B., 2014. Capturing change: a review of the implementation of Restorative Approaches and its outcomes within a local authority in North East England. *Online Educ. Res. J.* 5, 151.
- Marchal, B., Westhorp, G., Wong, G., Van Belle, S., Greenhalgh, T., Kegels, G., et al., 2013. Realist RCTs of complex interventions - an oxymoron. *Soc. Sci. Med.* 94, 124–128.
- Marcham, W.A., Aveyard, P., 2003. A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice. *Soc. Sci. Med.* 56 (6), 1209–1220.
- Pawson, R., Tilley, N., 1997. *Realistic Evaluation*. Sage.
- Pituch, K.A., Stapleton, L.M., 2012. Distinguishing between cross- and cluster-level mediation processes in the cluster randomized trial. *Socio. Methods Res.* 41 (4), 630–670.
- Raudenbush, S.W., Bryk, A.S., 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, second ed. Sage, Thousand Oaks, CA.
- Roeser, R.W., Midgley, C., Urdan, T.C., 1996. 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.* 88 (3), 408–422.
- Sawyer, M.G., Pfeiffer, S., Spence, S.H., Bond, L., Graetz, G., Kay, D., et al., 2010. School-based prevention of depression: a randomised controlled study of the beyond blue schools research initiative. *JCPP (J. Child Psychol. Psychiatry)* 51 (2), 199–209.
- Smith, J.D., Schneider, B.H., Smith, P.K., Ananiadou, K., 2004. The effectiveness of whole-school antibullying programs: a synthesis of evaluation research. *Sch. Psychol. Rev.* 33, 547.
- Van Belle, S., Wong, G., Westhorp, G., Pearson, M., Emmel, N., Manzano, A., et al., 2016. Can 'realist' randomised controlled trials be genuinely realist? *Trials* 17 (313).
- Vreeman, R.C., Carroll, A.E., 2007. A systematic review of school-based interventions to prevent bullying. *Arch. Pediatr. Adolesc. Med.* 161 (1), 78–88.
- Warren, E., Bevilacqua, L., Opondo, C., Allen, E., Mathiot, A., West, G., et al., 2019. Action groups as a participative strategy for leading whole-school health promotion: results on implementation from the INCLUSIVE trial in English secondary schools. *Br. Educ. Res. J.* 45 (5), 748–762.
- Warren, E., Melendez-Torres, G.J., Viner, R.M., Bonell, C.P., 2020. Using qualitative research within a realist trial to build theory about how context and mechanisms interact to generate outcomes: findings from the INCLUSIVE trial of a whole-school health intervention. *Trials* 21 (774).

Appendix 11: Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al.

Bonell et al. *Trials* (2016) 17:478
DOI 10.1186/s13063-016-1613-9

Trials

LETTER

Open Access



Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al.

Chris Bonell^{1*}, Emily Warren¹, Adam Fletcher² and Russell Viner³

Abstract

Background: Van Belle et al. argue that our attempt to pursue realist evaluation via a randomised trial will be fruitless because we misunderstand realist ontology (confusing intervention mechanisms with intervention activities and with statistical mediation analyses) and because RCTs cannot comprehensively examine how and why outcome patterns are caused by mechanisms triggered in specific contexts.

Methods: Through further consideration of our trial methods, we explain more fully how we believe complex social interventions work and what realist evaluation should aim to do within a trial.

Results: Like other realists, those undertaking realist trials assume that: social interventions provide resources which local actors may draw on in actions that can trigger mechanisms; these mechanisms may interact with contextual factors to generate outcomes; and data in the 'empirical' realm can be used to test hypotheses about mechanisms in the 'real' realm. Whether or not there is sufficient contextual diversity to test such hypotheses is a contingent not a necessary feature of trials. Previous exemplars of realist evaluation have compared empirical data from intervention and control groups to test hypotheses about real mechanisms. There is no inevitable reason why randomised trials should not also be able to do so. Random allocation merely ensures the comparability of such groups without necessarily causing evaluation to lapse from a realist into a 'positivist' or 'post-positivist' paradigm.

Conclusions: Realist trials are ontologically and epistemologically plausible. Further work is required to assess whether they are feasible and useful but such work should not be halted on spurious philosophical grounds.

Keywords: Randomized controlled trials, Realist evaluation, Scientific realism, Causation

Background

Van Belle et al. argue that our attempt to pursue realist evaluation via a randomised controlled trial (RCT) will be fruitless first, because we misunderstand realist ontology (confusing intervention mechanisms with intervention activities and with statistical mediation analyses) and second, because RCTs cannot comprehensively examine how and why outcome patterns are caused by mechanisms triggered in specific contexts [1]. We have found our ongoing debate with realist evaluators extremely useful in clarifying our thinking. In response to Van Belle et al.'s critique, we attempt to explain more fully how we believe

complex social interventions work and what realist evaluation should aim to do within a trial. Finally, we counter the argument that RCTs are inimical to realist enquiry.

Main text

A realist understanding of interventions and mechanisms

Let us first be clear about the nature of the 'Learning Together' intervention referred to in our earlier article [2] and how it is meant to work. We completely agree with Van Belle et al. that interventions comprise a series of resources. Learning Together aims to reduce bullying and aggression in secondary schools by providing schools with the following resources: (1) lesson plans and slides for a social and emotional skills curriculum; (2) a report of data on local student needs, a manual and an external facilitator; and (3) training sessions for staff

* Correspondence: Chris.Bonell@lshtm.ac.uk

¹London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK

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



© 2016 The Author(s). **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

on restorative practice. We offer these resources to schools with the hope that staff and students decide to use them in order to facilitate various actions such as: lessons on social/emotional skills; revisions to school policies and other locally decided school-level decisions; and restorative practice sessions in which staff and students respond to incidents of aggression and bullying. We theorise that through these actions various mechanisms may be triggered. We agree with Van Belle et al. that social interventions cannot introduce mechanisms directly but only via a process of participants acting on the resources provided. In the case of Learning Together, there are various intended mechanisms all involving the erosion of various 'boundaries' between and among staff and students, and between students' academic and broader development. We theorise that these boundaries will be eroded not by the intervention directly but via staff and students engaging in actions, supported by the intervention resources, to enhance relationships across the school and by reorienting school activities to focus on students' holistic development, which should include but not be limited to academic attainment.

Based on the theory of human functioning and school organisation [3] as well as on qualitative research on the school environment and young people's health [4], we theorised in our original article that the erosion of these boundaries will encourage more students, especially those from working class backgrounds, to feel committed to school, less committed to anti-school peer groups and less engaged in practices which go against school formal rules and informal norms, including bullying and aggression [2]. We further theorised that these mechanisms will be triggered and play out differently in different contexts. For example, we theorised that in schools where staff already give some priority to promoting students' overall wellbeing, the intervention activities are more likely to be implemented, so the intervention mechanisms, particularly those concerning the erosion of boundaries between students' academic and broader development, are more likely to be triggered. And we theorised that in schools with more working class students, the erosion of the boundaries will lead to a higher proportion of students becoming committed to school (because we theorise that the aforementioned boundaries in particular hamper the educational engagement of working class students). We hope this fuller description of our intervention and its mechanisms reassures readers that our understanding of the Learning Together intervention is compatible with realist ontology and realist evaluation practice.

A realist understanding of empirical evaluation

We continue to contend that our approach to research in general and evaluation in particular, is realist in

orientation. Realists suggest there is an 'empirical' realm consisting of the data researchers collect and analyse. This empirical realm provides a window, albeit an indirect one, on an 'actual' realm of occurrences apparent to participants. This actual realm in turn reflects a 'real' realm made up of structural mechanisms which are unobservable but which are the causes of the actual and empirical realms [5]. In terms of epistemology, realists believe they can identify objective truths describing the actual realm and can uncover the true causal mechanisms of the real realm based upon data from the empirical realm. Like Van Belle et al., we believe that even if intervention mechanisms do occur, they will not be directly observable. Boundaries and commitment to school lie in the realm of the real; they cause observable phenomena but are not themselves observable.

The activities our intervention is intended to promote are, in critical realist parlance, in the actual realm, as are the outcomes we hope will arise as a result of the intervention. Staff and students will be able to observe intervention activities, such as restorative sessions, as well as the behaviours, such as bullying, that the intervention is aiming to reduce. However, the data collected in the course of the RCT of Learning Together (like data collected in any form of research) are not a direct and unproblematic window into this actual realm. Our outcome evaluation is not collecting data on bullying or aggression directly. Rather, it is collecting data on student answers to questionnaires asking about their experiences of these practices. Similarly, our process evaluation is not collecting data on activities such as policy review and restorative sessions directly. Rather, it collects data in the form of notes that researchers make when they observe these sessions or in the form of the accounts of teachers or students when they are interviewed about their experiences of these activities. In critical realist terms, these data are in the realm of the empirical. We appreciate that all sorts of factors might mean that these data do not provide a full or unproblematic representation of events in the realm of the actual. Nonetheless, they should provide some guide as to what is happening.

Based on our theorising about how mechanisms (in the realm of the real) interact with context to produce outcomes (in the realm of the actual), we have hypothesised that in statistical analyses of outcome indicators (in the realm of the empirical), students in the schools randomly selected to implement the intervention will report less bullying and aggression than students in schools randomly selected to be controls [2]. We also hypothesised that in mediation analyses, the association between trial arm and the empirical indicators of bullying and aggression will be reduced by adjustment for indicators of increased student commitment to school. And we further hypothesised that in statistical moderation analyses,

baseline school-level aggregate indicators of staff priorities and school-level indicators of student-reported socio-economic status will moderate the association found between intervention arm and our measures of bullying and aggression.

The above are examples of how we intend to use statistical analyses (in the realm of the empirical) to test hypotheses about how, in the realm of the real, context and mechanisms interact to generate outcomes (known as context-mechanism-outcome or 'CMO configurations'). Van Belle et al. say very little about why realist RCTs are unlikely to be able to empirically test hypotheses about CMO. They assert, without reference to evidence or further argument, that "Given the need for randomisation and control in an RCT, only relatively few and simple CMO configurations can be tested at a time." We disagree. We aim to undertake analyses to test multiple CMOs including but not limited to those above. Some of these are based a priori on theory while others have been and will continue to be developed based on qualitative research within the trial. Testing these should be perfectly possible within an RCT. These analyses should help us gain a more vivid and nuanced understanding of how and why reductions in bullying and aggression are caused by mechanisms of boundary erosion and school engagement which are triggered and play out differently in diverse school contexts. But whether this is the case or not is ultimately an empirical question. It cannot be judged until we complete our analyses. We see nothing about random allocation which impede such analyses. Indeed, doing such analyses within an RCT has the crucial advantage of control of confounding. This is so important in public health because even important interventions will often have quite small effects for each individual participant, which can be impossible to distinguish from other confounding influences [6]. We agree that our ability to assess CMO configurations would be undermined if trials contained insufficient variety in characteristics of place or person because of excessively tight inclusion criteria for sampling clusters or individuals. But while this may sometimes occur in trials, it is not a necessary feature, particularly of pragmatic effectiveness trials. The other obvious impediments to the proposed analyses are measurement error and lack of statistical power. These are real challenges but they are in no way necessary or particular features of trials as opposed to any other designs.

Furthermore, we would like to stress that developing and empirically testing such hypotheses does not mean, as Van Belle et al. seem to suggest, that we are confusing the empirical with the real, reducing the causal mechanism of our intervention to statistical mediation analyses or reducing the context of the intervention to statistical moderation analyses. We are using crude and indirect

quantitative data (which exist in the realm of the empirical) as a way of indirectly testing whether our theories about mechanisms (which exist in the realm of the real) might be correct. Moreover, we are also using qualitative research to deepen our understanding of how mechanisms might work and using this to refine our theories and hypotheses. Like our quantitative research, our qualitative research examines empirical data (this time in the form of student and staff accounts) as an indirect window on the actual (participants' experiences) and the real (how mechanisms unfold through interactions between individual agency and social structures) realms.

And if our analyses do not support the above hypotheses, this will not mean we immediately conclude that the theorized mechanisms do not exist. Null results could indicate that the context causes the mechanisms to remain unactivated (for example, schools do not implement the intervention or implementation fails to trigger an erosion of boundaries) or the mechanism is activated but counteracted by other mechanisms (for example, government initiatives cause schools to buttress the boundaries between students' academic and broader development). We should stress that this approach is consistent with existing philosophy about how to interpret null results from RCTs of social interventions [7].

What is so bad about randomisation?

Although we think randomisation is extremely important technically in enabling us to assess statistical associations while minimizing bias, we think randomisation is philosophically trivial. We disagree that the use of randomisation will inevitably lead to our research lapsing from a realist into a positivist or post-positivist paradigm [8]. Randomisation ensures that the different groups we compare resemble each other as closely as possible in terms of all the characteristics (except from exposure to intervention resources) likely to affect outcomes, whether we know about these in advance or not. If we were to evaluate Learning Together by comparing rates of reported bullying and aggression between schools which chose to implement the intervention versus those which did not, there is a very strong likelihood that any reductions in indicators of bullying and aggression would wholly or partly (we would not know which) reflect baseline differences in the institutions and/or the individuals within them. Randomisation is merely a practical tool to reduce confounding. It does not fundamentally change the nature of the way we view or research the social world, or affect how we will use comparative empirical data to test hypotheses about mechanisms.

Proponents of non-randomised realist evaluation in fact often refer to the use of quantitative empirical data from different groups as exemplars of how to test hypotheses about intervention mechanisms. For example,

Pawson and Tilley refer to an evaluation of prisoner education as a means of reducing recidivism, in which evaluators compared rates of reoffending in intervention sites with expected rates. The latter were generated from historical 'usual treatment' data. Van Belle et al. do not attempt to explain why such non-random comparisons of quantitative data are legitimately realist whereas our comparisons of data from randomly allocated groups are not [9].

It is true that the originator of critical realism objected to the use of RCTs in social as opposed to natural scientific enquiry [10] but this betrayed his misunderstanding of how natural and social scientific research is done. Bhaskar argued that experimental manipulation is used to create closed systems to neutralise external forces and so isolate the mechanisms being tested. He suggests that experimental manipulation in social science is impossible because social systems are open [5]. We agree that experiments in the physical sciences do often try to control and isolate causal factors. However, experiments in the biological sciences often take a different approach, because in biology many systems (such as ecosystems and bodily systems) cannot be closed. Many RCTs in the fields of environmental science and clinical pharmacology do not use randomisation to remove all other mechanisms in order to isolate the mechanism under investigation. Rather, randomisation is used, in effect, to hold these other mechanisms constant so that the mechanism under investigation can be viewed in their full context, i.e. to understand its impact alongside these other mechanisms. In other words, these trials measure 'added value' not 'separated value'. The same principle applies to RCTs of social interventions. And in fact the same principle also applies to non-randomised comparative evaluations such as the realist review of prisoner education mentioned above. These designs use comparison groups to look for the added value of the intervention mechanisms against the backdrop of other influential mechanisms. The fact that these comparison groups are not assembled via random allocation does not alter this fact.

We would strongly agree with realists that RCTs are quite impractical for examining many questions in the social sciences - including perhaps the most interesting and important questions, such as what mechanisms account for the maintenance of class inequality [11] or secular reductions in violence [12]. It would obviously be impractical to randomly allocate people to different social classes or historical eras. And we would also acknowledge that for many of the most important public health interventions, control groups and/or random allocation are impractical, for example in the evaluation of the health impacts of smoking bans, alcohol taxes or seatbelt enforcement [13]. But in other cases, it is perfectly feasible to use RCTs to investigate public health

interventions and, in such cases, randomised designs should not be rejected based on dogma.

Some evaluations of social interventions draw on naturally occurring random allocation. For example, in the USA researchers have examined whether educational attainment is higher in charter schools than community schools [14]. Here, evaluation is facilitated by the fact that some oversubscribed charter schools determine entry by random ballot. The evaluations done so far have not been realist and have not assessed CMO configurations. But it is possible to imagine that naturally randomised experiments could assess how context interact with mechanisms to generate outcomes. Would the opponents of realist trials argue that in order to do so, evaluators would have to find other, non-random comparisons to avoid the taint of positivism?

Conclusions

Realist evaluation focuses on developing, refining and testing theories about how interventions provide resources which participants use to trigger mechanisms that interact with context to generate outcomes. This is extremely useful both in emphasising that evaluation studies should focus on the testing and refinement of intervention theory (rather than merely accrediting particular interventions as effective or not). It is also extremely helpful in providing a basis for understanding the importance of context, and for drawing on empirical evidence to consider how context might affect the implementation and effects of interventions in new settings. Realist evaluation has sometimes been described as methods neutral [15] and social science more generally as methodologically pluralist [16]. Indeed, Van Belle et al. state that realist evaluation should use "whatever data and analytic methods [are] appropriate to build, support, refute or refine plausible explanations that incorporate intervention, actors, outcomes, context and mechanisms". We hope to persuade realists that RCTs have a place in this analytic panoply.

Abbreviations

CMO: Context-mechanism-outcome; RCT: Randomised controlled trials

Acknowledgements

We would like to thank the students and staff of the schools participating in the study as well as the team delivering the intervention and the wider team of researchers. This project is funded by a grant from the UK National Institute for Health Research Public Health Research programme (grant 12/153/60). This report presents independent research commissioned by the UK 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 National Health Service (NHS), the NIHR, Medical Research Council (MRC), Central Commissioning Facility (CCF), NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC), the Public Health Research programme or the English Department of Health. The work was undertaken with the support of The Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement (DECIPHer), a UK Clinical Research Collaboration (UKCRC) Public Health Research Centre of Excellence. Joint funding (MR/K023231/1) from the British Heart Foundation, Cancer Research

UK Economic and Social Research Council, Medical Research Council, the Welsh Government and the Wellcome Trust, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.
Dedicated to the memory of Farah Jamal, 1986–2016.

Funding

There was no funding for this article. The trial is funded by a grant from the UK National Institute for Health Research Public Health Research programme (grant 12/153/60).

Availability of data and materials

Not applicable since this article reports no empirical findings.

Authors' contributions

CB wrote and revised the manuscript. EW, AF, and RV contributed to and edited the manuscript. All authors have read and approve of the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable since this article reports no empirical findings.

Ethics approval and consent to participate

Not applicable since this article reports no empirical findings.

Author details

¹London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK. ²Cardiff University, Glamorgan Building, King Edward VII, Cardiff CF10 3WT, UK. ³University College London Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK

Received: 28 July 2016 Accepted: 18 September 2016

Published online: 01 October 2016

References

- Van Belle S, Wong G, Westhorp G, Pearson M, Emmel N, Manzano A, Marchal B. Can "realist" randomised controlled trials be genuinely realist? *Trials*. 2016;17:313.
- Jamal F, Fletcher A, Shackleton A, Elbourne D, Viner R, Bonell C. The three stages of building and testing mid-level theories in a realist RCT: a case-example. *Trials*. 2015;16:466.
- 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.
- Jamal F, Fletcher A, Harden A, Wells H, Thomas J, Bonell C. The school environment and student health: a meta-ethnography of qualitative research. *BMC Public Health*. 2013;13:798.
- Bhaskar R. *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*. 2nd ed. Hemel Hempstead: Harvester Wheatsheaf; 1989.
- Cousens S, Hargreaves J, Bonell C, Armstrong B, Thomas J, Kirkwood BR, Hayes R. Alternatives to randomisation in the evaluation of public-health interventions: statistical analysis and causal inference. *J Epidemiol Community Health*. 2011;65:576–81.
- Cartwright N, Hardie J. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press; 2012.
- Bonell C, Fletcher A, Morton M, Lorenc T, Moore L. Methods don't make assumptions, researchers do: a response to Marchal et al. *Soc Sci Med*. 2013;94:81–2.
- Pawson R, Tilley N. *Realistic Evaluation*. London: Sage; 1997.
- Bhaskar R. *A Realist Theory of Science*. Leeds: Leeds Books; 1975.
- Bourdieu P, Wacquant L. *An Invitation to Reflexive Sociology*. Chicago: University of Chicago Press; 1992.
- Elias N. *The Civilizing Process. Sociogenetic and Psychogenetic Investigations*. Oxford: Basil Blackwell; 2000.
- Bonell CP, Hargreaves J, Cousens S, Ross D, Hayes R, Petticrew M, Kirkwood BR. Alternatives to randomisation in the evaluation of public health interventions: design challenges and solutions. *J Epidemiol Community Health*. 2011;65:582–7.
- Gleason P, Clark M, Clark Tuttle C, Dwoyer E, Silverberg M. *The Evaluation of Charter School Impacts. Final Report*. Washington DC: US Department of Education; 2010.
- Marchal B, Westhorp G, Wong G, Van Belle S, Greenhalgh T, Kegels G, Pawson R. Realist RCTs of complex interventions - an oxymoron. *Soc Sci Med*. 2013;94:124–8.
- Payne G, Williams M, Chamberlain S. Methodological pluralism in British sociology. *Sociology*. 2004;38:153–63.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



Appendix 12: Online appendices related to the QCA

ONLINE APPENDICES

Table 1: Truth table for overarching mechanisms

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

0	0	1	0	1	0	1	0.9109309	3	1 (13)
0	0	0	0	1	0	1	0.9565519	1	2 (36, 21)
0	0	1	1	0	0	1	0.9586992	9	1 (2)
0	1	0	1	1	0	1	0.9895403	4	1 (14)
0	1	0	1	0	1	1	0.9954019	8	1 (16)
0	1	0	0	1	0	1	0.9962664	2	2 (8, 37)
1	1	1	0	0	0	1	0.9964729	5	1 (27)
1	0	1	0	1	0	1	0.9991031	3	1 (10)
1	0	0	0	1	0	1	0.9991276	3	1 (34)
0	0	0	0	1	1	1		1	1 (3)
0	0	1	1	1	0	1		1	1(5, 11)
0	1	1	1	1	1	1		1	1 (32)
1	0	1	0	1	1	1		1	1 (28)
1	1	0	0	1	1	1		1	2 (25, 1)

1 1 | 1 0 1 | 1 | 1 | 1 1 (31)

Effective solutions in bold; contradictions indicated with *

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

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

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

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

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

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

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

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

Effective solutions in **bold**

Appendix 13: Copyrights for included publications

Appendix 13.1: Proof of copyright for “Are randomised controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services”

1/18/22, 7:27 AM

Rightslink® by Copyright Clearance Center



?
Help

Live Chat

Are randomised controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services

SPRINGER NATURE

Author: Chris Bonell et al

Publication: Trials

Publisher: Springer Nature

Date: Apr 19, 2018

Copyright © 2018, The Author(s).

Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

CC0 applies for supplementary material related to this article and attribution is not required.

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Terms and Conditions](#)
Comments? We would like to hear from you. E-mail us at customer@copyright.com

Appendix 13.2: Proof of copyright for “Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial”

1/21/22, 1:51 PM

Rightslink® by Copyright Clearance Center



- Home
- Help ▾
- Live Chat
- Sign in
- Create Account

Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): study protocol for a cluster randomised controlled trial



Author: Chris Bonell et al
Publication: Trials
Publisher: Springer Nature
Date: Sep 30, 2014

Copyright © 2014, Bonell et al.; licensee BioMed Central Ltd.

Creative Commons

The request you have made is considered to be non-commercial/educational. As the article you have requested has been distributed under a Creative Commons license (Attribution-Noncommercial), you may reuse this material for non-commercial/educational purposes without obtaining additional permission from Springer Nature, providing that the author and the original source of publication are fully acknowledged (please see the article itself for the license version number). You may reuse this material without obtaining permission from Springer Nature, providing that the author and the original source of publication are fully acknowledged, as per the terms of the license. For license terms, please see <http://creativecommons.org/>

BACK

CLOSE WINDOW

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Terms and Conditions](#)
Comments? We would like to hear from you. E-mail us at customer care@copyright.com

Appendix 13.3: Proof of copyright for “Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): update to a cluster randomised controlled trial protocol”

1/18/22, 7:28 AM

Rightslink® by Copyright Clearance Center



?
Help ▾

🗨️
Live Chat

SPRINGER NATURE

Initiating change locally in bullying and aggression through the school environment (INCLUSIVE) trial: update to cluster randomised controlled trial protocol

Author: Chris Bonell et al

Publication: Trials

Publisher: Springer Nature

Date: May 25, 2017

Copyright © 2017, The Author(s).

Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

CC0 applies for supplementary material related to this article and attribution is not required.

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Terms and Conditions](#)
Comments? We would like to hear from you. E-mail us at customer-care@copyright.com

Appendix 13.4: Proof of copyright for “Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial”

1/21/22, 1:53 PM

Rightslink® by Copyright Clearance Center



- Home
- Help
- Live Chat
- Sign in
- Create Account

Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial

THE LANCET

Author: Chris Bonell, Elizabeth Allen, Emily Warren, Jennifer McGowan, Leonardo Bevilacqua, Farah Jamal, Rosa Legood, Meg Wiggins, Charles Opondo, Anne Mathiot, Jo Sturgess, Adam Fletcher, Zia Sadique, Diana Elbourne, Deborah Christie, Lyndal Bond, Stephen Scott et al.

Publication: The Lancet

Publisher: Elsevier

Date: 8–14 December 2018

© 2018 The Author(s). Published by Elsevier Ltd.

Journal Author Rights

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK CLOSE WINDOW

© 2022 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | [Privacy statement](#) | [Terms and Conditions](#)
Comments? We would like to hear from you. E-mail us at customer care@copyright.com

Appendix 13.5: Proof of copyright for “Modifying the secondary school environment to reduce bullying and aggression: the INCLUSIVE cluster RCT”

1/18/22, 7:23 AM

Non Commercial Government Licence



Non-Commercial Government Licence for public sector information

delivered by



[Back to The National Archives](#)

You are encouraged to use and re-use the Information that is available under this licence freely and flexibly, with only a few conditions.

Using information under this licence

Use of copyright and database right material expressly made available under this licence (the 'Information') indicates your acceptance of the terms and conditions below.

The Licensor grants you a worldwide, royalty-free, perpetual, non-exclusive licence to use the Information for Non-Commercial purposes only subject to the conditions below.

This licence does not affect your freedom under fair dealing or fair use or any other copyright or database right exceptions and limitations.

You are free to:

copy, publish, distribute and transmit the Information;

adapt the Information;

exploit the Information for Non-Commercial purposes for example, by combining it with other information in your own product or application.

You are not permitted to:

exercise any of the rights granted to you by this licence in any manner that is primarily intended for or directed toward commercial advantage or private monetary compensation.

You must, where you do any of the above:

acknowledge the source of the Information by including any attribution statement specified by the Information Provider(s) and, where possible, provide a link to this licence;

If the Information Provider does not provide a specific attribution statement, you must use the following:

<https://www.nationalarchives.gov.uk/doc/non-commercial-government-licence/version/2/>

1/4

Contains information licensed under the Non-Commercial Government Licence v2.0.

If you are using Information from several Information Providers and listing multiple attributions is not practical in your product or application, you may include a URI or hyperlink to a resource that contains the required attribution statements.

ensure that any onward licensing of the Information – for example when combined with other information – is for Non-Commercial purposes only.

These are important conditions of this licence and if you fail to comply with them or use the Information other than for Non-Commercial purposes the rights granted to you under this licence, or any similar licence granted by the Licensor, will end automatically.



Exemptions

This licence does not cover the use of:

- personal data in the Information;
- Information that has not been accessed by way of publication or disclosure under information access legislation (including the Freedom of Information Acts for the UK and Scotland) by or with the consent of the Information Provider;
- departmental or public sector organisation logos, crests, military insignia and the Royal Arms except where they form an integral part of a document or dataset;
- military insignia
- third party rights the Information Provider is not authorised to license;
- other intellectual property rights, including patents, trade marks, and design rights; and
- identity documents such as the British Passport.

Non-endorsement

This licence does not grant you any right to use the Information in a way that suggests any official status or that the Information Provider and/or Licensor endorse you or your use of the Information.

No warranty

The Information is licensed 'as is' and the Information Provider excludes all representations, warranties, obligations and liabilities in relation to the Information to the maximum extent permitted by law.

The Information Provider is not liable for any errors or omissions in the Information and shall not be liable for any loss, injury or damage of any kind caused by its use. The Information Provider does not guarantee the continued supply of the Information.

Governing Law

This licence is governed by the laws of the jurisdiction in which the Information Provider has its principal place of business, unless otherwise specified by the Information Provider.

Definitions

In this licence the terms below have the following meanings:

'Information'

means information protected by copyright or by database right (for example, literary and artistic works, content, data and source code) offered for use under the terms of this licence.

'Information Provider'

means the person or organisation providing the Information under this licence.

'Licensor'

means any Information Provider which has the authority to offer Information under the terms of this licence or the Keeper of the Public Records, who has the authority to offer Information subject to Crown copyright and Crown database rights and Information subject to copyright and database right that has been assigned to or acquired by the Crown, under the terms of this licence.

'Non-Commercial purposes'

means not intended for or directed toward commercial advantage or private monetary compensation. For the purposes of this licence, 'private monetary compensation' does not include the exchange of the Information for other copyrighted works by means of digital file-sharing or otherwise provided there is no payment of any monetary compensation in connection with the exchange of the Information.

'Use'

as a verb, means doing any act which is restricted by copyright or database right, whether in the original medium or in any other medium, and includes without limitation distributing, copying, adapting, modifying as may be technically necessary to use it in a different mode or format.

'You'

means the natural or legal person, or body of persons corporate or incorporate, acquiring rights under this licence.

About the Non-Commercial Government Licence

The National Archives has developed this licence as a tool to enable Information Providers in the public sector to license the use and re-use of their Information under a common non-commercial licence. The National Archives invites public sector bodies owning their own copyright and database rights to permit the use of their Information under this licence where licensing under the default Open Government Licence is not appropriate.

The Keeper of the Public Records has authority to license Information subject to copyright and database right owned by the Crown.

This is version 2.0 of the Non-Commercial Government Licence. The National Archives may, from time to time, issue new versions of the Non-Commercial Government Licence. If you are already using information under a previous version of the Non-Commercial Government Licence, the terms of that licence will continue to apply.

Further context, best practice and guidance can be found in the UK Government Licensing Framework section on [The National Archives website](#).

[Go to the Welsh version of the licence.](#)

[Go to the version 1 of the licence.](#)

Appendix 13.6: Proof of copyright for “Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention”

1/18/22, 7:32 AM

Rightslink® by Copyright Clearance Center



?
Help ▾

🗨️
Live Chat



Examining intervention mechanisms of action using mediation analysis within a randomised trial of a whole-school health intervention

Author:

Chris Bonell, Elizabeth Allen, Charles Opondo, Emily Warren, Diana Ruth Elbourne, Joanna Sturgess, Leonardo Bevilacqua, Jennifer McGowan, Anne Mathiot, Russell M Viner

Publication: Journal of Epidemiology & Community Health

Publisher: BMJ Publishing Group Ltd.

Date: May 1, 2019

Copyright © 2019, BMJ Publishing Group Ltd. All rights reserved.

Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY 4.0](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Terms and Conditions](#)
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

[https://s100.copyright.com/AppDispatchServlet?publisherName=BMJ&publication=jech&oa=CC-BY-4.0&title=Examining intervention mechanisms of a...](https://s100.copyright.com/AppDispatchServlet?publisherName=BMJ&publication=jech&oa=CC-BY-4.0&title=Examining%20intervention%20mechanisms%20of%20a...) 1/1

Appendix 13.7: Proof of copyright for “Moderated mediation analyses to assess intervention mechanisms for impacts on victimisation, psycho-social problems and mental wellbeing: Evidence from the INCLUSIVE realist randomized trial”

1/21/22, 1:54 PM

Rightslink® by Copyright Clearance Center



- Home
- Help
- Live Chat
- Sign in
- Create Account



Moderated mediation analyses to assess intervention mechanisms for impacts on victimisation, psycho-social problems and mental wellbeing: Evidence from the INCLUSIVE realist randomized trial

Author: G.J. Melendez-Torres, Emily Warren, Russell Viner, Elizabeth Allen, Chris Bonell

Publication: Social Science & Medicine

Publisher: Elsevier

Date: June 2021

© 2021 Elsevier Ltd. All rights reserved.

Journal Author Rights

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW

© 2022 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | [Privacy statement](#) | [Terms and Conditions](#)
Comments? We would like to hear from you. E-mail us at customer-care@copyright.com

Appendix 13.8: Proof of copyright for “Action groups as a participative strategy for leading whole-school health promotion: Results on implementation from the INCLUSIVE trial in English secondary schools”

12/20/21, 5:35 PM

RightsLink Printable License

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Dec 20, 2021

This Agreement between London School of Hygiene & Tropical Medicine -- Emily Warren ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number 5213150987918

License date Dec 20, 2021

Licensed Content Publisher John Wiley and Sons

Licensed Content Publication British Educational Research Journal

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

Licensed Content Author Chris Bonell, Russell Viner, Farah Jamal, et al

Licensed Content Date Jul 12, 2019

Licensed Content Volume 45

Licensed 5

<https://s100.copyright.com/AppDispatchServlet>

1/6

Content
Issue

Licensed
Content 22
Pages

Type of use Dissertation/Thesis

Requestor
type Author of this Wiley article

Format Print and electronic

Portion Full article

Will you be
translating? No

Title Are realist randomised controlled trials possible? INCLUSIVE as a case study
of an emerging methodology

Institution
name London School of Hygiene & Tropical Medicine

Expected
presentation
date Apr 2022

London School of Hygiene & Tropical Medicine
Keppel Street

Requestor
Location London, WC1E 7HT
United Kingdom
Attn: Emily Warren

Publisher
Tax ID EU826007151

Total 0.00 GBP

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at <http://myaccount.copyright.com>).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order**, is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.
- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. **For STM Signatory Publishers clearing permission under the terms of the [STM Permissions Guidelines](#) only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts**, You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or

their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.
- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of

any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.
- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The [Creative Commons Attribution License \(CC-BY\)](#) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The [Creative Commons Attribution Non-Commercial \(CC-BY-NC\) License](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The [Creative Commons Attribution Non-Commercial-NoDerivs License](#) (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
<http://olabout.wiley.com/WileyCDA/Section/id-410895.html>

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.



[Appendix 13.9: Proof of copyright for “Using qualitative research to explore intervention mechanisms: findings from the trial of the Learning Together whole-school health intervention”](#)

1/18/22, 7:26 AM

Rightslink® by Copyright Clearance Center



?
Help ▾

🗨️
Live Chat

SPRINGER NATURE

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

Author: Emily Warren et al

Publication: Trials

Publisher: Springer Nature

Date: Sep 10, 2020

Copyright © 2020, The Author(s)

Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

CC0 applies for supplementary material related to this article and attribution is not required.

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Terms and Conditions](#)
Comments? We would like to hear from you. E-mail us at customer@copyright.com

Appendix 13.10: Proof of copyright for “Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al.”

1/18/22, 7:35 AM

Rightslink® by Copyright Clearance Center



?
Help ▾

🗨️
Live Chat

Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al.

SPRINGER NATURE

Author: Chris Bonell et al

Publication: Trials

Publisher: Springer Nature

Date: Oct 1, 2016

Copyright © 2016, The Author(s).

Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

CC0 applies for supplementary material related to this article and attribution is not required.

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Terms and Conditions](#)
Comments? We would like to hear from you. E-mail us at customer care@copyright.com

[https://s100.copyright.com/AppDispatchServlet?title=Realist trials and the testing of context-mechanism-outcome configurations%3A a response to Va...](https://s100.copyright.com/AppDispatchServlet?title=Realist%20trials%20and%20the%20testing%20of%20context-mechanism-outcome%20configurations%3A%20a%20response%20to%20Van%20Belle%20et%20al) 1/1

