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.(1) 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 like 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, some describing realist trials as an oxymoron.(1, 2) 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 epistemologically positivist and thus inimical to realist enquiry.(1-3) 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.(4-6)

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 reports 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,(7) 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 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.(8, 9) 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.(10-12)

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-hour 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; nd acurriculum materials for social and emotional skills lessons. School staff and students drew on these resources to implement various activities. Action groups comprised staff and students and 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 hours 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.(13)

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(14) 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; 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(15)) 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.(16) 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.(17, 18) 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.(18)

Informed by this qualitative research, we hypothesised that, in high capacity/inclusive ethos schools not faced by 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 Figure 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?

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.(14, 19) 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 the 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.(15) School belonging is measured using an eight-item subscale of the Beyond Blue School Climate Questionnaire with a four-step Likert scale averaged to construct a score (see Supplementary File 1 for items and scoring). This measure was developed in Australia (15) using questions from the Gatehouse,(20) Quality of School Life,(21) Patterns of Adaptive Learning,(22) Manitoba School Improvement Survey(23) and Psychological Sense of School Membership(24) instruments. Cronbach's alpha for the subscale of .85 was reported for the original Australian adolescent sample (personal communication, Lyndal Bond, 21 July 2011) and of 0.80 for the present study sample.(25)

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);(26) 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;(27) 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.(28) 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 'excellent' Ofsted ratings for leadership and administration at 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(15)) 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,(29) 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, convolving 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,(30) 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 Figure 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 outcome 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.

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 8,179 students, of whom 4,082 were in control schools and 4,097 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 (β=-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 (β=-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 (χ^2=31.900, *df*=4, *p*<0.0001).

Evidence for a similar pattern was found for schools that were below 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 (χ^2=12.486, *df*=4, *p*=0.014) and a significant indirect (β=-0.122, 95% CI [-0.214, -0.034]).

While school inclusivity at baseline did not moderate the mediational pathway through belonging to victimisation (χ^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 (β=-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.

Discussion

Summary of findings. We were able to draw on our prior qualitative research to develop hypotheses (see Figure 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 victimization, psychological problems and mental wellbeing. However, 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 were 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 Figure 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 hypothesize 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 Figure 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', reflecting 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.

Implications for research and policy. These findings largely support the hypotheses informed by our qualitative research and summarised in Figure 1.(17, 18) 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.(13) 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.(18, 31) 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 but merely 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, intervention might concentrate on delivering restorative practice whereas in schools with more capacity, more inclusive ethos and lower rates of baseline bullying, 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 previous discussed for example in terms of fidelity of function.(32) Our analyses were of quantitative indicators but were not naively positivist;(5) we recognised that these were imperfect empirical markers of underlying causal mechanisms which, though real, were not directly observable. Our use of a randomised 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.

References

1. Pawson R, Tilley N. Realistic evaluation: Sage; 1997.
2. Marchal B, Westhorp G, Wong G, Van Belle S, Greenhalgh T, Kegels G, et al. Realist RCTs of complex interventions - an oxymoron. Social Science and Medicine. 2013;94:124-8.
3. Van Belle S, Wong G, Westhorp G, Pearson M, Emmel N, Manzano A, et al. Can "realist" randomised controlled trials be genuinely realist? Trials. 2016;17(313).
4. Bonell C, Fletcher A, Morton M, Lorenc T. 'Realist Randomised Controlled Trials': a new approach to evaluating complex public health interventions. Social Science and Medicine. 2012;75(12):2299-306.
5. Bonell C, Moore G, Warren E, Moore L. 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. 2018;19:238.
6. Bonell C, Warren E, Fletcher A, Viner R. Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al. Trials. 2016;17(1):478.
7. Jamal F, Fletcher A, Shackleton N, Elbourne D, Viner R, Bonell C. The three stages of building and testing mid-level theories in a realist RCT: a theoretical and methodological case-example. Trials. 2015;16(1):466.
8. Smith JD, Schneider BH, Smith PK, Ananiadou K. The effectiveness of whole-school antibullying programs: A synthesis of evaluation research. School Psychology Review. 2004;33:547.
9. Vreeman RC, Carroll AE. A systematic review of school-based interventions to prevent bullying. Archives of Pediatrics & Adolescent Medicine. 2007;161(1):78-88.
10. Buckley S, Maxwell GM. Respectful schools: Restorative practices in education: A summary report. Wellington: Office of the Children's Commissioner and the Institute of Policy Studies, School of Government, Victoria University; 2007.
11. Kane JG, McCluskey LG, Riddell S, Stead J, Weedon E. Restorative Practices in Scottish Schools. Edinburgh: Scottish Executive; 2007.
12. 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 Educational Research Journal. 2014;5:151.
13. Markham WA, Aveyard P. A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice. Social Science & Medicine. 2003;56(6):1209-20.
14. Bonell C, Allen E, Warren E, McGowan J, Bevilacqua L, Jamal F, et al. Initiating change in the school environment to reduce bullying and aggression: a cluster randomised controlled trial of the Learning Together (LT) intervention in English secondary schools. The Lancet. 2018;392(10163):2452-64.
15. Sawyer MG, Pfeiffer S, Spence SH, Bond L, Graetz G, Kay D, et al. School-based prevention of depression: a randomised controlled study of the beyond blue schools research initiative. Journal of Child Psychology and Psychiatry. 2010;51(2):199-209.
16. 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. Journal of Epidemiology and Community Health http://dxdoiorg/101136/jech-2018-211443. 2019.
17. Warren E, Bevilacqua L, Opondo C, Allen E, Mathiot A, West G, et al. Action groups as a participative strategy for leading whole-school health promotion: results on implementation from the INCLUSIVE trial in English secondary schools. British Education Research Journal. 2019;45(5):748-62.
18. Warren E, Melendez-Torres GJ, Viner RM, Bonell CP. 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. 2020;21(774).
19. 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.
20. Bond L, Thomas L, Coffey C, Glover S, Butler H, Carlin JB, et al. Long-Term Impact of the Gatehouse Project on Cannabis Use of 16-Year-Olds in Australia. Journal of School Health. 2004;74(1):23-9.
21. Epstein JL, McPartland JM. The Concept and Measurement of the Quality of School Life. American Educational Research Journal. 1976;13(1):15-30.
22. Roeser RW, Midgley C, Urdan TC. Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. Journal of Educational Psychology. 1996;88(3):408-22.
23. Earl LM, Lee LE. Evaluation of the Manitoba School Improvement Program. Toronto, ON: University of Toronto; 1998.
24. Goodenow C. Classroom Belonging among Early Adolescent Students: Relationships to Motivation and Achievement. The Journal of Early Adolescence. 1993;13(1):21-43.
25. 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 Education Journal. 2016;76(2):206-20.
26. 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. Journal of School Health. 2007;77(2):75-9.
27. Goodman R. The Strengths and Difficulties Questionnaire: a research note. Journal of Child Psychology and Psychiatry. 2006;38(5):581-6.
28. 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(1):487.
29. Pituch KA, Stapleton LM. Distinguishing between cross- and cluster-level mediation processes in the cluster randomized trial. Sociological Methods & Research. 2012;41(4):630–70.
30. Raudenbush SW, Bryk AS. Hierarchical Linear Models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage; 2002.
31. Giddens A. Modernity and Self Identity: Self and Society in the Late Modern Age. Cambridge: Polity; 1991.
32. Hawe P, Shiell A, Riley T. Complex interventions: how "out of control" can a randomised controlled trial be? British Med Journal. 2004;328:1561-3.

Figure captions

Figure 1. Key hypotheses suggested by qualitative research.

Figure 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)			-0.191 (-0.305, -0.087)		-0.122 (-0.214, -0.034)		-0.380 (-0.676, -0.085)
Wald test (χ^2, *df*, *p*-value)		31.900, 4, <0.0001		12.486, 4, 0.014		8.686, 4, 0.069	

Note. Estimates are presented as coefficient (standard error) unless otherwise noted. *p<0.05, **p<0.01, ***p<0.001.

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)			-0.765 (-1.213, -0.352)		-0.454 (-0.784, -0.130)		-0.380 (-0.676, -0.085)
Wald test (χ^2, *df*, *p*-value)		26.322, 4, <0.0001		10.179, 4, 0.038		14.611, 4, 0.006	

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, *df*, *p*-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.