



REVIEW

Re-orientating systematic reviews to rigorously examine what works, for whom and how: Example of a realist systematic review of school-based prevention of dating and gender violence

Chris Bonell¹  | Bruce Taylor² | Vashti Berry³ | Sidnei R. Priolo Filho⁴ |
 Andrew Rizzo⁵ | Caroline Farmer³ | Ann Hagell⁶ | Honor Young⁷ |
 Noreen Orr³ | Naomi Shaw³ | Annah Chollet⁸ | Fraizer Kiff³ |
 Emma Rigby⁶ | G. J. Melendez-Torres³ 

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

²National Opinion Research Center, Chicago, Illinois, USA

³College of Medicine and Health, South Cloisters, University of Exeter, Exeter, UK

⁴Universidade Tuiuti de Paraná, Curitiba, Brazil

⁵College of Health and Human Performance, University of Florida, Gainesville, Florida, USA

⁶Association for Young People's Health, CAN Mezzanine, London, UK

⁷Centre for Development, Evaluation, Complexity and Implementation in Public Health Improvement (DECIPHer), Cardiff University, Cardiff, UK

⁸Department of Social Policy and Intervention, University of Oxford, Oxford, UK

Correspondence

Chris Bonell, Department of Public Health Environments and Society, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, WC1H 9SH, London, UK.
 Email: chris.bonell@lshtm.ac.uk

Funding information

National Institute for Health Research, Grant/Award Number: 17/151/05

Abstract

Conventional systematic reviews offer few insights into for whom and how interventions work. 'Realist reviews' examine such questions via examining 'context-mechanism-outcome configurations' (CMOCs) but are insufficiently rigorous in how evidence is identified, assessed and synthesised. We developed 'realist systematic reviews', addressing similar questions to realist reviews but using rigorous methods. We applied this to synthesising evidence on school-based prevention of dating and relationship violence (DRV) and gender-based violence (GBV). This paper reflects on overall methods and findings, drawing on papers reporting each analysis. Drawing on intervention descriptions, theories of change and process evaluations, we developed initial CMOC hypotheses: interventions triggering 'school-transformation' mechanisms (preventing violence by changing school environments) will achieve larger effects than those triggering 'basic-safety' (stopping violence by emphasising its unacceptability) or 'positive-development' (developing students' broader skills and relationships) mechanisms; however, school transformation would only work in schools with high organisational capacity. We used various innovative analyses, some of which aimed to test these hypotheses and some of which were inductive, drawing on available findings to augment and refine the CMOCs. Overall, interventions were effective in reducing long-term DRV but not GBV or short-term DRV. DRV prevention occurred most effectively via the 'basic-safety' mechanism. 'School-transformation' mechanisms were more effective in preventing GBV but only in high-income countries. Impacts on long-term DRV victimisation were greater when working with a critical mass of participating girls. Impacts on long-term DRV perpetration were greater for boys. Interventions were more effective when focusing on skills, attitudes and relationships,

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Research Synthesis Methods* published by John Wiley & Sons Ltd.

or lacking parental involvement or victim stories. Our method provided novel insights and should be useful to policy-makers seeking the best interventions for their contexts and the most information to inform implementation.

KEYWORDS

adolescents, dating and relationship violence, gender based violence, realist, systematic reviews, transferability

Highlights

What is already known

- Conventional systematic reviews offer few insights into for whom and how interventions work; 'realist reviews' examine such questions via examining 'context-mechanism-outcome configurations' (CMOCs) but are insufficiently rigorous in how evidence is identified, assessed and synthesised.

What is new

- We developed 'realist systematic reviews', addressing similar questions to realist reviews but using rigorous methods, and applied this to synthesising evidence on school-based prevention of dating and relationship violence (DRV) and gender-based violence (GBV).
- We found that, overall, interventions were effective in reducing long-term DRV but not GBV or short-term DRV.
- DRV prevention occurred most effectively via the 'basic-safety' mechanism; 'school-transformation' mechanisms were more effective in preventing GBV but only in high-income countries.
- Impacts on long-term DRV victimisation were greater when working with a critical mass of participating girls, while impacts on long-term DRV perpetration were greater for boys.
- Interventions were more effective when focusing on skills, attitudes and relationships, or lacking parental involvement or victim stories.

Potential impact for research synthesis methods readers

- Our method provided novel insights and should be useful to policy-makers seeking the best interventions for their contexts and the most information to inform implementation.

1 | INTRODUCTION

In this paper, we describe a method we have developed across several systematic reviews¹⁻⁵ of conducting 'realist' analyses within systematic reviews. We report on how we applied this method within a systematic review of school-based prevention of dating and relationship violence (DRV) and gender-based violence (GBV). We provide an overview of the methods and results of this review, and reflect on its implications for understanding and implementing interventions aiming to prevent DRV and GBV. We do not report in detail on the methods used or the empirical results from the various analyses conducted as part of the review because these are already published.⁶⁻⁹ Instead, we pull together these various analyses together to reflect on how

these methods and findings can be harnessed to provide a more nuanced understanding of how school-based DRV and GBV prevention works and the contexts in which it might work best, and then to reflect on the methodological value of our realist systematic review method.

Our method aims to develop a more nuanced understanding of how interventions work and the contexts (i.e., settings or populations) in which they might work best than do conventional systematic reviews. This method develops, tests, augments and refines hypotheses in the form of context-mechanism-outcome configurations (CMOC), that is, how intervention mechanisms interact with context to generate outcomes.¹⁰ But it does so in a more transparent and rigorous way than is achieved to date within 'realist reviews'.¹¹

Traditional systematic reviews report the overall effects of an intervention on an outcome for a certain population and comparator, pooling effect sizes from different studies.¹² This assumes there is a single true overall effect across studies in different contexts, which is unlikely to be true for many complex social interventions, defined as those with intervention components that interact with each other and with local context.¹³ Traditional reviews sometimes report subgroup analyses, pooling effect sizes for groups defined by setting or population.¹⁴ However, in the absence of theorising how intervention mechanisms interact with context to generate outcomes, these are unlikely to provide clear insights into how interventions work or might transfer across contexts.¹⁵ Traditional systematic reviews also generally do not synthesise evidence on intervention implementation or what factors affect this, again hindering consideration of transferability.¹²

DRV and GBV are important public-health problems with high prevalence among adolescents in all regions of the world, and significant consequences for current and future health and health inequalities.^{16,17} However, existing traditional systematic reviews offer few insights into what works, for whom and how. For example, a Cochrane systematic review reported a meta-analysis pooling intervention effects on DRV from multiple studies, finding no overall evidence of effectiveness and considerable heterogeneity in effects between studies, which remained unexplained.¹⁸ The review did not synthesise evidence on implementation or causal mechanisms. Other systematic reviews on the topic of DRV and GBV among young people also have not addressed these questions.^{19–22}

‘Realist reviews’ might potentially address these gaps.¹¹ Realist reviews synthesise various findings from diverse study designs oriented towards ‘theory tracking’ (defining CMOCs) and ‘theory testing’ (testing and refining CMOCs). These reviews do not use quality-assessment criteria because reviewers are interested, not in the overall quality of a study, but rather the validity of particular findings which are incorporated into a review, and this is deemed to require expert judgement rather than standardised checklists.¹¹ A reporting standard for realist reviews²³ argues:

‘Within any document, there may be several pieces of data that serve different purposes, such as helping to build one theory, refining another theory and so on. Therefore, the selection (for inclusion or exclusion) and appraisal of the contribution of pieces of data within a document cannot be based on an overall assessment of study or document quality.’ (p. 809)

Once realist reviews have reviewed literature to ‘track’ theory and define CMOCs, they then synthesise empirical evidence to test and refine these CMOCs. This occurs by reviewers assessing the plausibility of their CMOCs in the light of particular findings from empirical studies. Rather than identifying all pertinent evidence and statistically pooling data, realist reviews strive for ‘saturation’. They aim to include diverse evidence to offer different perspectives on the plausibility of the CMOCs. ‘Saturation’ is reached when no new insights emerge, as the originators of realist reviews explain¹¹:

‘A decision has to be made not just about which studies are fit for purpose in identifying, testing out or refining the programme theories, but also about when to stop looking – when sufficient evidence has been assembled to satisfy the theoretical need or answer the question. This test of saturation, comparable to the notion of theoretical saturation in qualitative research, can only be applied iteratively, by asking after each stage or cycle of searching whether the literature retrieved adds anything new to our understanding of the intervention and whether further searching is likely to add new knowledge.’ (p.28)

Thus, realist reviews aim to assess their hypotheses in a different way to traditional systematic reviews. Rather than examining whether statistical regularities in the data align with study hypotheses, realist syntheses focus on *narratives*, assessing the plausibility of their own narratives of CMOCs in the light of the various narratives suggested by included studies, hence their taking a purposeful approach to inclusion.

This distinctive orientation is clearly apparent in published realist syntheses.^{24,25} Unfortunately, there are no realist reviews of school-based prevention of DRV or GBV but realist reviews have been conducted in other areas of adolescent health promotion. For example, one review focused on how school tobacco policies (STP) influence student smoking. It included evidence based on whether studies provided rich, detailed description of how policies trigger mechanisms.²⁵ The review did not prioritise the inclusion of studies using designs offering more rigorous evidence of effectiveness. The review did not present each included study's methods or findings. Instead, it narratively described several possible CMOCs and then considered whether these aligned with the narratives apparent in the findings of included studies. A study mapping realist reviews more generally confirms that this is standard practice, reporting that few such

reviews report how studies are appraised and synthesised.²⁶

We agree with realist reviewers that evidence syntheses should examine not only overall effectiveness but also intervention mechanisms and how these interact with context to generate outcomes. But we propose a method which differs in several important ways from realist reviews as described above. First, although a narrative approach is appropriate when developing CMOCs, we believe that when testing CMOCs, it is important to examine whether empirical regularities (and not merely the narratives) as reported in empirical outcome evaluations align with what CMOCs would predict. This offers the most rigorous way to examine whether the pattern of effects aligns with hypotheses. Second, we believe that inclusion criteria for outcome evaluations need to refer to study designs because some offer more rigorous means of testing CMOC hypotheses than others. Lastly, we believe that reviews need to include all pertinent studies rather than merely a purposive subset, in order for quantitative analyses of the patterning of effects to be unbiased.

2 | SUMMARY OF MATERIALS AND METHODS

Here we provide a summary of methods, which are reported in detailed elsewhere.^{6,7} Our approach is summarised in Figure 1. The review was registered with PROSPERO (CRD42020190463) and followed PRISMA reporting criteria.⁷ We included randomised controlled trials (RCT) (with treatment-as-usual, waitlist or active control groups) and process evaluations of school-based prevention of DRV and/or GBV victimisation and/or perpetration among students aged 5–18 years. DRV was defined as physical, sexual and emotional violence in relationships between young people. GBV was defined as violence rooted in gender equality and sexuality within or outside dating relationships. We focused on RCTs because it is feasible and appropriate for school DRV/GBV interventions to be evaluated using this design, which offers the least biased estimates of effects. We searched 21 bibliographic databases in July 2020 from inception and without limitation on date or language.⁶ These searches were updated in June 2021. We also completed forwards and backwards citation checking on included studies and consulted with experts. Two reviewers piloted screening of successive batches of 100 titles/abstracts, discussing disagreements and calling on a third reviewer where needed. Once 90% agreement was reached, each title/abstract was reviewed independently. Studies not excluded were screened against the inclusion criteria

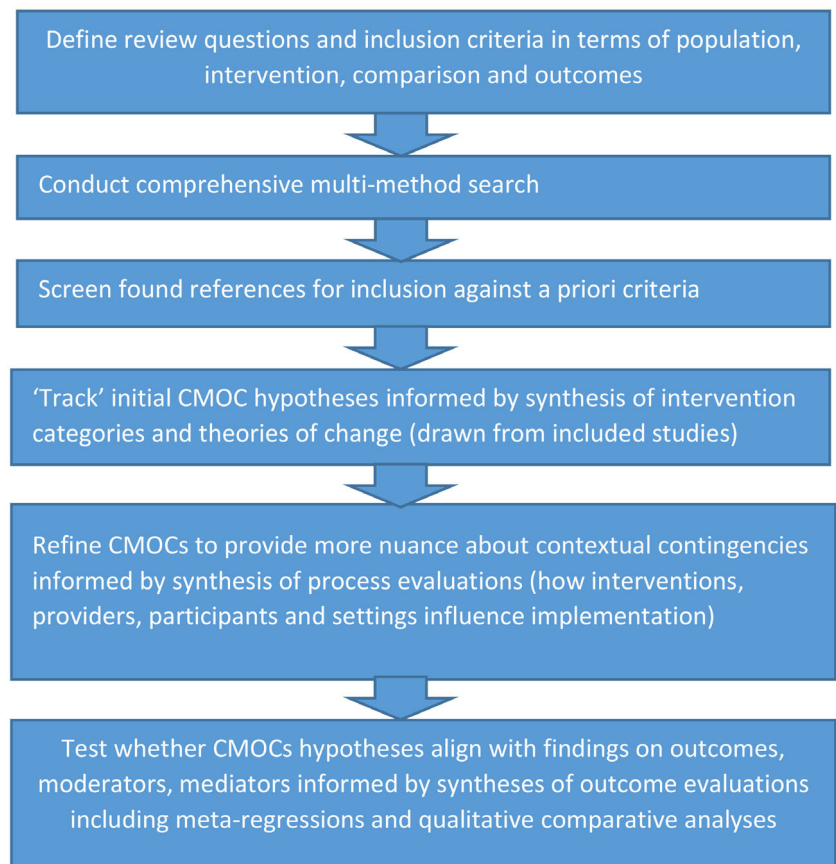
by two reviewers. Included studies were assigned to one or more evidence types (process, outcome or economic evaluation, mediation or moderation analysis). Data extraction is described elsewhere.⁶ Process evaluations were quality-assessed using the EPPI-Centre tool²⁷ and outcome evaluations using the Cochrane risk-of-bias tool²⁸ by two reviewers in duplicate.

Syntheses first aimed to ‘track’ and define initial CMOCs hypotheses by synthesising intervention descriptions, theories of change and process evaluations. We drew on intervention descriptions to categorise interventions. We synthesised theories of change as described in process and outcome evaluations by intervention category. We also drew on existing middle-range theory where this aligned with theories of change and enabled deeper insights. We then synthesised process evaluation to understand how features of interventions, providers, settings and recipients influence implementation and receipt again using meta-ethnographic methods.²⁹ These identified cases of reciprocal translation (the same concepts described in different terms between studies), refutational synthesis (contradictions in the concepts expressed in different studies) and line of argument (concepts from different studies allowed us to build a bigger picture of mechanisms than was available in any one study).

Our synthesis then moved on to test these initial CMOCs by assessing how well they aligned with quantitative evidence from outcome evaluations in terms of intervention overall effects, mediators, moderators and necessary conditions for effectiveness, drawing on narrative syntheses, statistical meta-analyses, meta-regressions, network meta-analyses, harvest plots and qualitative comparative analysis (QCA) methods.⁷ The analyses possible within a systematic review depend on the analyses reported by included studies, which means that not all of our CMOC hypotheses could be tested by our syntheses. It also means that some of the analyses we could perform were not orientated towards testing a prior CMOC but were more data-driven and inductive. We undertook these when we thought they might usefully inform our understanding of how and for whom interventions work.

Outcomes were classed as short-term (<1 year follow-up) or long-term (≥1 year follow-up). Narrative synthesis and meta-analyses examined overall effects. Meta-regression examined what study-level characteristics moderated intervention effects.¹² Network meta-analyses explored the relative effectiveness of intervention sub-categories.³⁰ QCA used Boolean logic (combinations of conditions linked by ‘and’, ‘not’ and ‘or’) to assess the necessary and sufficient conditions for intervention effectiveness.^{31,32}

FIGURE 1 Realist systematic review approach.



3 | OVERVIEW OF RESULTS

Full results are provided elsewhere.^{6–9} Searches identified 40,160 unique records, of which 793 were screened in full text. Of these, 68 outcome evaluations and 137 process evaluations were eligible for inclusion. Because our aim was to iteratively define and then test CMOCs, we report and reflect on findings in each section of the results below.

3.1 | Intervention descriptions

Synthesis of intervention descriptions were inductive and suggested various intervention categories, which are reported in detail in an earlier paper.⁹ In terms of mechanisms, many interventions aimed to equip potential perpetrators and victims with the basic capabilities (e.g., knowledge of harms) and motivations (e.g., that violence is unacceptable) needed to stop or avoid DRV/GBV, which we termed the ‘basic-safety’ mechanism. In contrast, some interventions aimed to promote a broader set of social skills (e.g., negotiating conflict) and healthy relationships which it was theorised would reduce DRV/GBV. We termed this the ‘positive-development’ mechanism. Student components in both

these sub-categories included activities such as: guided practising of skills; group discussions; individual reflection; visual/image- or narrative-based learning; or student competitions in class. Staff components usually involved training to build capacity for implementation.

Other interventions aimed to modify school social or physical environments to prevent DRV/GBV. We termed this the ‘school-transformation’ mechanism. Components addressing school social environments included visits from community organisations, changes to school policies, participative customisation of activities and school clubs. Components addressing the physical environment included posters in shared spaces and staff monitoring the safety of school spaces.

Overall, interventions could be categorised discretely as: student-level interventions (involving a curriculum or some other student-focused single component aiming to trigger the ‘basic-safety’ or ‘positive-development’ mechanisms); multi-component interventions (involving multiple student- or staff-level components aiming to trigger the ‘basic-safety’ or ‘positive-development’ mechanisms); and multi-level interventions (involving change at multiple levels including the individual, classroom and school environment context aiming to trigger the ‘school-transformation’ mechanism).

3.2 | Theories of change

Synthesis of theories of change was similarly inductive and informed refinements to the above mechanisms. These are reported in detail in an earlier paper.⁸ The ‘basic-safety’ mechanism focused on the prevention of negative behaviours (e.g., by identifying DRV or GBV as unacceptable behaviours). For example, the ‘DAT-E Adolescence’ intervention aimed to provide students with ‘basic-safety’ knowledge, attitudes and skills needed to stop perpetration or avoid victimisation.³³ The ‘positive development’ mechanism involved promoting a broader set of behaviours such as broader conflict-management and healthy-relationship skills. For example, the ‘DRV Curriculum’ intervention aimed to promote students’ ability to initiate and maintain healthy relationships.³⁴

The ‘school-transformation’ mechanism involved school staff transforming school environments to promote students’ school belonging and acceptance of pro-social norms via promoting student participation and relationships with staff. Theories of change for such interventions reciprocally translated not only with each other but also with an existing middle-range theory, the theory of human functioning and school organisation,³⁵ which we used to deepen our synthesis and provide overarching terminology. This theory is supported by evidence of school effects on substance use and violence, although not specifically DRV or GBV.² Drawing on this theory where it resonated with intervention theories of change, the school-transformation mechanism was theorised to involve ‘de-classification’ (i.e., eroding ‘boundaries’ and strengthening relationships between staff and students, the classroom and wider school, schools and local communities, and different professional roles within schools). Adding nuance to this picture and challenging some of the middle-range theories assumptions, refutational synthesis suggested that some interventions, particularly in high-crime areas, could be understood in terms of a mechanism increasing not eroding boundaries between the school and community to reject local pro-violence norms. Again informed by included studies and the theory of human functioning and school organisation, the school-transformation mechanism was also theorised to involve ‘reframing’, that is, increasing student participation in decisions at the level of the classroom (e.g., interactive, experiential learning) and the school (e.g., contribution to policies and decisions) so that these better align with student preferences. Using a terminology again informed by the theory of human functioning and school organisation which aligned with intervention theories, ‘de-classification’ and ‘reframing’ were in turn theorised to increase students’ sense of safety and belonging in school, and

acceptance of pro-social and anti-violence norms, and through this reduce their involvement in DRV and GBV.

Interventions varied in the extent to which they addressed all or some aspects of this ‘school-transformation’ mechanism. Multi-level interventions aimed to trigger such mechanisms at multiple levels of the school system including the individual, classroom and school. Classroom-level interventions aimed to trigger mechanisms at the level of staff-student relationships, reframing learning to be more interactive and increasing student commitment to learning.

We use these synthesised theories of change to define (or ‘track’ as realist term it) the mechanism and outcome element of CMOCs (Table 1; first column). We hypothesised that the ‘basic-safety’ and ‘positive-development’ mechanisms would achieve smaller and less sustained effects (because these involved fewer mechanisms of prevention and did not aim to permanently transform school environments). We hypothesised that multi-level interventions would trigger the school-transformation mechanism across multiple levels, and that these would achieve larger and more sustained effects because they would encourage enduring transformations in school environments. The synthesised theories of change did not at this stage offer strong insights into how such mechanisms might interact with different contexts to generate different outcomes. To consider this question, we turned to process evaluations.

3.3 | Synthesising process evaluations

Process evaluations were synthesised inductively and examined factors influencing implementation. These are reported in detail elsewhere.⁷ We found that, at the school level, implementation was facilitated by: school resources and infrastructure; school organisation and leadership capacity; reduced time constraints and competing priorities; and positive school perceptions of the importance of addressing DRV/GBV. Intervention characteristics that facilitated better implementation included the ease of delivery and the ease of intervention modification to the particularities of the setting.

Informed by the findings from these qualitative syntheses, we refined the wording of our CMOCs (Table 1 second column; Figure 2). We came to hypothesise that: multi-level interventions triggering the school-transformation mechanism across multiple levels would only work in school contexts with high organisational capacity; interventions triggering the ‘basic-safety’ and ‘positive-development’ mechanisms would work in all school contexts, including in more resource-poor contexts; and such interventions would

TABLE 1 Refinement of CMOCs.

Initial CMOC	CMOC refined as a result of process evaluation syntheses	CMOCs refined as a result of outcome, mediator and moderator syntheses	CMOCs refined as a result of QCA
<p>Multi-level interventions trigger a 'school transformation' mechanism involving de-classification and reframing across multiple levels to increase student commitment to school with changes to capabilities and motivations, generating reductions in DRV and GBV. Multi-level interventions achieve large effects and are highly sustainable because they effect enduring transformations in school organisations and social structures.</p>	<p>Multi-level interventions triggering a school transformation mechanism generally only work in schools with high capacity and resourcing.</p>	<p>Multi-level interventions aiming to trigger school transformation mechanism do not generally increase student commitment/belonging to school or change social norms and do not generate reductions in DRV/GBV. In some contexts (schools with high capacity and resourcing in high-income countries), the school transformation mechanism may be triggered and be sufficient to generate reductions in GBV.</p>	<p>The implementation of social structural components is important to trigger school transformation mechanisms to prevent GBV perpetration.</p>
<p>Some student-level (curriculum or other single component) interventions trigger a 'basic safety' mechanism, directly promoting individual- and group-level capabilities and motivations to avoid DRV and GBV. Single-level interventions achieve smaller effects and are less sustainable because these involved fewer mechanisms of prevention and do not aim to transform school organisation or social structures.</p>	<p>Student-level interventions triggering a basic safety mechanism work in all schools including in more resource-poor contexts because they require less school capacity and resourcing.</p>	<p>Student-level interventions triggering a basic safety mechanism directly promote individual-level capabilities and motivations (particularly negative attitudes to violence) are sufficient to generate reductions in DRV. Outcomes were greater among boys, those with previous experience of DRV perpetration and when delivered to school populations with a higher proportion of girls (perhaps because of greater aggregate student engagement with the intervention). These outcomes take time to generate possibly because of the time taken for individuals to apply new capabilities and motivations in new relationships. Single-level interventions aiming to trigger mechanisms directly promoting individual-level capabilities and motivations are less likely to generate reductions in GBV and only in high-income countries. GBV is a more public behaviour influenced by social norms that this mechanism is often not sufficient to modify.</p>	<p>The basic safety mechanism reduces victimisation via a reduction of perpetration. Mechanisms are more likely to be triggered when interventions involve single-gender components, involved a critical mass of girls, provide opportunities for guided practice of skills and attitudes, and focus on student relationships. Mechanisms are less likely to be triggered when interventions involve narrative or parent-involvement components.</p>
<p>Some student-level (curriculum) interventions trigger a 'positive development' mechanism directly promoting individual- and group-level capabilities and motivations to enact positive behaviours (such as conflict management and healthy relationships) which are accompanied by reductions in DRV and GBV. Single-level interventions achieve smaller effects and are less sustainable because these involved fewer mechanisms of prevention and do not aim to transform school organisation or social structures.</p>	<p>Student-level interventions triggering a positive development mechanism work in all schools including in more resource-poor contexts because they require less school capacity and resourcing.</p>	<p>Student-level interventions aiming to trigger positive development mechanisms are not sufficient to generate reductions in DRV/GBV.</p>	<p>-</p>

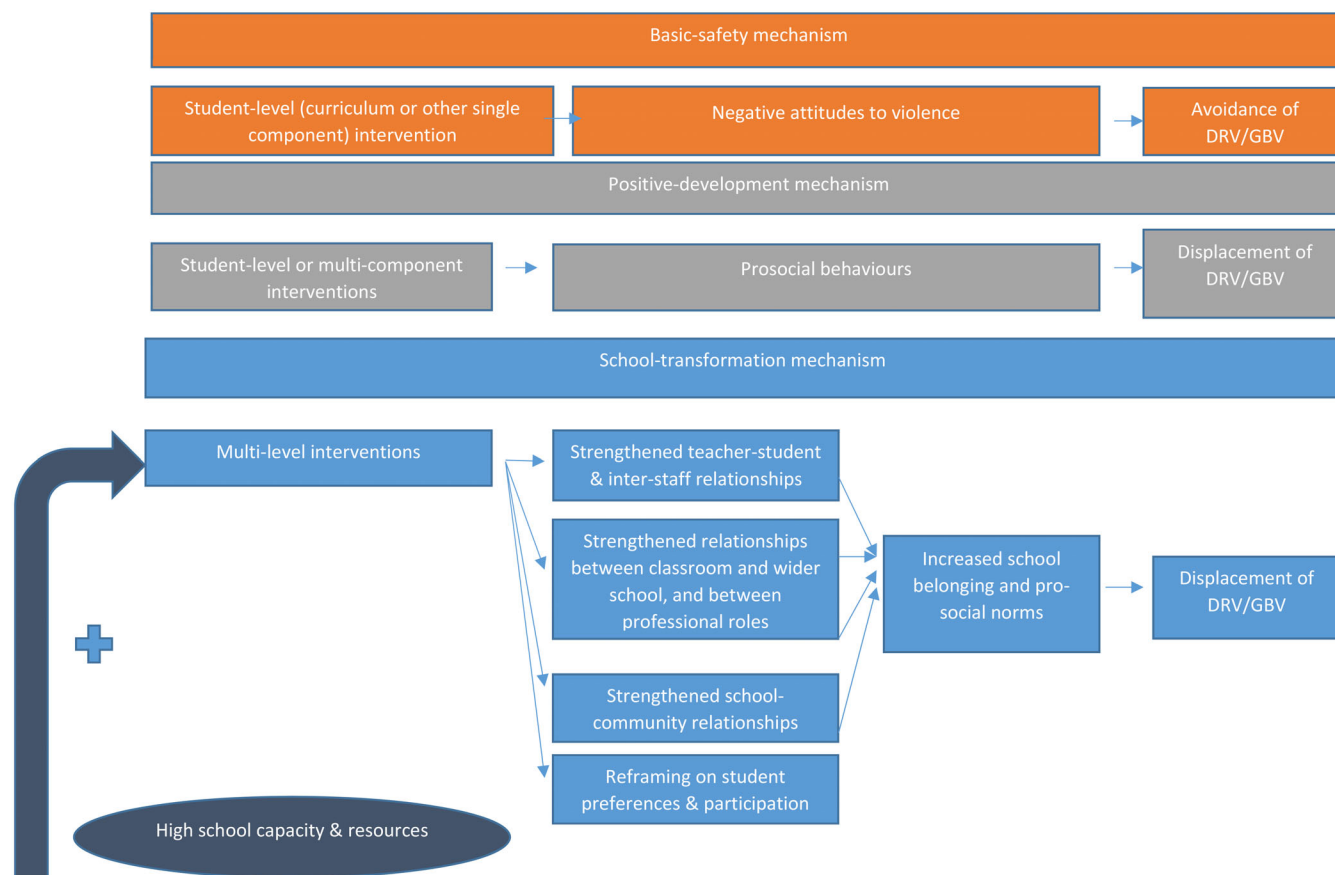


FIGURE 2 Context-mechanism-outcome configuration refined through synthesis of process evaluations.

be more effective when they were locally modifiable and interactive, and involved support from external agencies.

3.4 | Synthesising outcome evaluations

We conducted various quantitative analyses of outcome evaluations. Some of these aimed to test the above CMOCs where the analyses reported by included studies enabled this. Some of our analyses were more inductive, making use of the findings from included studies even where these did not directly speak to our hypotheses if we thought that these might nonetheless help us to augment or refine our CMOCs. All these analyses are reported in detail elsewhere.^{6,7} Drawing on inductive syntheses, we found that there were overall intervention effects on long-term (1 year or above) but not short-term (less than 1 year) DRV perpetration and victimisation. Forest plots and pooled estimates are reported elsewhere.⁶ This may be because many interventions required time to implement and benefit students. Outcomes may only have manifested when students initiated dating behaviours or entered new relationships.³⁶

There were no overall intervention effects on GBV victimisation or perpetration at either time-point. This might be because whereas DRV tends to be a behaviour occurring in the private context of dating and relationships, GBV can be a more public activity given the inherent performative aspects of gender roles and gender norms.³⁷ Consequently, while DRV might be more amenable to change via partners learning new capabilities and motivations, GBV may be more influenced by collective social norms, which are harder to modify. However, there was some evidence, from studies in high-income countries, of long-term effects on reduction of GBV victimisation and perpetration.⁷ This might be because GBV could be reduced by longer-term transformations in higher-capacity school systems.

We then tested our CMOC hypotheses about different sorts of interventions by examining whether intervention components could explain differences in effectiveness between studies via meta-regression but found no evidence for this. A similar finding arose from our network meta-analysis, which did not provide clear signals as to the differential contribution to effectiveness of intervention component classes. There was some evidence that single-component interventions were more effective for

long-term DRV victimisation and perpetration,⁷ possibly because simpler interventions might, as hypothesised in our CMOCs, be more feasible to deliver in more schools and so their effects accumulate. It may also be that simpler interventions allow a narrower but more sustained focus on key behaviours.

We then used narrative synthesis of study-level mediation analyses to test our hypotheses about the mechanisms of action. These indicated that there was some evidence that reductions in measures of student violence acceptance mediated intervention effects on long-term DRV victimisation and perpetration.⁷ There was inconsistent evidence that increased student knowledge mediated intervention effects on long-term DRV victimisation and perpetration. There was no evidence that improved student conflict-management skills, bystander actions or school belonging mediated intervention effects on long-term DRV victimisation or perpetration. This adds to the picture that interventions might reduce DRV primarily via impacts on individual capabilities and motivations (particularly attitudes towards violence) rather than via effects on school environments.

There was evidence from one study that an intervention reduced long-term GBV victimisation outcomes by improving school belonging.³⁸ A single study found evidence for student sense of school belonging occurring for intervention effects on some long-term GBV perpetration outcomes.³⁹ This evidence might suggest that, where GBV is prevented, this is most likely to occur via school-transformation mechanisms, although it is not clear what types of school characteristics might play a role and evidence from the analysis of intervention outcomes on GBV suggest that such mechanisms might rarely be triggered sufficiently to achieve significant reductions in GBV.

Our narrative syntheses of moderation evidence were largely inductive in orientation, led by what analyses were reported by included studies. These suggested that intervention effects on long-term DRV victimisation did not differ by gender, prior experience of DRV victimisation, dating history, age, ethnicity, acculturation or sexual orientation.⁷ There was some evidence from meta-regression that the proportion of the study sample that was female was associated with an increase in intervention effectiveness for preventing long-term DRV victimisation. This might be interpreted as evidence that school interventions are more likely to be effective in settings where a critical mass of female students encourage greater student engagement with intervention and the de-normalisation of violence.

There was evidence from multiple studies for gender moderating intervention effects on DRV perpetration, with greater effects for boys. There was weaker evidence

that intervention effects were greater for those with prior DRV perpetration. The finding that, for some DRV perpetration outcomes, effects were larger for boys suggests that many interventions were not gender-neutral and may have been interpreted by students as interventions aiming to reduce male perpetration of DRV and, also informed by the above findings on mediation, might have achieved these effects via changes in male attitudes to violence. There was weak evidence that such mechanisms were slightly stronger among those previously engaged in perpetration.

We tested and refined our CMOCs in the light of these syntheses (Table 1; third column; Figure 3). We concluded that multi-level interventions aiming to trigger the school-transformation mechanism do not generally increase student commitment/belonging to school or change social norms and do not generate reductions in DRV/GBV. In some settings (schools with high capacity and resourcing in high-income countries), the school-transformation mechanism may be triggered and be sufficient to generate reductions in GBV. We concluded that single-level interventions triggering the basic-safety mechanism which directly promote individual-level capabilities and motivations (particularly negative attitudes to violence) are sufficient to generate reductions in DRV. This is particularly so among boys, those with previous DRV perpetration and when delivered to school populations with a higher proportion of girls. However, these outcomes take time to manifest, possibly because of the time taken for individuals to apply new capabilities and motivations in new relationships. We concluded that single-level interventions aiming to trigger mechanisms directly promoting individual-level capabilities and motivations are less likely to generate reductions in GBV and only in high-income countries. GBV is a more public behaviour influenced by social norms, which the basic-safety mechanism is often not sufficient to modify. Finally, we concluded that single-level interventions aiming to trigger positive-development mechanisms are not sufficient to generate reductions in DRV or GBV.

We then turned to QCA analyses, with these again being inductive in orientation and dependent on what analyses included studies reported. Our QCAs were able to differentiate between the most effective and other interventions, with the exception of QCAs focused on short-term DRV perpetration.⁷ An important finding from our QCA was that a key condition for reduction of victimisation was reduced perpetration, across short-term and long-term DRV victimisation and short-term GBV perpetration. However, a number of other pathways to the reduction of victimisation were apparent in QCAs,

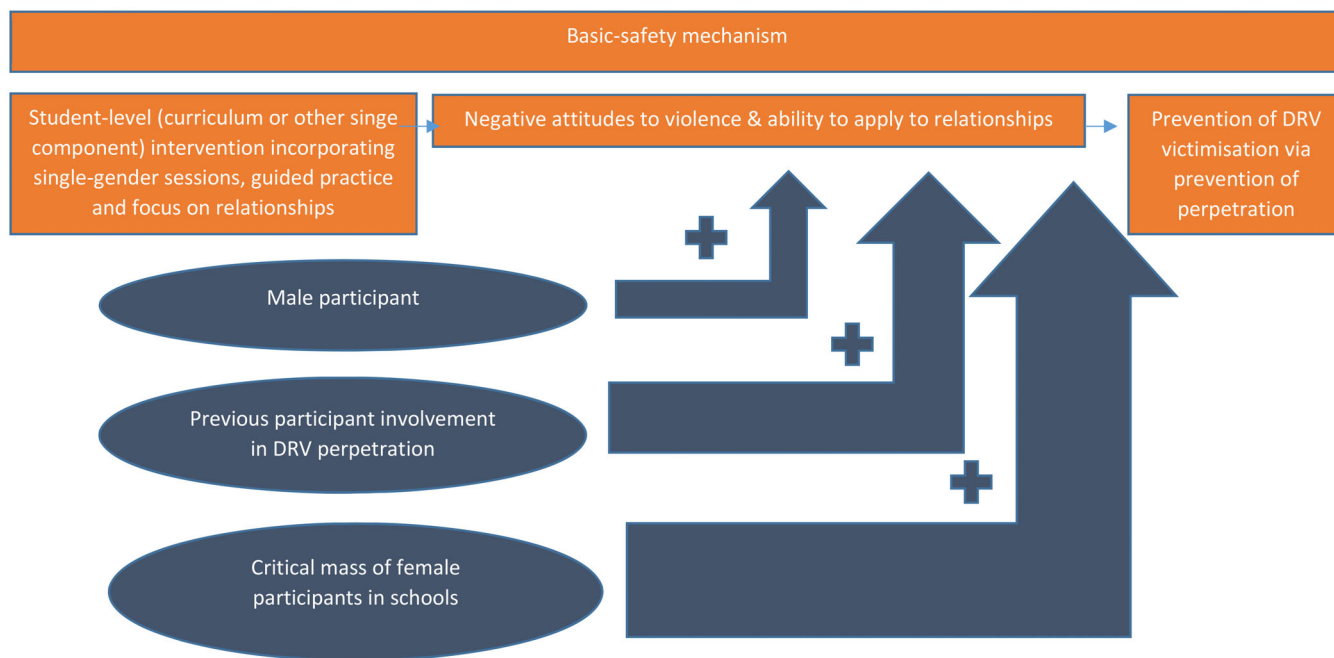


FIGURE 3 Basic-safety mechanism refined through our syntheses.

generally characterised by the inclusion of single-gender components or a critical mass of girls. A critical mass of girls was especially important where interventions involved more than one component. There was some evidence that the absence of some components was a necessary condition for most effectiveness. For long-term DRV victimisation, the absence of parental involvement was central to achieving effectiveness. For short-term GBV victimisation, the absence of victims telling their stories in school was an important part of causal pathways. It is possible that the absence of these components reduces opportunities to receive conflicting messages about the importance of preventing these outcomes.

But if reductions in victimisation are principally achieved through reductions in perpetration, how are reductions in perpetration achieved? Our QCA was unable to develop a satisfactory model for short-term DRV perpetration. However, for both long-term DRV perpetration and short-term GBV perpetration, interventions that were most effective incorporated a range of opportunities for guided practice of skills and attitudes, and interpersonal components focusing on student relationships. Importantly, the implementation of school environmental components was central to effectiveness for short-term GBV perpetration, but not for DRV perpetration.

We refined our CMOCs in the light of these QCAs (Table 1 fourth column; Figure 3). We concluded that the basic-safety mechanism reduces victimisation via reductions in perpetration. Mechanisms are more likely to be triggered when interventions involve single-gender components,

involve a critical mass of girls, provide opportunities for guided practice of skills and attitudes, and focus on student relationships. Mechanisms are less likely to be triggered when interventions involve victim-narrative or parent-involvement components. We concluded that implementation of school environmental components (e.g., changes to school policies, participative customisation of activities, and school clubs) is important to trigger school-transformation mechanisms to prevent GBV perpetration.

4 | CONCLUSIONS

4.1 | Summary of key findings

We defined CMOCs drawing on a synthesis of intervention descriptions and theories of change and then refined these informed by synthesis of process evaluations, providing more information about contextual contingencies. We initially hypothesised that multi-level interventions triggering a 'school-transformation' mechanism would achieve larger effects and be more sustainable than interventions triggering 'basic-safety' and 'positive-development' mechanisms. But we also hypothesised that the multi-level interventions would only work in schools with high organisational capacity whereas simpler interventions would work in all schools. We then conducted various syntheses of outcome evaluations (Table 1). Some of these were, where primary studies allowed, orientated towards hypothesis-testing but others were inductive,

dependent on what analyses primary studies. Together these two different approaches allowed us to test but also to augment and refine our CMOCs.

In undertaking these various quantitative analyses, we examined whether, across studies, markers of mechanism were associated with markers of outcomes, and on what markers of context these were contingent. This required a clear definition of which study types provided the best evidence to examine such associations and then ensuring that all such studies were included and their quality assessed. We found that interventions were effective in reducing DRV perpetration and victimisation (in the long but not the short term) but not GBV victimisation and perpetration. There was some evidence that the interventions we reviewed largely worked not by a school-transformation mechanism but via a basic-safety mechanism increasing student capabilities and motivations concerning the unacceptability of violence. There was evidence that this simpler basic-safety mechanism may have involved reductions in DRV perpetration among males and those with previous experience of perpetration.

We theorised that individual-level basic-safety mechanisms are more likely to effect changes in DRV rather than GBV perpetration because of the more private nature of DRV (meaning it is amenable to reduction via changes in partners' capabilities and motivations). However, the more public nature of GBV means that it might be influenced by social norms which interventions appear not to successfully address.³⁷ We found that interventions could be effective in preventing GBV but that this was only likely in high-income settings and required school-environmental components. It is possible that a school-transformation mechanism is needed to reduce GBV and that this requires a context of existing high school organisational capacity to deliver such components. Impacts on long-term DRV victimisation were greater when working with student populations with a critical mass of girls. Impacts on long-term DRV perpetration were greater for boys. Impacts on DRV victimisation occurred via impacts on perpetration. Interventions were more effective when they focused on skills, attitudes and relationships, or lacked parental involvement or victim stories.

4.2 | Limitations of the review

Our approach to using quantitative research to test our CMOCs might be open to criticism of taking a simplistically 'successionist' approach in which causation is inferred from constant conjunction.¹⁰ However, we contend that this would only be a valid criticism if we examined associations between causes and effects without reference to contextual contingencies. Instead, we took a generative approach,

aiming first to develop a rich and contextual understanding of how mechanisms might generate different outcomes in different contexts, and then testing whether this understanding aligns with contingent patterns of association among empirical data. We contend that this approach is no more successionist than realist reviews which rely on narrative conclusions from primary studies, which themselves are informed by statistical analyses of correlations and which generally do not attend to contextual contingencies. Furthermore, our synthesis drew not just on probabilistic statistical analyses of correlations but also on QCA analyses which examined more complex combinations of markers of context and mechanisms which co-occurred with markers of outcomes.

A more important limitation is that the analyses possible within a systematic reviewer are not entirely within the reviewers' control, as would be the case for analyses conducted within a primary study. In a trial, for example, we could ensure we collect data on the key potential moderators and mediators to allow thorough testing of hypotheses about mechanisms and contextual contingencies. We do not, however, have this control within a systematic review. We could test our prior CMOCs hypotheses only where what primary studies reported allowed this. However, we could complement this hypothesis-testing with a more inductive approach, summarising the findings from included studies and then considering what these might tell us about possible CMOCs. It is important to recognise the potential for multiple analyses of this sort to produce type-2 errors.

In the case of this review, a further limitation concerned our ability to test the middle-range theory that helped inform our CMOCs. The theory we chose resonated with our syntheses of intervention descriptions, theories of change and process evaluations but was quite broad and included some constructs which were not well defined. This, together with the more general limitation to hypothesis-testing, reduced our ability to test whether boundary erosion and de-classification did explain what empirical support we found for the school-transformation mechanism or whether other mechanisms not captured by this theory were responsible. Reviews will often not be able to definitively test middle-range theory when included studies are not designed for that purpose.

4.3 | Implications for policy and research

We draw a number of conclusions about intervention transferability. To prevent DRV, it may be less important to do something complex via multi-level interventions than to do something simpler well, via well-implemented

single-component interventions which might provide opportunities for guided practice of skills and attitudes, and focus on student relationships. Under-resourced schools and schools in areas of high deprivation may decide that they should focus on ensuring the basic-safety of students by suppressing harmful behaviours and postpone the encouragement of positive behaviours or implementing school environmental changes until they have the capacity to implement these well. School readiness and intervention choice provide important context. School readiness, whether defined as a receptive school climate, staff buy-in or strong school leadership, was linked with smoothing the path to implementation and unlocking a wider range of mechanisms beyond strictly student-directed mechanisms. Female critical mass may matter. In high-income but not low-income countries, it may be that interventions can reduce GBV perhaps by schools having more capacity to achieve transformations.

Thus, it is clear that, by conducting realist analyses within a systematic review, it is possible to develop conclusions that are much more nuanced, useful and rigorous than would be produced either by a conventional systematic review or by the current approach to a 'realist reviews'. By undertaking a mix of analyses aiming towards a coherent understanding of mechanisms (some of which test prior hypotheses and some of which augment these hypotheses based on inductive approach), we were able to identify promising intervention activities, the mechanisms via which these might generate benefits and the contextual contingencies affecting this. While uncertainties remain, for example as to the precise role of boundary erosion in these mechanisms, the uncertainties are much less than had we used existing approaches. The theories of change and process evaluations we included in our reviews were often individually quite limited but bringing them together allowed for a richer and more nuanced analysis. Our conclusions offer scientifically informative insights into possible mechanisms and how these might vary with context. Our broad searches and inclusion of studies regardless of language ensured that we drew on evidence from a diversity of contexts, facilitating our exploration of how mechanisms vary with context. Exploring such questions offers more nuanced suggestions about the scope and limits of intervention transferability. Our approach to synthesis achieves these ends while being clear as to which study designs are most useful to answer which questions and being transparent as to what evidence has informed which conclusions, and how this has been synthesised.

The question of which study designs to include in this type of review will depend on the questions asked. We judged that, for a review examining school-based universal interventions, it was feasible and appropriate to focus outcome assessment on RCTs. However, for reviews focused

on other interventions and settings less amenable to randomisation it may be appropriate to include quasi-experimental designs.⁴⁰

AUTHOR CONTRIBUTIONS

Chris Bonell: Conceptualization; investigation; funding acquisition; writing – original draft; methodology; writing – review and editing. **Bruce Taylor:** Conceptualization; funding acquisition; methodology; writing – review and editing. **Vashti Berry:** Conceptualization; funding acquisition; methodology; writing – review and editing. **Sidnei R. Priolo Filho:** Conceptualization; funding acquisition; methodology; writing – review and editing. **Andrew Rizzo:** Conceptualization; funding acquisition; methodology; writing – review and editing. **Caroline Farmer:** Conceptualization; methodology; writing – review and editing; funding acquisition. **Ann Hagell:** Conceptualization; funding acquisition; writing – review and editing; methodology. **Honor Young:** Conceptualization; funding acquisition; methodology; writing – review and editing. **Noreen Orr:** Conceptualization; methodology; funding acquisition; writing – review and editing. **Naomi Shaw:** Conceptualization; funding acquisition; methodology; writing – review and editing. **Annah Chollet:** Conceptualization; funding acquisition; methodology; writing – review and editing. **Fraizer Kiff:** Conceptualization; funding acquisition; methodology; writing – review and editing. **Emma Rigby:** Conceptualization; funding acquisition; methodology; writing – review and editing. **G.J. Melendez-Torres:** Conceptualization; investigation; funding acquisition; writing – original draft; methodology; writing – review and editing.

FUNDING INFORMATION

This study was funded by the National Institute for Health Research in England under its Public Health Research Board (17/151/05). 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.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Chris Bonell  <https://orcid.org/0000-0002-6253-6498>

G. J. Melendez-Torres  <https://orcid.org/0000-0002-9823-4790>

REFERENCES

- Bonell C, Dickson K, Hinds K, et al. The effects of positive youth development interventions on substance use, violence and inequalities: systematic review of theories of change, processes and outcomes. *Public Health Res.* 2016;4:1-218.
- 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.
- Meiksin R, Melendez-Torres GJ, Miners A, et al. E-health interventions targeting HIV/STIs, sexual risk, substance use and mental ill-health among men who have sex with men: systematic review. *Public Health Res.* 2022;10:1-322.
- Ponsford R, Melendez-Torres GJ, Miners A, Falconer J, Bonell C. Whole-school interventions promoting student commitment to school to prevent substance use and violence and improve educational attainment: a systematic review. *Public Health Res.* 2022; (Forthcoming).
- Tancred T, Melendez-Torres GJ, Papparini S, et al. Interventions integrating health and academic education in schools to prevent substance misuse and violence: systematic review and evidence synthesis. *Public Health Res.* 2018;7:1-244.
- Farmer C, Shaw N, Rizzo AJ, et al. School-based interventions to prevent dating and relationship violence and gender-based violence: systematic review and network meta-analysis. *Am J Public Health.* 2023;113:320-330.
- Melendez-Torres GJ, Orr N, Farmer C, et al. School-based interventions to prevent dating and relationship violence and gender-based violence (STOP-DRV-GBV): systematic review to understand characteristics, mechanisms, implementation and effectiveness. *Public Health Res.* 2022; (Forthcoming).
- Orr N, Chollet A, Rizzo A, et al. School-based interventions for preventing dating and relationship violence and gender-based violence: a systematic review and synthesis of theories of change. *Review of Educational Research.* 2022;10:e3382.
- Rizzo AJ, Orr N, Shaw N, et al. Exploring the activities and target audiences of school-based violence prevention programs: a systematic review and intervention component analysis. *Trauma, Violence Abuse.* 2022. doi:10.1177/15248380221134294
- Pawson R, Tilley N. *Realistic Evaluation.* Sage; 1997.
- Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review—a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy.* 2005;10:21-34.
- Chalmers, I., & Altman, D.G. (Eds.) (1995). *Systematic reviews.* London: BMJ Books.
- Moore G, Audrey S, Barker M, et al. *Process Evaluation of Complex Interventions: UK Medical Research Council (MRC) Guidance.* MRC Population Health Science Research Network; 2014.
- Lipsey MW. The primary factors that characterize effective interventions with juvenile offenders: a meta-analytic overview. *Vict Offenders.* 2009;4:124-147.
- Burchett HED, Kneale D, Blanchard L, Thomas J. When assessing generalisability, focusing on differences in population or setting alone is insufficient. *Trials.* 2020;21:286.
- Costa BM, Kaestle CE, Walker A, et al. Longitudinal predictors of domestic violence perpetration and victimization: a systematic review. *Aggress Violent Behav.* 2015;24:261-272.
- Loxton D, Dolja-Gore X, Anderson AE, Townsend N. Intimate partner violence adversely impacts health over 16 years and across generations: a longitudinal cohort study. *PLoS One.* 2017;12:e0178138.
- Fellmeth GLT, Heffernan C, Nurse J, Habibula S, Sethi D. Educational and skills-based interventions for preventing relationship and dating violence in adolescents and young adults. *Cochrane Database Syst Rev Art.* 2013;19:CD004534. doi:10.1002/14651858.CD004534.pub3
- De La Rue L, Polanin JR, Espelage DL, Pigott TD. A meta-analysis of school-based interventions aimed to prevent or reduce violence in teen dating relationships. *Rev Educ Res.* 2016;87:7-34.
- Kettrey HH, Marx RA, Tanner-Smith EE. Effects of bystander programs on the prevention of sexual assault among adolescents and college students: a systematic review. *Campbell Syst Rev.* 2019;15:e1013.
- Lundgren R, Amin A. Addressing intimate partner violence and sexual violence among adolescents: emerging evidence of effectiveness. *J Adolesc Health.* 2015;56:S42-S50.
- Stanley N, Ellis J, Farrelly N, Hollinghurst S, Downe S. Preventing domestic abuse for children and young people: a review of school-based interventions. *Children and Youth Services Review.* 2015;59:120-131.
- Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. *BMC Med.* 2013;11. doi:10.1186/1741-7015-11-21
- Defever E, Jones M. Rapid realist review of school-based physical activity interventions in 7-to 11-year-old children. *Child Youth Serv Rev.* 2021;8:52.
- Schreuders M, Nuyts PAW, van den Putte B, Kunst AE. Understanding the impact of school tobacco policies on adolescent smoking behaviour: a realist review. *Soc Sci Med.* 2017;183:e27.
- Berg RC, Nanavati J. Realist review: current practice and future prospects. *J Res Pract.* 2016;12:R1 <http://jrp.icaap.org/index.php/jrp/article/view/538/449>
- Shepherd J, Kavanagh J, Picot J, et al. The effectiveness and cost-effectiveness of behavioural interventions for the prevention of sexually transmitted infections in young people aged 13 to 19: a systematic review and economic evaluation. *Health Technol Assess Monogr.* 2010;14:1-206.
- Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ.* 2019;366:14898.
- Noblit G, Hare R. *Meta-Ethnography: Synthesizing Qualitative Studies.* Sage Publications Ltd; 1988.
- Dias S, Caldwell DMA. Network meta-analysis explained. *Arch Dis Child Fetal Neonatal Ed.* 2019;104:F8-F12.
- Melendez-Torres G, Sutcliffe K, Burchett HE, Rees R, Thomas J. Developing and testing intervention theory by incorporating a views synthesis into a qualitative comparative analysis of intervention effectiveness. *Res Synth Methods.* 2019;10:389-397.
- Thomas J, O'Mara-Eves A, Brunton G. Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example. *Syst Rev.* 2014;3:67.
- Munoz-Fernandez N, Ortega-Rivera J, Nocentini A, Menesini E, Sanchez-Jimenez V. The efficacy of the "Dat-E adolescence" prevention program in the reduction of dating

- violence and bullying. *Int J Environ Res Public Health*. 2019; 16:31.
34. Gage, A.J., Honoré, J.G., & Deleon, J. (2016). *Pilot Test of a Dating Violence-Prevention Curriculum Among High School Students: Emerging Evidence of Effectiveness in a Low-Income Country*. New Orleans: Tulane University.
35. Markham WA, Aveyard P. A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice. *Soc Sci Med*. 2003;56:1209-1220.
36. Wekerle C, Waechter RL, Leung E, Leonard M. Adolescence: a window of opportunity for positive change in mental health. *First Peoples Child Family Rev*. 2007;3:8-16.
37. Jewkes R, Flood M, Lang J. From work with men and boys to changes of social norms and reduction of inequities in gender relations: a conceptual shift in prevention of violence against women and girls. *The Lancet*. 2014;385:1580-1589. doi:[10.1016/S0140-6736\(14\)61683-4](https://doi.org/10.1016/S0140-6736(14)61683-4)
38. Espelage BK, Van Ryzin M, Holt MK. *Effects of a Middle School Social-Emotional Learning Program on Bullying, Teen Dating Violence, Sexual Violence, and Substance Use in High School*. National Institute of Justice, National Criminal Justice Reference Service; 2017.
39. Espelage LS, Polanin JR, Brown EC. The impact of a middle school program to reduce aggression, victimization, and sexual violence. *J Adolesc Health*. 2013;53:180-186.
40. Bonell C, Hargreaves J, Cousens S, et al. Alternatives to randomisation in the evaluation of public-health interventions: design challenges and solutions. *J Epidemiol Community Health*. 2011;65:582-587.

How to cite this article: Bonell C, Taylor B, Berry V, et al. Re-orientating systematic reviews to rigorously examine what works, for whom and how: Example of a realist systematic review of school-based prevention of dating and gender violence. *Res Syn Meth*. 2023;1-14. doi:[10.1002/jrsm.1644](https://doi.org/10.1002/jrsm.1644)