# Public Health Research

# Interventions integrating health and academic education in schools to prevent substance misuse and violence: a systematic review

*Tara Tancred, GJ Melendez-Torres, Sara Paparini, Adam Fletcher, Claire Stansfield, James Thomas, Rona Campbell, Suzanne Taylor and Chris Bonell*

# Interventions integrating health and academic education in schools to prevent substance misuse and violence: a systematic review

Tara Tancred,[1] GJ Melendez-Torres,[2] Sara Paparini,[1] Adam Fletcher,[2] Claire Stansfield,[3] James Thomas,[3] Rona Campbell,[4] Suzanne Taylor[1] and Chris Bonell[1]*

[1]Department of Social and Environmental Health Research, London School of Hygiene & Tropical Medicine, London, UK
[2]Development and Evaluation of Complex Interventions for Public Health Improvement, School of Social Sciences, Cardiff University, Cardiff, UK
[3]Evidence for Policy and Practice Information and Co-ordinating Centre, Social Science Research Unit, University College London Institute of Education, University College London, London, UK
[4]Development and Evaluation of Complex Interventions for Public Health Improvement, School of Social and Community Medicine, University of Bristol, Bristol, UK

*Corresponding author

This report should be referenced as follows:

Tancred T, Melendez-Torres GJ, Paparini S, Fletcher A, Stansfield C, Thomas J, *et al*. Interventions integrating health and academic education in schools to prevent substance misuse and violence: a systematic review. *Public Health Res* 2019;**7**(17).

# Public Health Research

**Criteria for inclusion in the *Public Health Research* journal**
Reports are published in *Public Health Research* (PHR) if (1) they have resulted from work for the PHR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

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

## PHR programme

The Public Health Research (PHR) programme, part of the National Institute for Health Research (NIHR), evaluates public health interventions, providing new knowledge on the benefits, costs, acceptability and wider impacts of non-NHS interventions intended to improve the health of the public and reduce inequalities in health. The scope of the programme is multi-disciplinary and broad, covering a range of interventions that improve public health. The Public Health Research programme also complements the NIHR Health Technology Assessment programme which has a growing portfolio evaluating NHS public health interventions.

For more information about the PHR programme please visit the website: http://www.nets.nihr.ac.uk/programmes/phr

## This report

# Abstract

## Interventions integrating health and academic education in schools to prevent substance misuse and violence: a systematic review

Tara Tancred,[1] GJ Melendez-Torres,[2] Sara Paparini,[1] Adam Fletcher,[2] Claire Stansfield,[3] James Thomas,[3] Rona Campbell,[4] Suzanne Taylor[1] and Chris Bonell[1]*

[1]Department of Social and Environmental Health Research, London School of Hygiene & Tropical Medicine, London, UK
[2]Development and Evaluation of Complex Interventions for Public Health Improvement, School of Social Sciences, Cardiff University, Cardiff, UK
[3]Evidence for Policy and Practice Information and Co-ordinating Centre, Social Science Research Unit, University College London Institute of Education, University College London, London, UK
[4]Development and Evaluation of Complex Interventions for Public Health Improvement, School of Social and Community Medicine, University of Bristol, Bristol, UK

*Corresponding author  chris.bonell@lshtm.ac.uk

**Background:** Schools struggle to timetable health education. Interventions integrating academic and health education to reduce substance use and violence offer promise. No current systematic reviews examine such interventions.

**Objectives:** To review evidence to explore the following questions: (1) what types of interventions integrating health and academic education in schools serving those aged 4–18 years have been evaluated? (2) What theories of change inform these interventions? (3) What factors facilitate or limit the successful implementation and receipt of such interventions, and what are the implications for the delivery of such implementations in the UK? (4) How effective are such interventions in reducing smoking and violence and the use of alcohol and drugs, and at increasing attainment? Does this vary by students' sociodemographic characteristics? (5) What factors appear to influence the effectiveness of such interventions?

**Data sources:** In total, 19 databases were searched from 18 November to 22 December 2015, updating searches for outcome evaluations for violence on 28 February 2018 and for substance use on 14 May 2018. References were extracted from included studies and authors contacted.

**Review methods:** Included studies reported on theories of change, and process or outcome evaluations of interventions that integrated academic and health education to reduce substance use and/or violence. References were screened on the title/abstract and then on the full report. Data extraction and appraisal used Cochrane, Evidence for Policy and Practice Information Centre and other established tools. Theories of change and process data were qualitatively synthesised. Outcome evaluations were synthesised narratively and meta-analytically.

**Results:** In total, 78,451 unique references were originally identified and 62 reports included. Search updates on 28 February and 14 May 2018 retrieved a further 2355 and 1945 references, respectively, resulting in the inclusion of six additional reports. Thirty-nine reports described theories, 16 reports (15 studies) evaluated process and 41 reports (16 studies) evaluated outcomes. Multicomponent interventions are theorised to erode 'boundaries' (strengthen relationships) between academic and health education, teachers and students,

behaviour in classrooms and in the wider school, and schools and families. Teachers, pro-social peers and parents are theorised to act as role models and reinforcers of healthy behaviours learnt in lessons. There was clear evidence that interventions are facilitated by supportive senior management and alignment with the schools' ethos, collaborative and supportive teaching environments, and positive pre-existing student, teacher and parent attitudes towards interventions. The barriers were overburdened teachers who had little time to both learn and implement integrated curricula. The strongest evidence for effectiveness was found for the reduction of substance use in school key stages (KSs) 2 and 3. For example, a meta-analysis for substance use at KS3 reported a mean difference of $-0.09$ (95% confidence interval $-0.17$ to $-0.01$). A meta-analysis for effectiveness in reducing violence victimisation in KS2 found no effect. There was mixed evidence for effects on academic outcomes, with meta-analysis precluded by methodological heterogeneity.

**Limitations:** Study quality was variable. Integration was sometimes not emphasised in theories of change.

**Conclusions:** These interventions are undertheorised but involve multiple forms of boundary erosion. There is clear evidence of characteristics affecting implementation. Interventions are likely to have the greatest impact on substance use. These programmes may be effective in reducing substance use but do not appear to reduce violence and findings on educational impacts are mixed.

**Future work:** Future evaluations should assess interventions with clearer theories of change and examine academic outcomes alongside violence and substance use outcomes.

**Study registration:** This study is registered as PROSPERO CRD42015026464.

**Funding:** The National Institute for Health Research Public Health Research programme.

# Contents

# List of tables

# List of figures

# List of abbreviations

| | |
|---|---|
| 4Rs | Reading, Writing, Respect and Resolution |
| A–B–C–D | Affective–Behavioural–Cognitive–Dynamic |
| ALPHA | Advice Leading to Public Health Advancement |
| ANOVA | analysis of variance |
| ASSIA | Applied Social Sciences Index and Abstracts |
| CI | confidence interval |
| DECIPHer | Development and Evaluation of Complex Interventions for Public Health Improvement |
| DRACON | DRAma = CONflict |
| DVD | digital versatile disc |
| EPPI-Centre | Evidence for Policy and Practice Information and Co-ordinating Centre |
| GRADE | Grading of Recommendations, Assessment, Development and Evaluations |
| I-LST | Infused Life Skills Training |
| IRR | incidence rate ratio |
| KAT | Kids and Adults Together |
| KS | key stage |
| LIFT | Linking the Interests of Families and Teachers |
| LST | Life Skills Training |
| MD | mean difference |
| NIHR | National Institute for Health Research |
| OR | odds ratio |
| PATHS | Promoting Alternative Thinking Strategies |
| PRISMA | Preferred Reporting Items for Systematic Reviews and Meta-Analysis |
| PSHE | personal, social and health education |
| RCT | randomised controlled trial |
| REC | Research Ethics Committee |
| RHC | Raising Healthy Children |
| SD | standard deviation |
| SE | standard error |
| SES | socioeconomic status |

# Plain English summary

Interventions that teach school students academic knowledge and skills alongside health knowledge and skills can aim to reduce student substance use and/or violence. They aim to equip young people with the skills and knowledge required to develop good relationships with teachers, parents and peers, and reduce risk-taking alongside the development of academic skills and knowledge. The authors of this report reviewed existing research on such interventions, exploring how they are expected to work, what factors affect their implementation and how they have an impact on substance use and violence. We defined what sort of evidence would be included in the review, carried out a comprehensive search and found 68 reports on interventions of interest, 41 of which provided evidence of effectiveness, most of them from the USA.

The review of theory about how such interventions are meant to work suggested that they aim to not only build links between health and academic learning but also build links between students and teachers within and beyond the classroom, as well as between students and their peers and parents. These links are supported through the development of students' social and emotional skills.

The review of factors affecting the delivery of integrated curricula emphasises the role of teachers, who need to have belief in the programmes, as well as enough time, resources and the capacity to deliver them. Furthermore, support from school administrators was very important.

The pooling of results from different studies of the effectiveness of integrated curricula suggest that these interventions may reduce substance use, particularly when they are implemented over longer periods of time. There is little evidence to suggest that they can reduce violence. There is inconclusive evidence around their ability to improve academic outcomes.

Currently, there is some evidence to support the implementation of integrated academic and health curricula for reduced substance use, but more needs to be learned about how these interventions work.

# Scientific summary

## Background and rationale

The review focused on substance use (i.e. alcohol consumption, smoking and drug use) and violence. The prevalence, harms and costs of these outcomes among young people mean that addressing them is a public health priority. Existing systematic reviews suggest that school curriculum-based health interventions can reduce alcohol consumption, smoking, drug use and violence, but in the UK these are increasingly difficult to deliver within constrained school timetables. In this context, schools may deliver health education in other subjects, integrating it with academic learning. Such interventions may either teach health education within other mainstream school subjects or provide specific health education lessons, ones that also provide teaching that covers academic, as well as health, knowledge and skills. This approach may allow for increased curriculum teaching time, be less prone to student resistance and prevention fatigue, and enable synergy and reinforcement between sessions provided in different subjects. However, existing interventions of this sort in the UK have not been informed by existing theory or evidence. Effects on substance use and violence are likely to be synergistic because each predisposes the other and has common risk factors.

No systematic review has examined evidence concerning interventions integrating health and academic education. Those exploring related interventions are dated and do not have comprehensive inclusion of integrated curricula. The marginalisation of student health and well-being education, especially in England, and the potential advantages of interventions integrating health and academic education to jointly achieve health and academic outcomes warrant an exploration of the available evidence.

## Aim and review questions

The aim was to systematically search for, appraise the quality of and synthesise evidence to address the following review questions:

1. What types of curriculum interventions that integrate health and academic education in schools and address substance use and violence have been evaluated?
2. What theories of change inform these interventions and what do these suggest about potential mechanisms and effects?
3. What characteristics of interventions, deliverers, participants and school contexts facilitate or limit successful implementation and receipt of such interventions, and what are the implications of these for delivery in the UK?
4. How effective are such interventions in reducing alcohol consumption, smoking, drug use and violence, and increasing academic attainment, when compared with usual treatment, no treatment or other interventions, and does this vary according to students' sociodemographic characteristics?
5. What characteristics of interventions, deliverers, school contexts and students appear to moderate or are necessary and sufficient for the effectiveness of such interventions?

## Methods

We carried out a multimethod systematic review of theories of change, process and outcomes of school-based curriculum interventions integrating health and academic education among students aged 4–18 years addressing substance use or violence. Academic education was defined as education in specific academic subjects, literacy, numeracy or study skills. The studies that were included addressed one or more of the

following primary review outcomes: smoking, alcohol use, legal or illicit drug use and violence (perpetration and victimisation). Academic attainment was also assessed as a secondary outcome. The review followed existing criteria for the good conduct and reporting of systematic reviews [Moher D, Liberati A, Tetzlaff J, Altman DG, Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Med* 2009;**6**:e1000097].

### *Searching electronic databases*

The search strategy involved terms concerning three core concepts: health education curricula (e.g. violence, smoking, drugs or alcohol education), integration with academic learning (e.g. integration within mathematics or literacy teaching), and population and setting (e.g. primary and secondary school-aged children). From 18 November to 22 December 2015, we searched the following databases: Applied Social Sciences Index and Abstracts (ASSIA), Australian Educational Index, BiblioMap (database of health promotion research), British Educational Index, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Database of Promoting Health Effectiveness Reviews, Dissertation Abstracts (UK theses, all dates; global theses 2010–15), Econlit, EResearch Index Citations, Health Technology Assessments, International Bibliography of the Social Sciences, MEDLINE, NHS Economic Evaluation Database, PsycINFO, Social Policy and Practice including Child Data & Social Care Online, Social Science Citation Index/Web of Knowledge and Trials Register of Promoting Health Interventions. We updated searches for outcome evaluations using PsycINFO and the Cochrane Central Register of Controlled Trials. Searches for outcome evaluations relating to violence were updated on 28 February 2018 and searches relating to substance use were updated on 14 May 2018.

### *Searching other resources*

The following 32 websites were searched to identify relevant studies: Cambridge Journals [URL: www.cambridge.org/core/ (accessed 12 January 2016)], Centers for Disease Control and Prevention: Smoking & Tobacco Use [URL: www.cdc.gov/tobacco/index.htm (accessed 12 January 2016)], Child and Adolescent Research Unit [URL: www.cahru.org/ (accessed 12 January 2016)], Childhoods Today [URL: www.childhoodstoday.org/ (accessed 12 January 2016)], Children in Scotland [URL: https://childreninscotland.org.uk (accessed 12 January 2016)], Children in Wales [URL: www.childreninwales.org.uk/ (accessed 12 January 2016)], Community Research and Development Information Service [URL: https://cordis.europa.eu/home_en.html (accessed 14 January 2016)], Database of Educational Research [Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)] [URL: https://eppi.ioe.ac.uk/webdatabases/SearchIntro.aspx (accessed 14 January 2016)], Drug and Alcohol Findings Effectiveness Bank [URL: https://findings.org.uk/ (accessed 14 January 2016)], Google [URL: www.google.com (accessed 14 January 2016)], Google Scholar [URL: www.scholar.google.com (accessed 14 January 2016)], Government of Wales [URL: http://gov.wales/?lang=en (accessed 18 January 2016)], Government of Scotland [URL: www.gov.scot/ (accessed 18 January 2016)], Joseph Rowntree Foundation [URL: www.jrf.org.uk/ (accessed 18 January 2016)], National Criminal Justice Reference Service [URL: www.ncjrs.gov/ (accessed 18 January 2016)], National Society of the Prevention of Cruelty to Children [URL: www.nspcc.org.uk/ (accessed 18 January 2016)], National Youth Agency [URL: https://nya.org.uk/ (accessed 18 January 2016)], National Institute for Health Research (NIHR) Clinical Research Network Study Portfolio [URL: www.nihr.ac.uk/research-and-impact/nihr-clinical-research-network-portfolio/ (accessed 19 January 2016)], Northern Ireland Executive [URL: www.northernireland.gov.uk/ (accessed 19 January 2016)], OpenGrey [URL: www.opengrey.eu/ (accessed 19 January 2016)], Personal Social Services Research Unit [URL: www.pssru.ac.uk/ (accessed 19 January 2016)], Project Cork [URL: www.dartmouth.edu/~cork/ (accessed 21 January 2016)], University College of London Institute of Education Digital Education Resource Archive [URL: http://libguides.ioe.ac.uk/dera (accessed 21 January 2016)], University of Illinois at Urbana Champaign [URL: http://illinois.edu/ (accessed 21 January 2016)], US Centre for Substance Abuse Prevention [URL: www.samhsa.gov/accessed (accessed 21 January 2016)], Social Issues Research Centre [URL: www.sirc.org/accessed (accessed 21 January 2016)], The Campbell Library [URL: www.campbellcollaboration.org/library.html (accessed 21 January 2016)], The Children's Society [URL: www.childrenssociety.org.uk/ (accessed 21 January 2016)], The Open Library [URL: https://openlibrary.org/ (accessed 22 January 2016)], The Schools and Students' Health Education Unit Archive [URL: http://sheu.org.uk/ (accessed 22 January 2016)], World Health

Organization International Clinical Trials Registry Platform [URL: www.who.int/ictrp/en/ (accessed 23 January 2016)] and Young Minds: Child & Adolescent Mental Health [URL: https://youngminds.org.uk (accessed 21 January 2016)].

### Study selection
Studies were screened by the title and abstract by four reviewers. Each reviewer initially screened sets of 50 of the same studies. A 90% agreement rate was required before proceeding to independent screening by the title and abstract. Full reports were obtained for studies not excluded by the title and abstract using the same process of piloting.

### Data extraction and management
References were stored in EPPI-Reviewer version 4.0 (Evidence for Policy and Practice Information and Coordinating Centre, University of London, London, UK) and data were extracted using coding tools for theory, process or outcome reports. Data extraction tools were piloted on five studies (two theory reports, two process evaluations and one outcome evaluation) and refined. For studies describing a theory of change, we extracted data on description of the theory of change, the rationale for integrating health and academic education, links to other theories and how the theory differs from others included in the study. For process and outcome evaluations, we extracted data on study location, intervention/components, description of integration, intervention development, timing of intervention and evaluation, target population, provider and provider organisation, research questions or hypotheses, timing of evaluation, sampling methods and sample size at baseline and follow-up, sociodemographic characteristics of participants at baseline and any follow-ups, and data collection and analysis.

For outcome studies, when additional data were needed to calculate effect sizes, we contacted authors for the relevant information. When authors did not provide the relevant information, we used the best approximation available.

### Quality appraisal
The quality of each study was independently assessed by two reviewers, with differences in opinion resolved by discussion without the need for recourse to a third reviewer. The quality of studies reporting on theory was assessed on clarity (of definition of constructs and pathways), plausibility (of pathways, the theory being informed by empirical evidence), testability (evidence of empirical testing), ownership (of theory by relevant stakeholders) and generalisability (of theory to different contexts with evidence of having done so).

The quality of process evaluations was assessed based on whether or not efforts had been made to increase rigour of data collection and data analysis, the extent to which the study findings were grounded in the data, the extent to which the study privileged the perspectives of youth participants, and the breadth and depth of the findings. Reviewers then judged both the reliability and the usefulness of the findings as low, medium or high.

Outcome evaluations were assessed for risk of bias in seven domains: sequence generation, allocation concealment, blinding, completeness of outcome data, whether or not clustering was accounted for, other sources of bias, and the suitability of the control group. Each study was then defined as having a low, high or unclear risk of bias.

### Synthesis of theoretical data
First, we synthesised theories of change for each individual intervention included in the review. Second, we synthesised theories across all interventions to explore points of reciprocal resonance, refutation and/or complementarity potentially leading to the development of a line-of-argument synthesis. This led us to employing a mix of methods: line-by-line coding and thematic synthesis for the 'within-intervention' theories and meta-ethnography for the 'across-intervention' theories.

### Synthesis of process data

Process evaluations reported qualitative, quantitative or mixed results and were synthesised qualitatively using thematic synthesis methods applied to any results.

### Synthesis of outcome data

We undertook both narrative synthesis and meta-analytic synthesis of the results of outcome evaluations. Our narrative synthesis included both end-point measurements and trajectory estimates for each intervention separately. Effect sizes from included study reports were converted into standardised mean differences (Cohen's *d*) using all available information as presented for each study. Effect estimates adjusted for covariates were used when these were presented alongside unadjusted estimates. In interpreting the results of meta-analyses, the standard rule for the interpretation of Cohen's *d* was followed: 0.2 is a small effect, 0.5 is a medium effect and 0.8 is a large effect. Negative effect sizes indicate a positive effect (e.g. a reduction in substance use). Data transformation and imputation were carried out as necessary and a multilevel meta-analysis with random effects was used at both the outcome and study level. A standard three-level model was used, with level one being the 'hypothetical' participants who contributed to the effect sizes, level two being the within-study outcome-specific effect size estimates with sampling error and level three being the 'between-study' level. A 'matrix' of key stage (KS) against type of outcome was created. Findings were then meta-analysed within each cell of the matrix where appropriate. For each model, an overall effect size was estimated and expressed as a standardised mean difference with a 95% confidence interval. $I^2$ was estimated at the study level using the variance components implied by the multilevel model.

### Stakeholder analysis

One-to-one consultations were conducted to reflect on the findings with policy and practice stakeholders. Young people were also consulted via the Advice Leading to Public Health Action young people's public health research advisory group based in the Centre for Development and Evaluation of Complex Public Health Interventions for Public Health Improvement. Views were sought regarding the potential feasibility and acceptance of integrated academic and health education within the UK. Emerging hypotheses were also explored, largely around implementation characteristics.

### Ethics considerations

This project was approved by the Research Ethics Committee (REC) of University College of London Institute of Education (ethics approval reference REC 746). The project complied with the Social Research Association's ethics guidelines and guidance from the National Co-ordinating Centre for Public Engagement.

## Results

### Included studies

Original searches identified 78,451 unique references from which 62 reports were included. Update searches retrieved an additional 2355 and 1945 references (on 28 February 2019 and 14 May 2018, respectively), yielding an additional six reports of outcome evaluations. Thirty-nine reports described theories, 16 reports (15 studies) evaluated process and 41 reports (16 studies) evaluated outcomes.

### What types of curriculum interventions that integrate health and academic education in schools and address substance use and violence have been evaluated?

Health curricula were either partially or fully integrated within an academic class. Fully integrated curricula use the same learning activities to achieve health and academic learning objectives (e.g. a programme that uses English literature lessons to teach themes about bullying, aiming to reduce both violence in children and improve literacy). Partially integrated programmes have separate learning activities that address health and academic learning objectives separately but within one overall package.

### What theories of change inform these interventions and what do these suggest about their potential mechanisms and effects?

The interventions within this review aimed to integrate and, thus, erode boundaries between health and academic education. Role-modelling and reinforcement of risk avoidance by teachers and pro-social peers promoted through interventions was important, particularly alongside the development of positive teacher–student and pro-social peer relationships (interpreted as erosion of boundaries between students and teachers). Theories of change also emphasised multilevel interventions with classroom work, supported by other components delivered at multiple levels (e.g. the overall school environment and the family). This was interpreted as erosion of boundaries between classrooms and schools and between schools and families. Such work was theorised to ensure that learning and reinforcement of positive behaviours occurred beyond the classroom. In turn, it was theorised that these interventions would provide students with various assets necessary to reduce engagement in substance use and violence as well as to increase academic attainment.

### What characteristics of interventions, deliverers, participants and school contexts facilitate or limit successful implementation and receipt of such interventions, and what are the implications of these for delivery in the UK?

Key facilitators of integrated health and academic curricula were supportive senior management, alignment of the intervention with the school's ethos, positive teaching environment and positive pre-existing student, teacher and parent attitudes towards interventions. Important barriers were overburdened teachers, with little time to both learn and implement integrated curricula. Reflections from stakeholders, as part of our consultation process, suggested a broad alignment with the above factors and the importance of government support for such programming, as well as having effective teacher training with ready-made resources that do not add to the teacher workload or prove burdensome in promoting good implementation in the UK. There were further comments about the differences in primary and secondary schools, with the general agreement that implementation would be more feasible and more logistically possible in primary schools. It was not possible to draw on the above factors to determine which interventions, reported on by studies included in this review, are most appropriate for the UK context.

### How effective are such interventions in reducing alcohol consumption, smoking, drug use and violence, and increasing academic attainment when compared with usual treatment, no treatment or other interventions, and does this vary according to students' sociodemographic characteristics?

The strongest evidence for the effectiveness of interventions integrating health and academic education was for the reduction of substance use in schools at KS2 and 3. A meta-analysis for the effectiveness of these interventions in reducing violence and victimisation in KS2 did not find an effect. It was not possible to undertake an analysis based on sociodemographic characteristics. There was mixed evidence about the effects of these interventions on academic outcomes, the reporting of which was generally poor.

### What characteristics of interventions, deliverers, school contexts and students appear to moderate or are necessary and sufficient for the effectiveness of such interventions?

Studies provided insufficient detail on such factors, precluding analysis.

## Conclusions

This form of intervention is undertheorised but involves multiple forms of boundary erosion. There is clear evidence of characteristics affecting implementation. Interventions are likely to have the greatest impact on substance use. These programmes may be effective in reducing substance use but do not appear to reduce violence and findings on educational impacts are mixed. These differences may simply reflect the particular studies reviewed or the differences in how open these outcomes are to modification among school-aged children.

## Study registration

This study is registered as PROSPERO CRD42015026464.

## Funding

# Chapter 1 Background

Parts of this chapter are reproduced from Bonell et al.[1] Contains information licensed under the Non-Commercial Government Licence v2.0.

## Description of the problem

This review focuses on substance use (i.e. alcohol consumption, smoking and drug use) and violence because these are important, intercorrelated outcomes that are addressed by interventions sharing common theories of change.[2–5] Alcohol has been suggested to be the most harmful substance in the UK.[6] Treating alcohol-related diseases costs the NHS in England an estimated £3.5B annually.[7] The total annual societal costs of alcohol use in England are estimated at £21B.[8] Alcohol-related harms are strongly stratified by socioeconomic status (SES).[9] Early initiation of alcohol use and excessive drinking are linked to later heavy drinking, alcohol-related harms[10,11] and poor health.[12] Alcohol use among young people is associated with truancy, exclusion and poor attainment, as well as unsafe sexual behaviour, unintended pregnancies, youth offending, accidents/injuries and violence.[13]

Preventing young people from taking up smoking is another key public health objective with 80,000 deaths due to smoking each year.[14] In 2005–6, smoking cost the NHS £5.2B and wider costs amounted to £96B.[15,16] A total of 40% of smokers start smoking in secondary school[17] and early initiation is associated with heavier and more enduring smoking and greater mortality.[18,19] Smoking among young people is a major source of health inequalities.[17]

Among UK 15- to 16-year-olds, 25% have used cannabis and 9% have used other illicit drugs.[18] Early initiation and frequent use of 'soft' drugs may be a potential pathway to more problematic drug use in later life.[20] Drugs such as cannabis and ecstasy are associated with an increased risk of mental health problems, particularly among frequent users.[20–23] Young people's drug use is also associated with accidental injury, self-harm, suicide[24–26] and other 'problem' behaviours.[27–30]

The other primary outcome of the review is violence. The prevalence, harms and costs of violence among young people mean that addressing this is a public health priority.[31,32] One UK study found that 10% of young people aged 11–12 years reported carrying a weapon and 8% admitted to attacking someone with intent to hurt them seriously.[33] By the age of 15–16 years, 24% of students reported that they have carried a weapon and 19% of students reported attacking someone with the intention to hurt them seriously.[33] There are also links between aggression and antisocial behaviours in youth and violent crime in adulthood.[34,35] As well as leading to further health inequalities, the economic costs to society of youth aggression, bullying and violence are high. For example, the total cost of crime attributable to conduct problems in childhood has been estimated at about £60B per year in England and Wales.[36]

In the UK, many schools are reducing the provision of lessons that address health[37–39] because schools now increasingly focus on narrow attainment targets and school inspections have only a limited focus on schools' promotion of student health and personal development.[40] In addition, in England, personal, social and health education (PSHE) is not a statutory subject.[39]

## Description of the intervention

Existing systematic reviews suggest that school curriculum-based health interventions can reduce alcohol consumption,[41] smoking,[42] drug use[43] and violence,[44–46] but these are increasingly difficult to deliver within constrained school timetables. In this context, many schools integrate health education with academic learning.[47] Such interventions may either teach health education within other mainstream school subjects

or provide specific health education lessons (ones that are also providing teaching that covers academic, as well as health, knowledge and skills). Even without the marginalisation of PSHE, this approach may be more effective because it could allow for increased curriculum teaching time,[47,48] it may be less prone to student resistance to health messages,[4] and it may enable synergy and reinforcements between sessions provided in different subjects.[2] On the other hand, integration may lead to a delivery of the intervention by staff unqualified or unwilling to address health, or to a delivery of the intervention with only a cursory treatment of health topics. Existing UK interventions use a range of innovative approaches to integrate health and academic education,[49,50] but have not been informed by existing theory or evidence. For example, the British Heart Foundation's 'Money to Burn'[49] and the Ariel Trust's 'Plastered'[50] interventions incorporate education about the respective risks of smoking and drinking into mathematics lessons. Money to Burn includes activities such as calculating how much a smoker would typically spend on cigarettes per day, per week or per year, and Plastered includes activities such as calculating units of alcohol consumed and presenting statistics on attitudes to alcohol as a basis for whole-class discussions about the harmful consequences of alcohol consumption.

The UK can learn from theory and evidence being generated in other countries. For example, the Reading, Writing, Respect and Resolution (4Rs)[3,51–57] intervention aims to integrate the learning of social and emotional skills with literacy skills for children in US elementary schools to reduce violence and antisocial behaviour. Lessons use a discussion of children's literature as a basis for education about managing emotions and negotiating disagreements non-violently. A randomised controlled trial (RCT) reported a significant reduction in aggression and improved academic attainment.[3,51,52]

In terms of theorised mechanisms, such interventions may either fully integrate health and academic education, so that the same learning activities address health and academic learning objectives together, or partially integrate health and academic education, so that separate learning activities aim to address health and academic learning objectives within one overall package. Such interventions could address substance use or violence by developing social and emotional skills, such as self-awareness, self-regulation, motivation, empathy and communication,[58] healthier social support or norms among students,[3,51,52] knowledge of the costs[49] and consequences[50] of substance use, media literacy skills to critique tobacco and alcohol advertising, and modifying students' social norms about substance use.[2,4,49,59,60] This category of intervention is likely to involve theories of change that are distinctive from conventional health education because such interventions aim to integrate health promotion into academic learning, and may aim to promote developmental cascades involving the interplay of cognitive and non-cognitive skills.[3,61] Effects on substance use and violence are likely to be synergistic because each predisposes the other and has common risk factors.[5]

## Rationale for the current study

No systematic review has examined theory or empirical evidence concerning interventions integrating health and academic education. The reviews cited above,[41–46] some of which are now quite old, are focused on school-based interventions, but the interventions included in the reviews are overwhelmingly those delivered in specific PSHE lessons or their international equivalents. Some of these reviews do include some interventions that integrate health and academic education, but they omit important studies and do not specifically analyse or draw conclusions about the effects of this category of intervention. Furthermore, these reviews have not synthesised evidence on intervention theories of change or process evaluations and have not considered components or mechanisms and, thus, cannot provide information about the intended mechanisms or the feasibility and acceptability of interventions, or their transferability to the UK.

These are important gaps to investigate because of the marginalisation of health education in the UK and the potential advantages of interventions integrating health and academic education (as well as the risk that these interventions might actually be less effective than conventional discrete health education), and the distinctive approaches and theories of change of this category of interventions. There is thus a

good rationale for a new systematic review focused on this category of interventions. This review focuses on substance use (i.e. alcohol consumption, smoking and drug use) and violence because our scoping searches and logic model (see appendices 1 and 2 of the protocol[1]) suggest that these interventions have the most potential in reducing the risk of these behaviours. Substance use and violence are closely intertwined,[5] and the theories of change underlying interventions addressing these outcomes appear to be similar. As explained in *Appendix 1*, which lists deviations from the protocol,[1] we originally intended to consider interventions that weaved health education into any existing academic lesson other than biology. However, we subsequently decided to include any interventions that integrated health and academic education regardless of whether the lessons in which they occurred were existing, timetabled academic lessons or new health lessons but with an academic component, and we decided against the exclusion of biology as an academic subject used to deliver health and well-being information. These modifications were made because it is the integration of health and academic education that is the focus of this review rather than the place in the school timetable in which this education occurs. This modification also gave our review greater international utility, as the division between health education and other lessons that exists in the UK is not as clear in many other countries (e.g. the USA). In the USA, health education is very commonly delivered in social studies lessons, which is an academic subject but in which there is often no integration of health and academic education.

## Review aim and objectives

To search systematically for, appraise the quality of and synthesise evidence on curriculum interventions that integrate health and academic education in schools to prevent substance use and violence. These aims have been addressed by focusing on the following objectives:

1. to conduct electronic and other searches
2. to screen found references and reports for inclusion in the review
3. to extract data from, and assess the quality of, included studies
4. to develop a typology of interventions and synthesise theories of change and process evaluations
5. to consult with policy/practice and youth stakeholders on the typology and theory/process synthesis to inform amendments and plans for synthesis outcome data
6. to synthesise outcome evaluation data and undertake meta-regression and/or qualitative comparative analyses
7. to draw on these syntheses to draft a report addressing our research questions
8. to consult with policy/practice and youth stakeholders on the draft report to inform amendments and dissemination.

## Review questions

The following questions were addressed:

1. What types of curriculum interventions that integrate health and academic education in schools and address substance use and violence have been evaluated?
2. What theories of change inform these interventions and what do these suggest about their potential mechanisms and effects?
3. What characteristics of interventions, deliverers, participants and school contexts facilitate or limit successful implementation and receipt of such interventions, and what are the implications of these for delivery in the UK?
4. How effective are such interventions in reducing alcohol consumption, smoking, drug use and violence, and increasing academic attainment when compared with usual treatment, no treatment or other interventions, and does this vary according to students' sociodemographic characteristics?
5. What characteristics of interventions, deliverers, school contexts and students appear to moderate or are necessary and sufficient for the effectiveness of such interventions?

# Chapter 2 Review methods

## About this chapter

Parts of this chapter are reproduced from Bonell *et al.*[1] Contains information licensed under the Non-Commercial Government Licence v2.0.

Parts of this chapter are reproduced or adapted from Bonell *et al.*[62] Contains information licensed under the Non-Commercial Government Licence v2.0.

This section outlines the methods used in this systematic review, which were described a priori in the research protocol.[1] The study is also registered with the PROSPERO registry of systematic reviews (reference number CRD42015026464) and is available from www.crd.york.ac.uk/PROSPEROFILES/26464_PROTOCOL_20160011.pdf (accessed 9 September 2018). Although there are no existing checklists for a complex, multimethod review such as the one undertaken, we have adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidance[63] and the PRISMA checklist can be found in *Appendix 2*.

### Design

The project is a multimethod systematic review of known existing research. This chapter describes the review methods. *Chapter 3* describes the flow of studies through the review as well as the characteristics of included studies (addressing research question 1), study reports and interventions. This is followed by four chapters presenting our various syntheses.

1. A thematic synthesis of the literature describing the theories of change of interventions included in this review (addressing research question 2).
2. A thematic synthesis of process evaluations exploring what factors facilitate or limit successful implementation and receipt (addressing research question 3).
3. A narrative synthesis and meta-analysis of the effects of interventions on substance use and violence outcomes (addressing research question 4). The evidence reviewed did not enable us to examine how effects varied according to students' sociodemographic characteristics.
4. An analysis of intervention components assessing key features of interventions integrating academic and health education (this reports a post hoc analyses not included in the original protocol[1]).

The evidence reviewed did not allow us to examine research question 5, concerning what characteristics of interventions, deliverers, school contexts and students appear to moderate or are necessary and sufficient for the effectiveness of such interventions. *Chapter 7* of this report draws on all the component syntheses to develop overall conclusions.

### Criteria for considering studies for this review

The criteria and definitions used for considering which studies to include in this review are outlined below. These inclusion criteria were operationalised into exclusion criteria to inform the screening of found studies (see inclusion and exclusion criteria in the protocol[1]). The results of this screening process are detailed in *Chapter 4*.

#### Types of participant

Studies were included where the majority of participants were children and young people aged 4–18 years attending schools. Interventions that targeted specific subpopulations defined in terms of health outcomes, such as autistic children, children with learning disabilities or children with known behavioural problems, were excluded. This is a clarification of the original protocol,[1] which was not explicit about including only

studies addressing general student populations (see *Appendix 1*). The rationale is that theories of change for interventions targeting subpopulations defined by specific health outcomes will differ from those addressing a general student population, and so would introduce inappropriate heterogeneity into the review.

## Types of intervention

School-based health curriculum interventions integrating health and academic education targeting young people aged 4–18 years were included. Academic education was defined as education in specific academic subjects, such as literacy, numeracy or study skills. It did not include education on social conduct in the classroom, relationships with peers or staff, attitudes to education, school or teachers or aspirations and life goals. Study reports needed to state explicitly that the intervention aimed to integrate health and academic education. To have been included, integration could have taken the form of either health education being fully integrated with academic education, so that the same activities address learning in each domain (i.e. the earlier example where education on smoking and alcohol was woven into existing timetabled mathematics lessons), or partial integration, when interventions combine activities separately addressing health and academic learning within one overall package (e.g. a social and emotional skills curriculum aiming to prevent violence that also included sessions aiming to improve literacy or study skills). We originally intended to exclude interventions integrating health education with biology but decided against this because it would have been an arbitrary exclusion (see *Appendix 1*). Interventions could be delivered by teachers or other school staff, such as teaching assistants, but could also be delivered by external providers, for example from the health, voluntary or youth service sectors. Our definition excluded interventions that were delivered in mainstream subject lessons but did not aim to integrate health and academic education, trained teachers in classroom management without student curriculum components or were delivered exclusively outside classrooms.

## Types of outcome

Studies that addressed one or more of the following primary review outcomes were included: smoking (e.g. salivary cotinine, carbon monoxide levels, self-reported use of cigarettes), alcohol use (e.g. self-reported alcohol consumption via questionnaires or diaries), legal or illegal drug use (e.g. self-reported drug use) and violence [self-reported violence perpetration (e.g. carried a weapon, got into a fight) or victimisation].

Types of outcome were informed by existing systematic reviews focused on substance use and violence among young people.[41–46] The outcome measures of these reviews could have drawn on dichotomous or continuous variables and self-report or observational data, they could have used measures of frequency (e.g. monthly, weekly or daily), the number of episodes of use or an index constructed from multiple measures. Alcohol measures could have examined alcohol consumption or problem drinking. Drug outcomes could have examined drugs in general or specific illicit drugs, including drug convictions. Measures of violent and aggressive behaviour could have examined the perpetration or victimisation of violence, including convictions for violent crime. Of note is that we included some outcomes that were a composite, such as items examining physical violence alongside non-physical (e.g. verbal or emotional) violence, but excluded other composite outcomes, such as physical violence alongside other items examining non-violent behaviours (e.g. damage to property). The section on the synthesis of outcome evaluations (see *Chapter 6*) describes how measures were combined.

Although not an inclusion criterion, we also assessed academic attainment as a secondary outcome [e.g. student standardised academic test scores, intelligence quotient (IQ) tests or other validated scales and school academic performance].

## Types of studies

To address research question 1, we have presented a summary of all included studies, regardless of evaluation type, in *Chapter 3*. To address research question 3, we included studies reporting on process evaluations this included studies reporting on the planning, delivery, receipt or causal pathways of interventions using

quantitative and/or qualitative data. These studies could have reported exclusively on process evaluations or on process alongside outcome data. To address research question 4, we included studies reporting on outcome evaluations using RCTs allocating schools, classes or individuals. Control participants could have been students, classes or schools allocated randomly to a control group in which no or usual school health and academic education was delivered or to a control group that included another 'active' intervention. To address research question 2, we drew on included process and outcome evaluations as defined above, which also included descriptions of intervention theories of change or logic models. To address research question 5, we aimed to draw on syntheses of all of the above study types.

### Date
Studies were not restricted by date of research or publication.

### Language
No language restrictions were placed on searches. Studies were not excluded based on language.

## Search strategy

### Database search strategy

#### Search terms
A sensitive search strategy was developed and tested by an experienced information scientist (CS). Test searches were undertaken in PsycINFO in order to further refine the search. Definitive searches were run between 18 November and 22 December 2015. Key search terms were determined by the review question and the inclusion criteria and were developed and tested against reports already known to the research team. The search strategy involved developing strings of terms and synonyms to reflect three core concepts in the review:

1.  health education curricula (e.g. violence, smoking, drugs or alcohol education)
2.  integration with academic learning (e.g. integration within mathematics or literacy teaching)
3.  population and setting (e.g. primary and secondary school-aged children).

These concepts were combined in searches as follows: concept 1 and concept 2 and concept 3.

It was identified that relevant research is described with very diverse terminology, and that a broad range of search terms would be needed to identify relevant literature. Thus, a range of free-text and database-controlled vocabularies was used. The free-text searches were used across all databases and the search strategies were adapted within each database to incorporate database-specific controlled terms. Full details of the search strings used for each database can be found in *Appendix 3*.

#### Databases
The following 19 electronic bibliographic databases were searched from inception to between 18 November and 22 December 2015, with the precise date of the search varying between databases (see *Appendix 3*). These databases were selected to draw on research literature from the fields of education, economics, social sciences and health and health behaviour. The list of databases that were originally intended to be searched was amended (see *Appendix 1*) on the advice, informed by initial pilot searches, of the information scientist (CS):

- Applied Social Sciences Index and Abstracts (ASSIA) via ProQuest
- Australian Educational Index via ProQuest
- BiblioMap (database of health promotion research) via Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)
- British Educational Index via EBSCO*host*
- Cochrane Central Register of Controlled Trials via The Cochrane Library
- Cochrane Database of Systematic Reviews via The Cochrane Library

- Database of Abstracts of Reviews of Effects via The Cochrane Library
- Database of Promoting Health Effectiveness Reviews via EPPI-Centre
- Dissertation Abstracts (UK theses, all dates; global theses 2010–15) via ProQuest
- Econlit via EBSCO*host*
- Educational Research Index Citations via EBSCO*host*
- Health Technology Assessment Database via The Cochrane Library
- International Bibliography of the Social Sciences via ProQuest
- MEDLINE via Ovid
- NHS Economic Evaluation Database
- PsycINFO via OVID
- Social Policy and Practice Including Child Data & Social Care Online via OVID
- Social Science Citation Index via Web of Knowledge
- Trials Register of Promoting Health Interventions via EPPI-Centre.

## Other search sources

The following 32 websites were also searched to identify relevant studies:

- Cambridge Journals [URL: www.cambridge.org/core/ (accessed 12 January 2016)]
- Centers for Disease Control and Prevention: Smoking & Tobacco Use [URL: www.cdc.gov/tobacco/index.htm (accessed 12 January 2016)]
- Child and Adolescent Research Unit [URL: www.cahru.org/ (accessed 12 January 2016)]
- Childhoods Today [URL: www.childhoodstoday.org/ (accessed 12 January 2016)]
- Children in Scotland [URL: https://childreninscotland.org.uk (accessed 12 January 2016)]
- Children in Wales [URL: www.childreninwales.org.uk/ (accessed 12 January 2016)]
- Community Research and Development Information Service [URL: https://cordis.europa.eu/home_en.html (accessed 14 January 2016)]
- Database of Educational Research (EPPI-Centre) [URL: https://eppi.ioe.ac.uk/webdatabases/SearchIntro.aspx (accessed 14 January 2016)]
- Drug and Alcohol Findings Effectiveness Bank [URL: https://findings.org.uk/ (accessed 14 January 2016)]
- Google [URL: www.google.com (accessed 14 January 2016)]
- Google Scholar [URL: www.scholar.google.com (accessed 14 January 2016)]
- Government of Wales [URL: http://gov.wales/?lang=en (accessed 18 January 2016)]
- Government of Scotland [URL: www.gov.scot/ (accessed 18 January 2016)]
- Joseph Rowntree Foundation [URL: www.jrf.org.uk/ (accessed 18 January 2016)]
- National Criminal Justice Reference Service [URL: www.ncjrs.gov/ (accessed 18 January 2016)]
- National Society of the Prevention of Cruelty to Children [URL: www.nspcc.org.uk/ (accessed 18 January 2016)]
- National Youth Agency [URL: https://nya.org.uk/ (accessed 18 January 2016)]
- National Institute for Health Research (NIHR) Clinical Research Network Study Portfolio [URL: www.nihr.ac.uk/research-and-impact/nihr-clinical-research-network-portfolio/ (accessed 19 January 2016)]
- Northern Ireland Executive [URL: www.northernireland.gov.uk/ (accessed 19 January 2016)]
- OpenGrey [URL: www.opengrey.eu/ (accessed 19 January 2016)]
- Personal Social Services Research Unit [URL: www.pssru.ac.uk/ (accessed 19 January 2016)]
- Project Cork [URL: www.dartmouth.edu/~cork/ (accessed 21 January 2016)]
- University College of London Institute of Education Digital Education Resource Archive [URL: http://libguides.ioe.ac.uk/dera (accessed 21 January 2016)]
- University of Illinois at Urbana Champaign [URL: http://illinois.edu/ (accessed 21 January 2016)]
- US Centre for Substance Abuse Prevention [URL: www.samhsa.gov/accessed (accessed 21 January 2016)]
- Social Issues Research Centre [URL: www.sirc.org/accessed (accessed 21 January 2016)]
- The Campbell Library [URL: www.campbellcollaboration.org/library.html (accessed 21 January 2016)]
- The Children's Society [URL: www.childrenssociety.org.uk/ (accessed 21 January 2016)]
- The Open Library [URL: https://openlibrary.org/ (accessed 22 January 2016)]

- The Schools and Students' Health Education Unit Archive [URL: http://sheu.org.uk/ (accessed 22 January 2016)]
- World Health Organization International Clinical Trials Registry Platform [URL: www.who.int/ictrp/en/ (accessed 23 January 2016)]
- Young Minds: Child & Adolescent Mental Health [URL: https://youngminds.org.uk (accessed 21 January 2016)].

Several of the above websites were not included in the original protocol[1] but were added to the list later on (see *Appendix 1*).

Dependent on the functionality of each website's interface, searches were undertaken using Google's (Google Inc., Mountain View, CA, USA) function to search a website, using the following search: '(drugs OR alcohol OR smoking OR violence OR bullying OR weapons) AND (primary school OR secondary school OR elementary school OR junior high OR high school) AND (lessons OR curriculum OR classes OR classroom OR health education OR health literacy OR health promotion)'. Websites were searched using the Google site function to enable more advanced search functioning than is typically possible using the search within each website. This search function enables all relevant webpages under a specific domain to be searched at the same time. Furthermore, results are returned within Google, which enables easy numeration of references found. For websites that are databases themselves (e.g. EPPI-Centre Database of Educational Research, NIHR Clinical Research Network Study Portfolio, OpenGrey, University College of London Institute of Education Digital Education Resource Archive and The Open Library) the following search was used: '(drugs OR alcohol OR smoking OR bullying OR violence) AND (primary school OR elementary school OR high school OR secondary school) AND (curriculum OR integration OR lessons OR education)'. All webpages were then explored for their relevance and included if they met our inclusion criteria.

Subject experts in the field who were known to us a priori were contacted to identify possible studies for inclusion, including unpublished or ongoing research. See *Appendix 4* for the experts contacted, a template of the e-mail sent to them and the included studies found via this search method. The reference lists of all included studies were searched for further relevant studies. The protocol stated that we would hand-search journals when these were not indexed on databases to be searched, but that published included study reports found only via reference checking.[1] No such studies were found and so no hand-searches were conducted.

### Update searches

We noted that all reports of outcome evaluations were retrieved from a combination of Cochrane Central Register of Controlled Trials and PsycINFO searches. Given the date of our original searches, we reran tailored versions of our original search strategy in both of these databases to search for outcome evaluations. Searches for outcome evaluations relating to substance use were undertaken on 14 May 2018, and searches for outcome evaluations relating to violence were undertaken on 28 February 2018. Findings from these searches were subject to the same study selection, data extraction and synthesis methods as findings from the main searches.

### *Information management*

All citations identified by our searches were uploaded and managed during the review process using the EPPI-Centre's specialist online review software, EPPI Reviewer version 4.0 (Evidence for Policy and Practice Information and Coordinating Centre, University of London, London, UK). This software records the bibliographic details of each study, where studies were found and how, reasons for their inclusion or exclusion, descriptive and quality assessment codes, text about each included study and the data used and produced during synthesis. The software also enabled us to store and track electronic documents [e.g. portable document files (PDFs)].

## *Study selection*

An exclusion criteria worksheet was prepared, informed by our inclusion criteria (see above) and with guidance notes. For the screening of references identified from electronic databases, this was piloted by four reviewers (CB, TT, GMT and AF) who screened 50 references in pairs on title and abstracts. Pilot screening results were discussed by pairs of reviewers involved in screening to ensure consistency in applying the criteria. This process was invaluable because, despite being guided by clear inclusion criteria, decisions about whether or not an intervention aimed to integrate health and academic education were not always easy. The definition of integration in terms of full and partial integration of health and academic education was also helpful in making clear decisions about which studies to include or exclude. A 90% agreement rate was required before proceeding to screening by single reviewers, with each reviewer screening discrete subsets of the full set of references. Full reports were obtained for those references judged as meeting our inclusion criteria based on title and abstract, or for which there was insufficient information from the title and abstract to judge inclusion or exclusion. These reports were then screened to determine inclusion. The reviewers piloted the procedure by screening full reports and working in pairs screening 50 reports each and discussing any differences in opinion. A 90% agreement rate was required before proceeding to independent screening of the full set of references. The principal investigator (CB) reviewed all studies identified as potentially includable in the review as a final check to determine inclusion, identify which review question they answered and group multiple reports from the same study. Citations identified from websites were screened online based on their title, title and abstract or full text, when available. Potentially includable studies were cross-referenced with the electronic searches imported to EPPI-Reviewer to identify any unique references. As is customary with searching of this type, only the included references, not the excluded ones, were recorded.

## *Data extraction*

### Tools

Data were extracted using coding tools developed for the review components relating to each review question (see *Appendices 6–8*). Each tool drew on and supplemented the codes used in the EPPI-Centre classification system for health promotion and public health research.[62] For studies describing a theory of change,[3] we extracted data on the description of the theory of change, rationale for integrating health and academic education, links to other theories and how the theory differs from others included in the study. For process and outcome evaluations, we extracted data using a modified version of an existing tool,[62] including items on study location, intervention/component description, description of integration, intervention development, timing of the intervention and evaluation, target population description, provider and provider organisation characteristics, research questions or hypotheses, timing of the evaluation, sampling methods and sample size at baseline and follow-up, sociodemographic characteristics of participants at baseline and any follow-ups, and data collection and analysis. After piloting and refinement, two reviewers working independently extracted study reports before meeting to agree on coding. For studies reporting on outcome evaluations, we also extracted data on research design, the nature of the control group(s), the unit of allocation, the generation and concealment of the allocation, blinding, baseline equivalences between control and intervention groups, the adjustment/control of clustering and confounding, the use of intention-to-treat analyses, outcome measures and the evidence of reliability and validity, and effect sizes overall and by age, sex, SES, and ethnic subgroup.

### Data extraction process

Data extraction tools were piloted on five studies (two theory reports, two process evaluations and one outcome evaluation). Reviewers met to compare extraction and identify any differences that might inform refinements of the coding tools or how these were applied. All study reports were then extracted by two reviewers working independently in parallel, before meeting to discuss and agree on their coding to ensure quality and consistency in their interpretations.

## Missing data

When additional data were needed to calculate effect sizes, we contacted the authors for the relevant information. If authors did not provide the relevant information, we used the best approximation available, generally by using other information from within the same study (see *Appendix 8* for a template of the e-mail sent to authors to obtain additional data).

### *Quality assessment*

### Quality assessment process

The quality of each study was independently assessed by two reviewers with differences in opinion resolved by discussion, without the need for recourse to a third reviewer, except for reports of theory. For theory reports, there were frequent differences of opinion. Discussion between the reviewers concluded that these differences were inevitable and could not be addressed through consultation with a third reviewer. Despite our best efforts in developing quality criteria informed by previous research (see *Quality assessment tool for theory studies*), some criteria could not be applied to make objective judgements. As a result, for theories of change we decided to report each reviewer's independent judgements (see *Appendix 9*).

### Quality assessment tool for theory studies

The quality of studies reporting on theory was assessed using a tool adapted from a previous review[64] informed by other recent work on theory synthesis.[65] Quality was assessed in terms of:

1. clarity

   - constructs defined
   - clear pathways from intervention inputs to outcomes

2. plausibility/feasibility

   - theory is logical with plausible pathways from intervention inputs to behaviour change
   - theory is supported by existing empirical evidence

3. testability

   - evidence of empirical testing of theory

4. ownership

   - theory has been developed with practitioners
   - theory has been developed with community members

5. generalisability

   - theory is presented as generally applicable to different contexts
   - theory describes how it is applicable to different contexts
   - authors present empirical evidence of the generalisability of the theory.

### Quality assessment tool for process evaluations

Process evaluations were assessed using the standard Critical Appraisal Skills Programme and EPPI-Centre tools[66] developed to assess qualitative studies and assess the rigour of whether or not the sampling strategy was indicated, the data collection methods were indicated (including any statements around increasing the rigour of data collection), the methods of data analysis were indicated (including any statements around efforts made to improve reliability of findings and reduce bias) and the extent to which the study findings were grounded in the data, privileged the perspectives of youth participants and the

breadth and depth of findings. The tool was used in this review to assess process evaluations regardless of whether these drew on qualitative or quantitative data, because the criteria were judged applicable to both. A final step in quality assessment was to assign studies two types of 'weight of evidence'. First, reviewers assigned a weight (i.e. low, medium or high) to rate the reliability or trustworthiness of the findings (i.e. the extent to which the methods employed were rigorous/could minimise bias and error in the findings). Second, reviewers assigned an additional weight (i.e. low, medium or high) to rate the usefulness of the findings for shedding light on factors relating to the review questions. Guidance was given to reviewers to help them reach an assessment on each criterion and the final weight of evidence. To be judged as highly reliable, studies needed to have taken steps to ensure rigour in at least three of the first four criteria. Studies were judged as medium when scoring only two out of the first four criteria and low when scoring only one or none out of the first four criteria. To achieve a rating of 'high' on usefulness, studies needed to be judged to have privileged the perspectives of participants and to present findings that achieved both breadth and depth. Studies that were rated as 'medium' on usefulness only partially met this criteria, and 'low' rated studies were judged to have sufficient but limited findings. Quality was used to determine the qualitative weight given to findings in our synthesis, with none of the themes represented solely by studies judged as low on both dimensions.

## Quality assessment tool for outcome evaluations

Outcome evaluations were assessed for risk of bias using the tool modified from the questions suggested in the Cochrane Handbook for Systematic Reviews of Interventions.[67] For each study, two reviewers independently judged the likelihood of bias in seven domains: sequence generation, allocation concealment, blinding (of participants, providers or outcome assessors), completeness of outcome data, whether or not clustering was accounted for, other sources of bias and the suitability of the control group. Each study was subsequently allocated a score of 'high risk', 'low risk' or 'unclear risk' within each domain. We used the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach as described in the Cochrane Handbook for Systematic Reviews of Interventions[67] to present the quality of evidence (see *Table 3*). The downgrading of the quality of a body of evidence for a specific outcome was based on five factors: limitations of the study, indirectness of the evidence, inconsistency of results, imprecision of results and publication bias. The GRADE approach specifies four levels of quality (i.e. high, moderate, low and very low).

## *Synthesis of results*

### Intervention descriptions

We first produced a basic descriptive categorisation of interventions based on a narrative review of intervention descriptions in all study reports in order to report on their targeted school level, providers, components, the subject within which the intervention is delivered and the extent of integration of health and academic education.

### Theories of change

We then synthesised theories of change. There is a growing literature on synthesising theories of change of interventions. Methods of theory synthesis tend to draw on methodological approaches to synthesising empirical qualitative data (or 'qualitative synthesis').[68] Methods of qualitative synthesis, in turn, draw on methods of primary qualitative data analysis.[69] Broadly, common principles and practices apply to different approaches to qualitative analysis and synthesis, including data immersion, an emphasis on depth, iterative coding, triangulation among multiple researchers, and the purposive inclusion of 'deviant' cases in the analysis.

There are several examples of systematic reviews and syntheses of theories of change within the public health literature.[70] There are challenges when applying methods used to synthesise empirical qualitative data, which are generally context specific, to the synthesis of theory, which is generally abstracted from context.[71] However, we considered that synthesis could be achieved by treating theory data as primary data in itself. We considered that synthesising theories of change for interventions that have similar premises and similar stated outcomes could also be conducted through meta-ethnography.[71,72]

For the theory synthesis reported here, we noted that two levels of analysis would be possible and useful. First, it was possible to synthesise theories of change for each individual intervention included in the review. These may have been reported on only once or by multiple authors and across multiple reports. Second, it was possible to synthesise theories across all interventions to explore points of reciprocal resonance, refutation and/or complementarity potentially leading to the development of a line of argument 'representing different stages along the same causal pathway'.[73] This led us to employ a mix of methods: line-by-line coding and thematic synthesis[69] for the 'within intervention' theories, and meta-ethnography for the 'across-intervention' theories. Specific methods are detailed below.

For the thematic synthesis of individual intervention theories, two reviewers first read and reread two reports deemed by two reviewers to be of high quality (i.e. having quality scores > 50%) (see *Appendix 9*), by different authors and focused on the same intervention. Line-by-line codes were applied and memos were written to identify and explain the content of the description of theory. Codes were then grouped, organised into frameworks and exchanged and compared between the reviewers to develop an overall set of codes (see *Appendix 10* for an example of this). This set of codes was then applied to any other study reports for the intervention in question, keeping track of and comparing any modification with the coding framework made as a result of the coding of subsequent reports. Having judged that this piloted process was appropriate, it was then repeated for each intervention.

In this way, we synthesised theories for each individual intervention. Thematic synthesis identified commonalities, differences of emphasis and contradictions within the single or multiple reports for each individual intervention theory of change, drawing on the coding described above.[3] Any apparent inconsistencies were resolved by discussion. A third reviewer helped achieve reconciliation when necessary. When only one report described the theory of change for an intervention, that was taken to represent the theory of change for that intervention. This process enabled us to develop summaries of the theories of change for each intervention, describing intervention inputs, mechanisms of change, underlying assumptions, and proximal and distal outcomes. These individual theories of change were undertaken as a preliminary step, prior to undertaking an across-intervention synthesis, and are not presented in this report.

For meta-ethnography across intervention theories, we used a meta-ethnographic approach to synthesise across individual intervention theories of change to develop an overarching theory of change for the overall category of interventions that integrate health and academic education to prevent substance use and violence (see *Appendix 11*). We determined that meta-ethnography was an appropriate method for this synthesis given it is commonly applied to synthesising areas of literature that are diverse yet share commonalities, as we judged was the case with the included theories of change for interventions integrating health and academic education.[68]

In this approach we considered the key concepts extracted during the coding exercise as our primary data, treating these concepts as 'first-order constructs'.[74] These were considered first-order constructs because they represented the authors' key theoretical concepts, and so were distinctive from authorial points within an empirical study report, which are rightly regarded as second-order constructs.[70,75] We then generated second-order concepts that were our interpretations of authors' views.[71,72,76] Finally, we developed third-order constructs through a process of reciprocal translation (a dynamic, iterative process during which concepts from each of the previous syntheses were 'translated' into one another). We then adapted an approach from Britten *et al.*'s[76] worked example of meta-ethnography to build up the line of argument of the overall synthesis of the theory of change for the interventions included in our systematic review.

Developing a line of argument was challenging because intervention theories in many cases described similar notions but offered limited explanations for key concepts and assumptions. In developing our third-order constructs and overarching synthesis, it was helpful to use an existing theoretical framework to identify commonalities across what initially might have appeared to be a set of disparate concepts. Informed by the notion of 'boundary erosion' proposed by Markham and Aveyard[77] as a key mechanism by which schools may promote student health, we integrated concepts arising from the individual syntheses

into a set of third-order constructs and so developed an overall line of argument. A few themes did not initially fit with the concept of boundary erosion as presented by Markham and Aveyard.[77] This stimulated us to refine and expand the concept of boundary erosion so that it could encompass these apparently divergent themes. These are presented alongside our line of argument synthesis in *Appendix 11, Table 14*.

## Process evaluations

Process evaluations reported qualitative, quantitative or mixed results. Process evaluations were synthesised qualitatively using thematic synthesis methods[68,78] applied to both qualitative and quantitative results. Although some quantitative studies did examine correlates of implementation, these findings were too heterogeneous to meta-analyse statistically. Instead, textual reports of quantitative results were subject to thematic synthesis after first checking that they were consistent with the quantitative data presented in study reports.

Thematic synthesis proceeded via the following process: studies were first read and reread by two reviewers and the two reviewers then carried out line-by-line coding of process data, drawing inductive codes from the process data itself (see *Appendix 12* for a coding template for all process studies). Coding focused on textual reports that included verbatim qualitative data excerpts and author interpretations of these.

Reviewers wrote analytic memos throughout to describe emerging 'meta-themes'. Each reviewer developed an emerging coding structure of hierarchically arranged codes applied in the course of the analysis. The two reviewers then compared their coding to agree on a common structure that formed the basis for the synthesis. This approach to coding was first piloted on two studies before proceeding, without modification, to the remainder of the reports. As the overall analysis was developed, the reviewers referred to tables summarising the methodological quality of each study to ensure that the synthesis reflected study quality.

## Outcome evaluations

We undertook both narrative and meta-analytic syntheses of the results of outcome evaluations. Our narrative synthesis included both end-point measurements and trajectory estimates. Many evaluations compare participants on differences in change over time, but meta-analytic methods have not yet been developed for these models. Thus, we narratively synthesised trajectory evidence from interventions separately.

Effect sizes from included study reports concerning substance use (i.e. smoking, alcohol or drugs) or violence, as defined in the protocol,[1] were extracted into a Microsoft Excel® (Microsoft Corporation, Redmond, WA, USA) spreadsheet and converted into standardised mean differences (MDs) (Cohen's *d*) using all available information as presented for each study. Because all evaluations were cluster randomised trials, some baseline imbalance on individual participant characteristics was likely. Thus, we used effect estimates adjusted for covariates when these were presented alongside unadjusted estimates. In interpreting the results of meta-analyses, we followed the standard rule for the interpretation of Cohen's *d*: 0.2 is a small effect, 0.5 is a medium effect and 0.8 is a large effect. Negative effect sizes indicate a positive effect (e.g. a reduction in substance use).

Because of the variation in reporting across studies, some degree of data transformation and imputation was necessary (see *Appendix 13*). Odds ratios (ORs) were converted to standardised MDs assuming the logistic transformation. When appropriate, we used the test statistic values from *t*-tests and *F*-tests to estimate the standardised MD. Clustering was accounted for when possible and necessary, using information provided in the evaluation and through contact with authors.

Most studies reported several substance use and violence outcomes at several measurement time points. As indicated in the protocol,[1] we used multilevel meta-analysis as set out by Cheung[79] and Van den Noortgate *et al.*[80] with random effects at both the outcome and study level. Multilevel meta-analysis accounts for dependencies between outcomes from the same study by partitioning the variance ($\tau^2$) between outcomes into a within-study and a between-study level. The final effect size estimate includes all of the information that the

multiple effect size estimates contribute while correcting for the non-independence of multiple effect size estimates from each study.

A standard three-level model was used, with level one being the 'hypothetical' participants who contributed to the effect sizes, level two being the within-study outcome-specific effect size estimates with sampling error and level three being the between-study level. Because several evaluations were reported as multiple papers, the between-study level does not reflect clustering by paper but rather by evaluation.

We aimed to estimate several different models. Because interventions are often measured at several time points in the developmental trajectory, we created a 'matrix' of key stage (KS) (*Table 1*) against type of outcome. We then meta-analysed findings within each cell of the matrix when appropriate (e.g. substance use for students in KS3). We set out to estimate models for substance use, violence perpetration and violence victimisation. Within substance use, we considered omnibus outcomes (e.g. count of substances used, any substance use), alcohol outcomes, smoking outcomes and illicit drug use outcomes separately, and then as a combined model to examine the global impact of these interventions on substance use.

For each model, we estimated an overall effect size expressed as a standardised MD with a 95% confidence interval (CI). We estimated $I^2$ at the study level using the variance components implied by the multilevel model. Interpretation of $I^2$ at the level of the study is most comparable to the interpretation of $I^2$ in 'standard' meta-analyses that include one effect size per study.

We intended to estimate meta-regression models to examine how intervention effects varied by participant SES, sex and ethnicity, to examine how intervention effects varied by area deprivation and in order to test hypotheses on other moderators of effects. These other hypotheses were derived from the syntheses of theory and process evaluations, and consultations with young people and policy/practitioner stakeholders. However, such analyses were not possible because of the absence of meaningful heterogeneity in effects between studies as well as the lack of consistency of reporting of subgroup effects within studies. We also intended to run a qualitative comparative analysis to examine the causal combinations of conditions that predict intervention effectiveness. However, this was not possible because of the generally poor description of interventions. We also did not find sufficient studies (≥ 10 per outcome) to draw funnel plots to assess the presence of possible publication bias. Full details of these methods may be found in our protocol.[1]

### Note on terms used in the synthesis of outcome evaluations

Included interventions were conducted in a diversity of settings. For the sake of clarity, we used intervention descriptions to map when interventions were implemented in children's educational progression to the UK system (i.e. years 1 to 13). Because included evaluations often tested multiyear interventions, we generally describe points of follow-up in terms of when measurements occurred relative to the start of the intervention. For example, a measurement taken in the first summer term after an intervention's start would be described as 'at the end of the first intervention year', and a measurement taken at the start of the school year following the initiation of the intervention would be described as 'at the start of the second intervention year'. For each set of outcomes synthesised, we present a schematic depicting when interventions were implemented and when outcome measurements were presented in evaluations.

**TABLE 1** Key stages by year of schooling

| KS | Years included |
|----|----------------|
| 1 | 1–2 |
| 2 | 3–6 |
| 3 | 7–9 |
| 4 | 10–11 |
| 5 | Sixth form (years 12–13) |

### *User involvement*

We conducted two sets of one-to-one consultations to reflect on our findings with young people and with policy and practice stakeholders. Individual consultation was used instead of group meetings because of the impossibility of finding a date when all could attend a meeting. The first set of consultations was conducted from December 2016 to June 2017. As per the protocol,[1] we aimed to discuss the validity of our typology of interventions and synthesis of process evaluations and identify the feasibility and acceptability of such interventions in the UK. Our original intention was to then use these discussions to determine which specific interventions should be included in a secondary analysis of outcome data focused on interventions most relevant for the UK. However, it was clear from our typology of interventions and synthesis of process evaluations, as well as from consultation with stakeholders, that it was not possible to identify a discrete subset of interventions that were relevant to the UK. All interventions were potentially relevant to the UK with adaptation. The adaptations required would concern the detail of the intervention materials rather than the overall intervention approaches and theories of change.

Stakeholders' views were sought about the potential feasibility and acceptance of integrated academic and health education within the UK (see *Appendix 14*). We asked the following general questions:

1. Could this type of intervention be delivered within the UK?
2. If so, which intervention characteristics (e.g. delivered within primary or secondary schools, delivered within specific academic classes, integration being 'full' or 'partial', intervention being facilitated by teachers, etc.) would be the most appropriate?
3. Do you think schools would be receptive to this type of intervention in the UK?
4. What factors would enable or inhibit these types of interventions in the UK?

In addition, based on some preliminary findings from the review, as well as on the review team's previous experiences, we explored the following hypotheses with stakeholders:

1. Interventions of this sort will, in general, be more feasible in primary rather than secondary schools because the timetable is more flexible and the emphasis is on core skills (e.g. literacy) that such interventions often address.
2. Interventions will be more feasible when they target an academic subject that is part of the national curriculum for students of that year group.
3. Interventions that aim to integrate health with academic subjects that not all schools or students will study will have much less reach (e.g. drama).
4. Interventions that focus on students beyond year 9 will be less attractive to schools because they will be perceived as reducing curriculum time for exam preparation.
5. Interventions that aim to fully integrate health into existing academic education will be more feasible than those which provide discrete new curricula that include separate health and academic components.
6. Interventions that are to be delivered in regular, timetabled health education lessons (e.g. PSHE) will be less feasible because many schools, especially at the secondary phase, do not run such lessons.
7. Interventions that include the whole school alongside classroom elements will be feasible in principle but the more complex the school-level components are the less likely it is that they will be implemented with fidelity.
8. Intervention components that aim to reach out to parents will generally not be well delivered, especially at the secondary level.
9. Interventions that require teachers to attend training away from school for > 1–2 days will not be feasible because schools will be reluctant to release teachers for this amount of time.

The second round of consultations occurred in September/October 2017, with the aim of reviewing the validity and usefulness of our syntheses to inform how our research outputs are structured and disseminated. Participants were given a summary of study findings (see *Appendix 14*), inclusive of outcome evaluation synthesis results. Furthermore, we discussed stakeholder views around next steps in terms of knowledge translation, replication studies and the creation of new interventions.

Consultations involved the following adult policy stakeholders: the Ariel Trust (Paul Ainsworth), Mentor UK (Michael O'Toole), Public Health Wales (Mary Charles), the School Health Research Network (Joan Roberts), Public Health England (Claire Robson), the PSHE Association (Jonathan Baggaley), the Association for Young People's Health (Ann Hagell and Emma Rigby), London Fields Primary School (Sindee Bass), St. Saviour's Church of England Junior School (Nick Bonell), Barnhill Community High School (Greig Pilkington) and Newstead Wood School (Jonathan Lewis).

Young people were consulted via the Advice Leading to Public Health Advancement (ALPHA) young people's public health research advisory group based in the Centre for Development and Evaluation of Complex Interventions for Public Health Improvement (DECIPHer) for Public Health Improvement, a collaboration between the universities of Cardiff, Bristol and Swansea. Ten young people were consulted, half of whom were female and half of whom were male, and the age range was from 14 to 18 years.

### Ethics arrangements

This project was approved by the Research Ethics Committee (REC) of University College of London Institute of Education (ethics approval reference number REC 746). The project complied with the Social Research Association's *Ethical Guidelines*[81] and guidance from the National Co-ordinating Centre for Public Engagement.[82]

# Chapter 3 Included studies

## About this chapter

Parts of this chapter are reproduced or adapted from Bonell et al.[62] Contains information licensed under the Non-Commercial Government Licence v2.0.

This chapter reports the results of our systematic search and screening process, and gives a brief overview of the included studies, study reports and interventions. It categorises interventions in order to address research question 1.

## Results of the search

A total of 78,451 references were identified from the searches. Of these, 1472 (2%) references were identified as duplicates. The remaining 76,979 references were screened on title and abstract and, of these, 76,277 (99%) were excluded using the criteria listed in the protocol.[1]

When piloting the process for screening on title and abstract, initial screening agreements between reviewers were consistently < 90% on whether or not a study should be excluded. Agreement was lower on the question of which particular criterion should be cited in excluding a particular reference, varying from 28% to 62% among six different pairs of reviewers. Discussion between reviewers established that this reflected the multiple criteria that could be used to exclude many studies and, therefore, a somewhat arbitrary choice of which particular criterion to cite in each case. Given that agreements were < 90% on whether to exclude or not, we moved to a system of one reviewer independently screening each reference, as set out in our protocol.[1]

Of the 702 references that were not excluded after screening on title and abstract, we were able to obtain the full-text reports of 690 (98%) references, the remainder not being accessible online or through interlibrary loans. Piloting the application of the same exclusion criteria used at the title and abstract screening stage on the full reports, screening between two pairs of reviewers also reached < 90% agreement on whether or not a study should be excluded, and so an individual reviewer (TT) moved to independent screening of documents, consulting with the principal investigator (CB), when necessary. CB made a final check of all the studies included. This procedure led to a further 628 studies being excluded at this stage in the review screening process.

The remaining 62 reports[3,51–57,60,83–135] deemed eligible for inclusion in the review were coded according to the review question they answered. Sixteen distinct interventions were examined by studies included in the review. Only three interventions had only one report,[108,133,135] and 13 interventions were described in the remaining 59 reports.[3,51–57,60,83–107,109–132,134] Furthermore, 27 study reports provided answers to more than one review question.[3,52,53,57,84,86–88,92–98,100,103,107,108,110,113,118,125,127,133–135] In presenting the number of studies and study reports, we have not double-counted those that address more than one of our review questions. When appropriate, we clarify whether a study or report addressed more than one of our review questions. Figure 1 summarises the flow of references, reports and studies through the review, providing a breakdown of the exclusion criteria at both title and abstract and full document stages and the number of studies included in each synthesis. Table 2 provides an overview of the interventions included in this review that were subject to process or outcome evaluation.
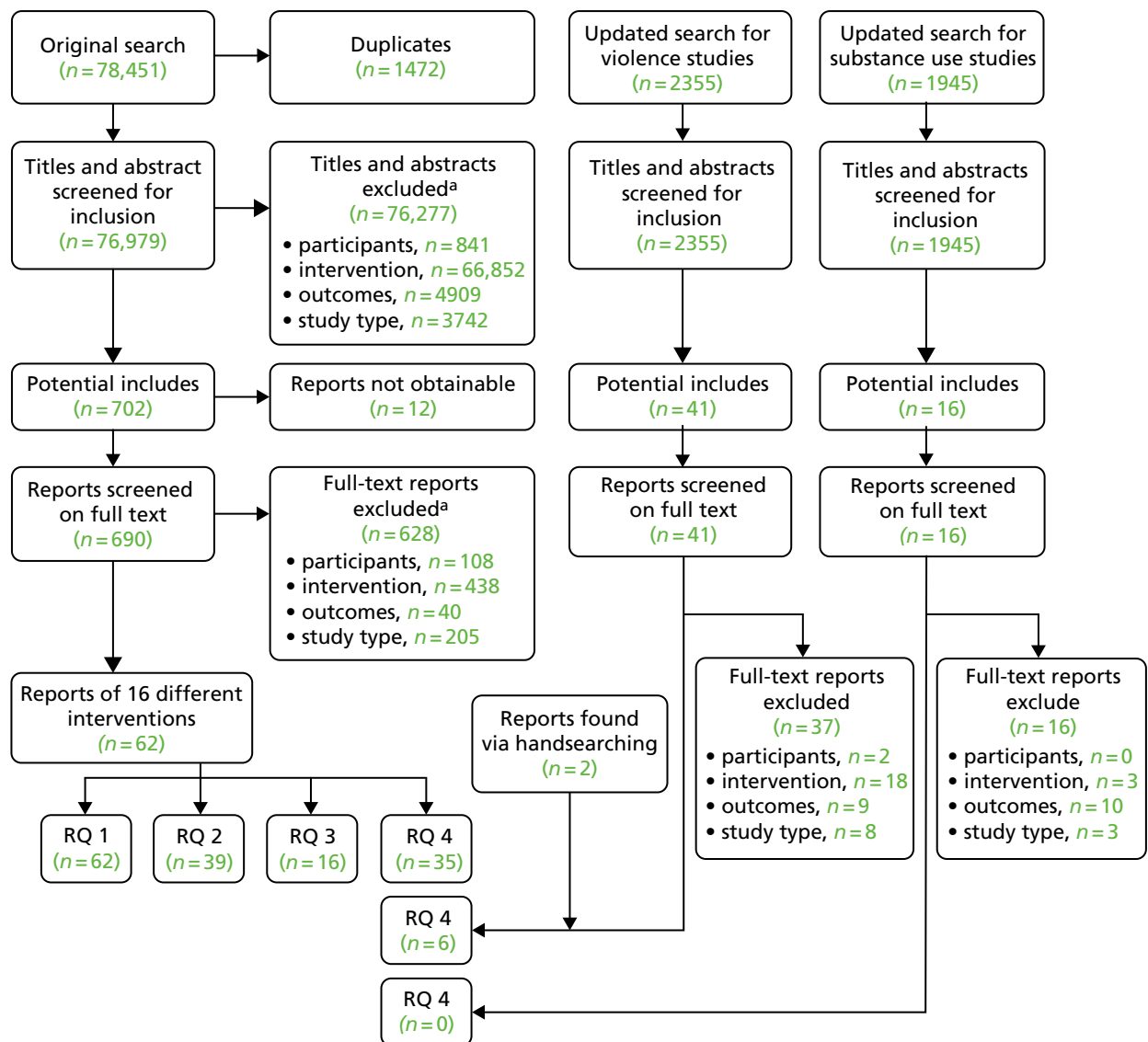
**FIGURE 1** Flow of studies in the review: main searches. RQ, research question. a, Note that these do not sum to 628 references, as some studies were excluded based on more than one criterion.

Updated searches for violence outcomes found a total of 2355 records, of which 41 were screened at full text. Six records relating to three evaluations were included. Updated searches for substance use outcomes found a total of 1945 records, of which 16 were screened at full text. No records were included from this update.

## Interventions evaluated

Overall, 19 interventions were studied in 68 reports,[3,51–57,60,83–137,139–141] of which one[55] reported findings about three separate interventions. The 19 interventions and the corresponding reports for each are summarised below according to the research question that the study helped to answer. These are summarised as 'included theory studies' that helped to answer research question 2, 'included process evaluations' that helped to answer research question 3 and 'included outcome evaluations' that helped to answer research question 4. All reports helped to answer research question 1. See *Appendix 15* for a listing of all included reports by the research question they answer.

**TABLE 2** Overview of reports included in this review

| Interventions examined in the review | Included theory studies (author, year) | Included process evaluations (author, date, location) | Included outcome evaluations (author, year, location) |
|---|---|---|---|
| 4Rs | Aber et al., 2011[53] | Sung, 2015[57] (New York, NY, USA) | Aber et al., 2011[53] (New York, NY, USA) |
| | Brown et al., 2010[54] | | |
| | Flay et al., 2009[55] | | Jones et al., 2010[52] (New York, NY, USA) |
| | Jones et al., 2008[56] | | Jones et al., 2010[51] (New York, NY, USA) |
| | Jones et al., 2010[52] | | |
| | Jones et al., 2011[3] | | Jones et al., 2011[3] (New York, NY, USA) |
| | Sung, 2015[57] | | |
| Bullying Literature Project | Couch, 2015[96] | None | Couch, 2015[96] (California, CA, USA) |
| | Wang et al., 2015[134] | | Wang et al., 2015[134] (California, CA, USA) |
| Bullying Literature Project–Moral Disengagement | None | None | Wang et al., 2017[136] |
| DRACON | Malm and Löfgren, 2007[119] | O'Toole and Burton, 2005[120] (Brisbane, QLD, Australia) | None |
| English Classes (no name) | Holcomb and Denk, 1993[108] | Holcomb and Denk, 1993[108] (Houston, TX, USA) | None |
| Hashish and Marijuana | Zoller and Weiss, 1981[135] | Zoller and Weiss, 1981[135] (Haifa, Israel) | None |
| I-LST | Bechtel et al., 2006[84] | Bechtel et al., 2006[84] (Pennsylvania, PA, USA) | Smith et al., 2004[128] (Pennsylvania, PA, USA) |
| | | | Vicary et al., 2006[132] (Pennsylvania, PA, USA) |
| KAT | Segrott et al., 2015[127] | Rothwell and Segrott, 2011[126] (Wales, UK) | Segrott et al., 2015[127] (Wales, UK) |
| | | Segrott et al., 2015[127] (Wales, UK) | |
| Learning to Read in a Healing Classroom | None | None | Torrente et al., 2015[137] |
| | | | Aber et al., 2017[138] |
| LIFT | DeGarmo et al., 2009[98] | None | DeGarmo et al., 2009[98] (Oregon, OR, USA) |
| | Eddy et al., 2000[100] | | |
| | Eddy et al., 2015[99] | | Eddy et al., 2000[100] (Oregon, OR, USA) |
| | Reid et al., 1999[125] | | Eddy et al., 2003[101] (Oregon, OR, USA) |
| | Reid and Eddy, 2002[124] | | |
| | | | Reid et al.,1999[125] (Oregon, OR, USA) |
| | | | Stoolmiller et al., 2000[131] (Oregon, OR, USA) |

**TABLE 2** Overview of reports included in this review (*continued*)

| Interventions examined in the review | Included theory studies (author, year) | Included process evaluations (author, date, location) | Included outcome evaluations (author, year, location) |
|---|---|---|---|
| Peaceful Panels | Wales, 2013[133] | Wales, 2013[133] (Athens, GA, USA) | None |
| Positive Action | Beets *et al.*, 2008[86] | Beets, 2007[85] (Hawaii, HI, USA) | Bavarian *et al.*, 2013[83] (Chicago, IL, USA) |
| | Beets *et al.*, 2009[87] | Beets *et al.*, 2008[86] (Hawaii, HI, USA) | Beets *et al.*, 2009[87] (Hawaii, HI, USA) |
| | Flay, 2009[55] | | |
| | Flay and Allred, 2010[102] | Malloy *et al.*, 2015[118] (Chicago, IL, USA) | Lewis, 2012[113] (Chicago, IL, USA) |
| | Lewis, 2012[113] | | Lewis *et al.*, 2012[114] (Chicago, IL, USA) |
| | Malloy *et al.*, 2015[118] | | Lewis *et al.*, 2013[115] (Chicago, IL, USA) |
| | | | Li *et al.*, 2011[116] (Chicago, IL, USA) |
| | | | Snyder *et al.*, 2010[129] (HI, USA) |
| | | | Snyder *et al.*, 2013[130] (HI, USA) |
| PATHS | Crean and Johnson, 2013[97] | Ransford *et al.*, 2009[123] (Pennsylvania, PA, USA) | Crean and Johnson, 2013[97] (USA) |
| | Flay *et al.*, 2009[55] | | |
| | Greenberg and Kusché, 2006[106] | | |
| | Kusché and Greenberg, 2012[112] | | |
| RHC | Brown *et al.*, 2005[92] | None | Brown *et al.*, 2005[92] (Seattle, WA, USA) |
| | Catalano *et al.*, 2003[95] | | Catalano *et al.*, 2003[95] (Seattle, WA, USA) |
| Roots of Empathy | Cain and Carnellor, 2008[94] | Cain and Carnellor, 2008[94] (WA, Australia) | None |
| | Gordon, 2003[105] | | |
| | Hanson, 2012[107] | Hanson, 2012[107] (Isle of Man and Western Canada) | |
| Second Step | None | None | Espelage *et al.*, 2013[139] |
| | | | Espelage *et al.*, 2015[140] |
| | | | Espelage *et al.*, 2015[141] |
| Steps to Respect | Brown *et al.*, 2011[93] | Low *et al.*, 2014[117] (California, CA, USA) | Brown *et al.*, 2011[93] (CA, USA) |
| | Frey *et al.*, 2009[103] | | Frey *et al.*, 2005[104] (Pacific Northwest, USA) |
| | | | Frey *et al.*, 2009[103] (Pacific Northwest, USA) |

**TABLE 2** Overview of reports included in this review (*continued*)

| Interventions examined in the review | Included theory studies (author, year) | Included process evaluations (author, date, location) | Included outcome evaluations (author, year, location) |
|---|---|---|---|
| The Gatehouse Project | Bond and Butler, 2010[88] | Bond *et al.*, 2001[89] | Bond *et al.*, 2004[91] (Melbourne, VIC, Australia) |
| | Patton *et al.*, 2000[122] | | Bond *et al.*, 2004[90] (Melbourne, VIC, Australia) |
| | Patton *et al.*, 2003[121] | | Bond and Butler, 2010[88] (Melbourne, VIC, Australia) |
| | | | Patton *et al.*, 2006[60] (Melbourne, VIC, Australia) |
| Youth Matters | Jenson and Dieterich, 2007[110] | None | Jenson and Dieterich, 2007[110] (Denver, CO, USA) |
| | | | Jenson *et al.*, 2010[111] (Denver, CO, USA) |
| | | | Jenson *et al.*, 2013[109] (Denver, CO, USA) |

DRACON, DRAma = CONflict; I-LST, Infused Life Skills Training; KAT, Kids and Adults Together; LIFT, Linking the Interests of Families and Teachers; PATHS, Promoting Alternative Thinking Strategies; RHC, Raising Healthy Children.

The 4Rs programme aimed to integrate social–emotional learning and conflict resolution with lessons in literacy centred on children's books.[3,51–53] One of the driving ideas behind this intervention was to 'reintroduce' social–emotional learning to the curriculum in the face of crowding out by more explicitly academic lessons. The intervention is designed to be implemented school-wide in the US equivalent of years 1–6, but the evaluation cohort was enrolled in year 4. The intervention, which was extensively manualised, was implemented by classroom teachers and supported by intensive introductory training (25 hours in duration) and ongoing coaching and classroom observation. In each year of the intervention, the course materials were divided into seven units, with 21–35 lessons across the units and a benchmark of one lesson delivered per week.[3] In each unit, students were introduced to a children's book, and teachers then followed-up each book with between three and five social–emotional learning lessons. The intervention was underpinned by both 'multilevel program theory' and 'developmental cascades theory';[53] that is, 4Rs is designed to address multiple levels of social relationships and is expected first to affect 'proximal' outcomes, such as hostile attribution biases, to be followed by improvements in more 'distal' outcomes, such as reduced aggression.

The Bullying Literature Project[96,134] aimed to prevent bullying in the USA-equivalent of years 4 and 5 via integrating lessons on how to cope with bullying and how to intervene in peers' bullying behaviours with children's literature. The intervention was conducted over five sessions of between 35 and 45 minutes each in one school term and implemented by school psychologists (i.e. not classroom teachers). The Bullying Literature Project, which was underpinned by social learning theory, was based on combining instructional elements with role playing and modelling the positive behaviours demonstrated by characters in the chosen texts.[96,134] The interventionists read the lesson's story to children and then engaged them in activities, including role plays and writing. The intervention developers noted that by giving students the opportunity to practice new skills and observe how peers react to bullying scenarios, students could also learn to reduce their own bullying behaviours. In addition to classroom lessons, data collected on students' reported bullying experiences and behaviours were presented to both students and school staff, and parents were informed of the intervention and provided with a list of suggested readings and strategies to reinforce learning from the intervention.

A second iteration of this intervention, the Bullying Literature Project–Moral Disengagement,[136] included additional content relating to moral disengagement and sought to teach students how to intervene as bystanders rather than 'walking away' from others' bullying behaviours.

The goals of DRAma = CONflict (DRACON)[119,120] are to use drama to develop cognitive understanding of conflict and bullying, to empower students to manage their own conflict both personally and within the broader school community. DRACON is delivered to students in primary and secondary schools (ages 7–16 years). Conflict literacy is taught through 'enhanced forum theatre' and other drama techniques. The peer learning classes taught by the students are typically not delivered in drama classes but in other classes (e.g. English) because of the perceived opportunities in the curriculum to discuss conflict and bullying. Health and academic education are said to be integrated by conflict literacy being taught through 'enhanced forum theatre' and other drama techniques. It is unclear from the study report, but it appears that there are nine cycles, each of which runs for 8 weeks with two 100-minute classes per week within the school year. Cycle two onwards has the aim of combining drama and peer teaching by older classes with their younger peers. These classes also aim to develop group communication, familiarity, and empathy. The DRACON project's theory of change[119,120] is informed by Coleman's model of adolescent development,[142] Kolb's modes and characteristics of experiential learning[143] and Winslade and Monk's approaches to peer mediation.[144]

The English Classes[108] intervention targets students in the USA-equivalent of years 9 and 10 in the UK (aged 13–15 years). Teachers are trained and, working in pairs, develop integrated health/English material, with a specific emphasis on the prevention of drug and alcohol use. It is unclear from the study report, but it seems that the intervention is expected to be delivered in (potentially) every English class, with health lessons integrated whenever appropriate. Health topics are infused into English classes. English was chosen because it was felt to be the subject during which non-traditional concepts could be discussed and because it is taken by all students. No details on the timing and 'dose' of the intervention are provided. The theoretical basis of the English Classes intervention is vague but is described as being informed by learning theory[145] and the notion of the impact of integrated subject matter.[146]

The goal of the Hashish and Marijuana programme[135] that is delivered to students in Israeli upper high school (aged 17–18 years) is to develop scientific knowledge of hashish and marijuana and to strengthen students' problem-solving and decision-making skills. Lessons about hashish and marijuana are delivered through lectures in chemistry classes and complemented by group work, class discussions, independent study, projects, games, field trips and any other curricula additions that develop social and decision-making skills. The intervention is entirely integrated into chemistry classes, where lessons around hashish and marijuana take place, teaching the students about the chemical aspects of the drugs. Behaviour change is also addressed through more participatory teaching methods. No information on 'dose' and no theoretical basis are provided for the Hashish and Marijuana intervention.[135]

Infused Life Skills Training (I-LST)[128,132] was based on an existing intervention, Life Skills Training (LST), but was modified to be incorporated alongside other classes. Intervention delivery was undertaken by classroom teachers with students in the USA-equivalent of year 8 and this delivery lasted for 3 years. In I-LST, which targeted drug use outcomes, teachers from participating schools worked with trial staff to map all the standard LST principles onto standard academic curricula (i.e. not as a dedicated module) for the relevant year, including how academic objectives could be met alongside intervention objectives. The goal of this approach was to facilitate repeated exposure to drug prevention content and to reduce the need for time to be set aside for prevention education. Thus, instead of a fixed set of lessons, adaptation and integration were different in every school, with the goal of covering all relevant LST principles in at least one subject area. Key content from the 'standard' LST intervention includes drug education, decision-making, communication and assertiveness, and self-management.

Kids and Adults Together (KAT)[127] was an intervention centred on classroom activities, including a family event and take-home materials, to be implemented over approximately 20 hours of contact time and over at least 1 week of exposure. The intervention was trialled primarily on students in UK school years 5 and 6, with some students from year 4, and was delivered by classroom teachers with support from school staff. Integration with academic education was achieved by delivering KAT lessons and messages about health and alcohol consumption ('not too much, not too soon') in literacy, maths and science classes, although extensive details were not provided.[126,127] Initial pilot studies suggested that teachers would adapt material to local context and fit; thus, teachers were encouraged to adapt the specific materials provided for the intervention to their own contexts and needs. The culminating event of the intervention was a family event during which students exhibited the work that they had completed in different classes about alcohol use. Students were also provided with take-home materials [which included an informative digital versatile disc (DVD)] to reinforce information provided in school.

Learning to Read in a Healing Classroom[137,138] was trialled in the Democratic Republic of Congo. It was intended as a comprehensive intervention to include social–emotional learning through literacy lessons. This intervention included principles from 4Rs, both didactic classroom content to teach social–emotional learning through reading and comprehensive teacher development through training and ongoing 'teacher circles' alongside school-level strategies, developed and led by teachers, to improve the environment for learning. Outcomes examined in this intervention primarily related to mental health and violence.

Linking the Interests of Families and Teachers (LIFT)[98,100,101,125,131] was a universal preventive intervention targeting an entire school year and intervening in 'moment-to-moment social interaction processes' as a starting point to prevent longer-term child conduct problems, such as drug use or aggression, including through the promotion of social skills and by preventing association with antisocial peers. Integration with academic education was achieved via inclusion of study skills content alongside twice-weekly lectures of 1 hour's duration over 10 weeks. In these lectures, students are introduced to social and problem-solving skills, they practice these skills and then have free play followed by review and rewards. The intervention was delivered by study staff rather than classroom teachers. In addition to lessons, a reward system for positive playground behaviour, parent training (e.g. training on proactive parenting and positive reinforcement) in weekly meetings for 6 weeks and parent–teacher communication (e.g. via a recorded answerphone message from teachers for parents to receive updates on the intervention).

The Peaceful Panels[133] intervention aims to reduce and prevent bullying and improve peer relations at school. The Peaceful Panels intervention targets students in the USA-equivalent of years 9 and 10 in schools in England and Wales (aged 13–15 years). Throughout what appear to be all art classes within a school year, students participate in antibullying lessons (from the Second Step[139–141] intervention for eighth-grade students on empathy and communication in handling a grievance) and comic-making lessons. Students then prepare artwork to demonstrate their understanding of how to resolve conflict. Conflict resolution lessons are integrated fully within art classes, and creation of art is used as a medium to reflect on learning. The Peaceful Panels[133] programme is described as being rooted in Stuhr's 1994 paradigm of Social Reconstructionist Art Education[147] and critical pedagogy.[148]

The Positive Action intervention was evaluated in two included trials: Chicago[83,113–116] and Hawaii.[87,130] The Positive Action intervention is designed as a whole-school intervention to affect multiple domains, including academic attainment, social and character development and student behaviour. Although the intervention is designed to occur from the USA-equivalent of year 1 through to year 13, the cohorts evaluated in Chicago received the intervention from year 4 to year 9[83] and the cohorts evaluated in Hawaii received the intervention from year 2 or 3 through to year 9.[87] The classroom component of Positive Action, intended to be delivered by classroom teachers, includes 140 lessons per academic year in the initial years of the intervention, with each lesson being about 20 minutes. Across units, academic achievement and study skills are explicitly linked to positive health behaviours, and students are encouraged to link academic learning to real-life situations. Teachers are also supported to link Positive Action lessons to curricular content.

The specific units of Positive Action include supporting students to improve their health (e.g. via physical activity, self-regulation and study skills for self-management), social skills (e.g. empathy), honesty with self and others, and tools for self-improvement (e.g. goal-setting). Throughout the intervention, teachers and students are encouraged to communicate with each other via discussions and activities. The classroom component is accompanied with a school-wide climate intervention led by the principal/head teacher or a specific intervention co-ordinator.

Promoting Alternative Thinking Strategies (PATHS)[97] was a social–emotional learning intervention delivered over 3 years to students starting in the USA-equivalent of year 4. Lessons in PATHS are divided into six units over 3 years, with lessons two or three times weekly over the duration of the school year provided by classroom teachers with support from programme staff. Each lesson is intended to be between 20 and 30 minutes. Although teachers are encouraged and supported to integrate PATHS activities into traditional instruction (and this is included in the instructional manual), this appears to be largely determined by individual teachers. As part of this programme, teachers use a variety of visual materials (e.g. posters, pictures) to teach students 'interpersonal problem-solving skills', 'self-control', 'emotional understanding', 'positive self-esteem' and 'healthy relationships'. Students were assigned homework relating to PATHS as a vehicle for parental involvement. A school-wide component focusing on planned events reinforced the classroom-level intervention and teachers received training in the intervention, including weekly consultations with programme staff.

Raising Healthy Children (RHC)[92,95] was a comprehensive intervention designed to target risk and protective factors for later problem behaviours. RHC was evaluated in a cohort of students starting in the equivalent of English and Welsh schools' year 2 or 3. Oriented from a 'social development' perspective, the intervention linked academic, cognitive and social skills with substance use avoidance, with the theory that improved bonding with 'pro-social others' (e.g. non-substance-using peers and family) would lead to reduced substance use in adolescence. The programme integrated academic and health education in two ways. First, participating classroom teachers received continuing professional development (from year 2 to year 8), focusing on academic skills for students (e.g. reading), social and problem-solving skills, and classroom management. Second, student-level interventions integrated academic and health education in years 5 to 7 by providing academic enrichment and other extracurricular opportunities for students after school, including substance use refusal skills alongside tutoring, study clubs and workshops, and summer camps in the early years of the intervention. These integrating components were offered alongside parenting workshops between years 2 and 9 and, when necessary, specific families had access to a 'school–home co-ordinator' who was also an expert in the intervention. Thus, the intervention was delivered by both classroom teachers and additional programme staff. Because of the intensive longitudinal nature of the intervention exposure varied, but students, on average, had 28.3 contacts with the student and family intervention strategies.

Roots of Empathy[94,105,107] is an intervention that brings a visiting baby and mother (and/or father) into a classroom. Through the students' exploration of the baby's interactions with his/her parent(s) a platform for learning empathy is provided. It is implemented for students in the equivalent of English and Welsh schools' years 2 to 10 (ages 7–15 years). Visits are facilitated by a certified Roots of Empathy instructor who works with the class before, during and after each baby visit. Students build empathy by observing the baby's development and milestones and by interacting with the baby, learning about its needs and temperament; through this, students learn messages of social inclusion, respect, how to build consensus and how to contribute to a safe and caring classroom while developing emotional intelligence. Throughout the school year, a neighbourhood baby will visit the class with its parent(s) nine times. Connections with literacy, writing, art, music, mathematics and science are made. Lessons may involve literature to help develop a theme to encourage children to explore their own emotions and experiences. Children may also make artwork to reflect these feelings, which is often compiled and used in follow-up activities. The Roots of Empathy intervention is described by its authors as being informed by the Collaboration for Academic, Social, and Emotional Learning's framework for the design of the curriculum, but no further detail is provided on the theory of change.[149]

The Second Step[139–141] intervention has been evaluated in multiple trials but the reports included in this review focused on an intervention specifically describing integration with academic lessons through homework. This intervention includes 15 weeks of classroom lessons taught weekly or every 2 weeks throughout the school year for 3 years. Teachers are supported by professional development training to deliver intervention content (alongside videos), which includes topics on bullying, problem-solving, emotional regulation and empathy. Teachers also receive plans on how to support the integration of Second Step content into academic lessons. Modelling, role play and coaching are included in the intervention. Students receive homework to reinforce skills and use group and collaborative work to practice these skills.

Steps to Respect intervention was evaluated in two studies: one testing a 2-year version of the intervention when outcomes were in part ascertained via playground observation[104] and the other testing a 1-year version of the intervention when outcomes were primarily ascertained via a questionnaire.[93] Both evaluations included children starting in the equivalent of English and Welsh schools' year 4, which is when the intervention was intended to be delivered, and were similar in intervention delivery and content (i.e. integration of academic and health education, staff training and individual coaching components). However, the second evaluation examined outcomes only after 1 year. Academic and health education were integrated in this intervention via literature-based lessons informed by cognitive–behavioural techniques, with three consecutive yearly modules. The goal of these lessons was to promote positive norms around eliminating bullying and to encourage the development of social–emotional skills. Over a period of 12–14 weeks, students received bi-weekly lessons in social–emotional skills totalling at about 1 hour per week for about 10 weeks, followed by a unit comprised 8–10 lessons based on appropriate literature to further explore antibullying themes and pro-social skills. In addition, staff were trained in the intervention and school-wide antibullying policies were identified. When students were identified as engaging in or victimised by bullying behaviours, they received 'brief individual coaching' to focus on skills for the prevention of future bullying episodes. The intervention was provided by classroom teachers and school staff, with staff training by external programme staff.

The Gatehouse Project[60,88,90,91] was a multiyear, whole-school intervention that included classroom-based and professional development with the aim of 'whole-school change' to support the social–emotional and behavioural development of adolescents. The programme began when students were in the Australian and USA equivalents of year 9. Academic and health education were integrated via lessons in English classes and also through PSHE classes inspired by stories and poems, among other materials. As part of these lessons, students practised strategies for addressing situations that provoke difficult emotional responses. The framework for the lessons was informed by principles of cognitive–behavioural therapy, primarily relating to self-management and the link between thoughts and feelings. Lessons were intended to be provided over a 10-week period in each of 2 intervention years by classroom teachers, who were encouraged to adapt the programme to their own lessons through an understanding of the principles underlying the intervention. More broadly, intervention developers described integration as occurring by linking students' health and well-being to educational outcomes.[88] Although drug use behaviours were a key outcome of this study, drug use in particular was not a focus of lessons delivered in classrooms. Whole-school components of the intervention included specific targeting of intervention strategies based on combinations of risk and protective factors in each school and the creation of school-wide adolescent health teams. In addition, a school liaison team provided specific support to schools included in the study and supported teachers in implementing the intervention via weekly meetings. The overarching foci of the intervention, as stated by the developers, were on 'building a sense of security and trust; enhancing communication and social connectedness; and building a sense of positive regard through valued participation in aspects of school life'.[88]

Finally, Youth Matters[109–111] was an intervention intended to reduce bullying behaviours and victimisation. In this intervention, intended to be implemented in the USA, students of the UK-equivalent of years 5 and 6 read a story between 30 and 40 pages in length and then engaged in either 'issues' modules or 'skills' modules. Issues modules considered peer norms relating to bullying and skills modules involved the development of skills to both prevent and respond to bullying. Modules culminated in group projects shared with the school community. Academic and health education were integrated via the story that

began each module, as these texts were designed to address both types of education. Each of the four modules was 10 weeks long, with a weekly lesson of about 90 minutes in duration. In the Youth Matters trial, the intervention was delivered over four semesters with one module per semester, and the interventionists were education specialists on the trial staff rather than classroom teachers.[109] Similarly to RHC,[92,95] the Youth Matters intervention was based on a 'social developmental' model that emphasised bonding with pro-social peers, establishment of pro-social norms and values, and development of skills to prevent and respond to negative behaviours as key to positive development. Because the study population included a high number of Hispanic American students, materials were adapted (although specific methods are not discussed) and translated for use in Spanish-speaking classrooms.[110]

## Study characteristics

A descriptive overview of the 68 study reports[3,51–57,60,83–135] representing 30 empirical studies of 19 interventions included in the review is provided below. As indicated in *Table 1*, many of the studies are linked and have been counted as one empirical study; for example, a study presenting results of an intervention at 1 year, another presenting further follow-up of the study after 2 years of implementation and another at 3 years (see Jones *et al.*,[52] Jones *et al.*[51] and Jones *et al.*[3] for examples of this). In some instances, one report may focus on studies of more than one intervention, as in Flay *et al.*,[55] or it may include reports of two different empirical studies in different sites of the same intervention, as in Hanson.[107] Altogether, data are provided from 30 separate empirical studies included within these 68 study reports.[3,51–57,60,83–135] The summaries below include details of the rate of report publication, geographical location of each empirical study, which age groups were targeted by the interventions, providers of the interventions, intervention components and the characteristics of the integration of health and academic education involved in each intervention.

### Rate of report publication

*Figure 2* provides a breakdown of rate of publication, based on the 68 reports included in the review and according to which review question they answered. Only six studies were published between 1981 and 2000,[100,108,122,125,131,135] and 60 from 2001 to 2015,[3,51–57,60,83–99,101–107,109–121,123,124,126–130,132–134,137,139–141] reaching peaks in 2009, 2010, 2013 and 2015, each with six publications per year. An additional two papers were
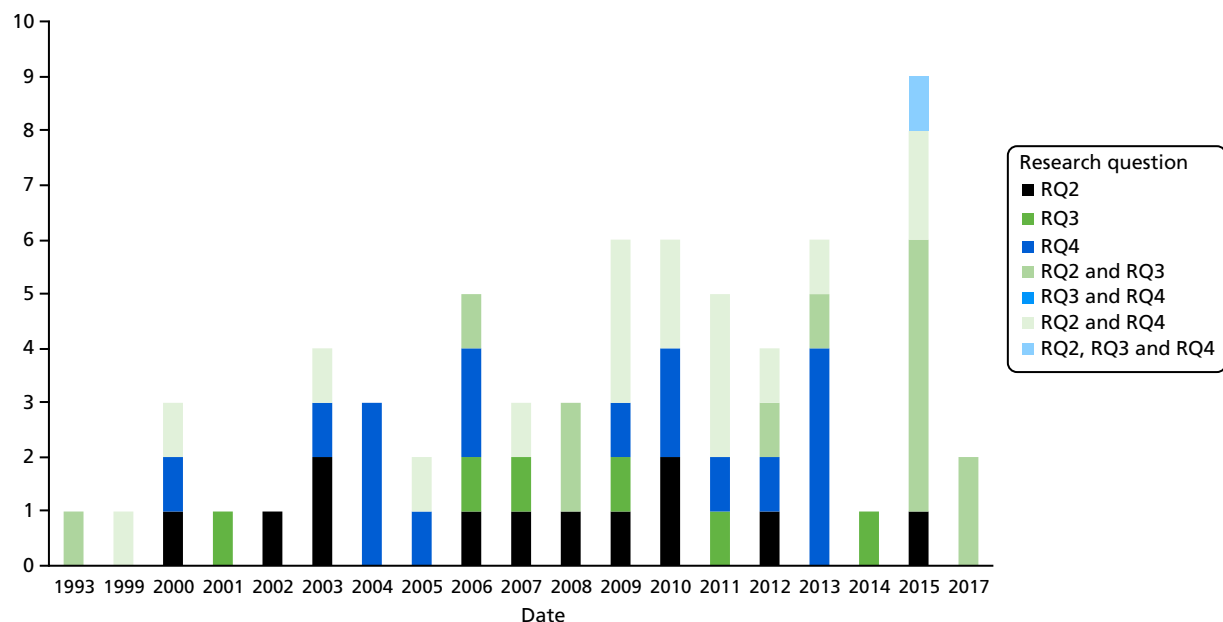


**FIGURE 2** Rate of study publication by research question, by year. RQ, research question.

published in 2017.[136,138] Of the 68 studies, 12 (19%) published only on research question 2 (theory),[54–56,99,102,105,106,112,119,121,122,124] five (8%) published only on research question 3 (process),[89,117,120,123,126] 23 (27%) published only on research question 4 (outcomes),[51,60,87,90,91,101,104,109,111,114–116,128–132] another 17 (25%) published on research questions 2 and 4 (theory and outcome),[3,52,53,83,88,92,93,95–98,100,103,110,113,125,134] 10 (16%) published on research questions 2 and 3 (theory and process),[57,84–86,94,107,108,118,133,135] and none was published on process and outcomes together. Only one publication[127] answered research questions on theory, process and outcomes.

### Geographical location of included studies
Of the 30 empirical studies of 19 interventions indicated in *Table 1*, 23 (77%) reported on a study conducted in the USA.[3,51–54,56,57,83–87,92,93,95–101,103,104,108–111,113–118,123–125,128–134] There were four studies conducted in Australia (13%),[60,88–91,94,119–122] two studies conducted in the UK (7%)[107,126,127] and one study each from Canada (3%)[107] and Israel (3%).[135] An additional study was carried out in the Democratic Republic of Congo (3%) (*Figure 3*).[137,138]

### Phase of schooling targeted by interventions
There was a fairly even split between interventions aimed at primary school students (*n* = 9)[3,51–57,93,96–101,103,104,106,109–112,117,123–127,131,134,136–138] and secondary school students (*n* = 6),[60,84,88–91,108,121,122,128,132,133,135,139–141] with four interventions targeting both primary and secondary school students (*Figure 4*).[55,83,85–87,92,94,95,102,105,107,113–116,118–120,129,130] Note that primary school is defined as encompassing years 1–6 and secondary school as years 7–12. These year groups were given different labels by authors in different settings (e.g. elementary school and high school for studies in US settings).

### Providers of interventions
Not surprisingly, teachers of the academic curriculum were also the primary facilitators for 13 interventions.[3,51–57,83,85–87,93,96–104,106,108,112–120,123–125,129–131,133–141] Five interventions engaged teachers and external facilitators together, often with the external facilitator acting in more of a mentoring role with the intention that the teacher might take over fully at a later date.[60,84,88–92,94,95,105,107,121,122,126–128,132] Only one intervention, Youth Matters, was fully externally facilitated.[109–111]
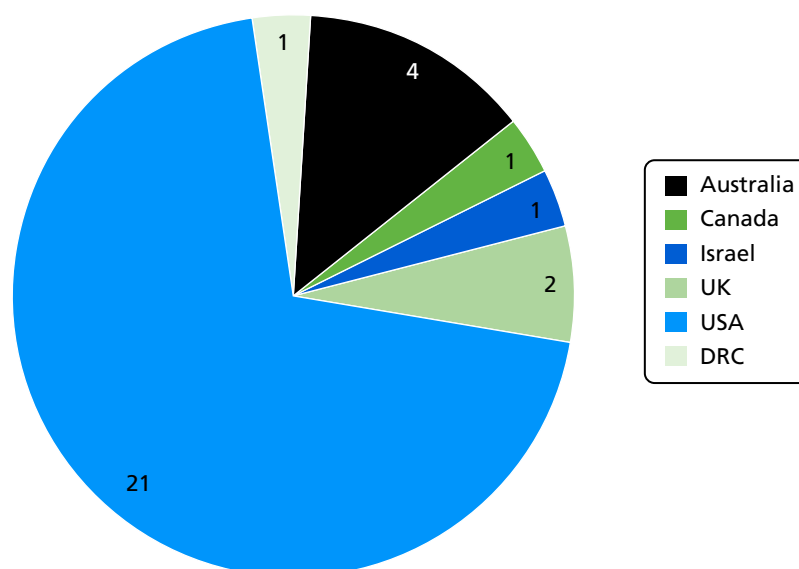


**FIGURE 3** Spread of studies by country. DRC, Democratic Republic of Congo.
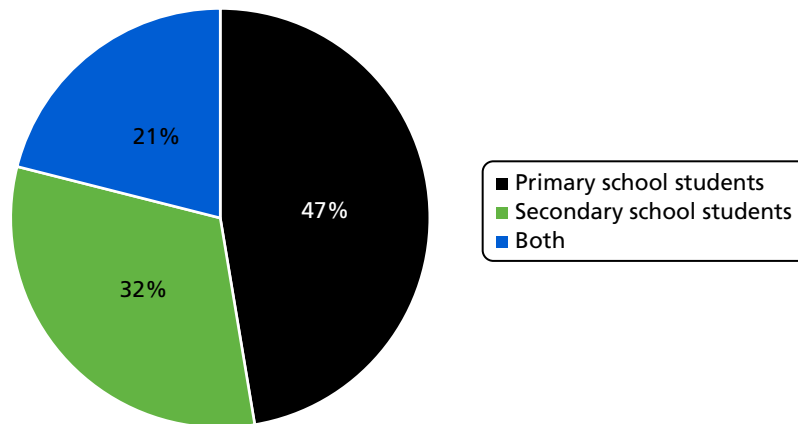
**FIGURE 4** Intervention target populations.

## Intervention components

Twelve interventions had only a classroom component (*Figure 5*).[3,51–57,84,94,96,97,102,105–112,119,120,123,128,132–136,139–141] One intervention had classroom and family components,[126,127] four had classroom and school-wide components[55,60,83,85–91,93,102–104,113–118,121,122,129,130,137,138] and two had classroom, school and family components.[92,95,98–101,124,125,131]

Classroom-only strategies sought principally to deliver content in the classroom context (e.g. 4Rs,[3,51–57] the Bullying Literature Project,[96,134] I-LST[84,128,132] and Youth Matters[109–111] were all principally focused on classroom-based activities). Although participants in Youth Matters[109–111] presented their work in school-wide fora, it was not clear from the intervention's theory of change whether or not this was central to the intervention's outcomes (i.e. the goal was not to effect school-wide change).

Classroom and whole-school change strategies often linked academic and health education to school-level initiatives with the goal of reflecting messages about social–emotional learning and pro-social behaviour throughout the school, by altering the school environment and by providing school-level resources to support classroom-level learning. The Gatehouse Project[88–91,122] and Positive Action[55,83,85–87,102,113–116,118,129,130] both focused on school climate beyond the classroom. The Gatehouse Project[88–91,122] included the development of an adolescent health team in each school. This adolescent health team, which was supported by intervention developers and evaluators, was tasked with the dual foci of enhancing adolescent attachment to school and creating a locally relevant intervention strategy based on school-level profiles of risk and protective factors. Similarly, Positive Action[55,83,85–87,102,113–116,118,129,130] included the formation of a 'climate intervention' led by the head teacher or a designated co-ordinator to facilitate school-level change in pro-social norms and
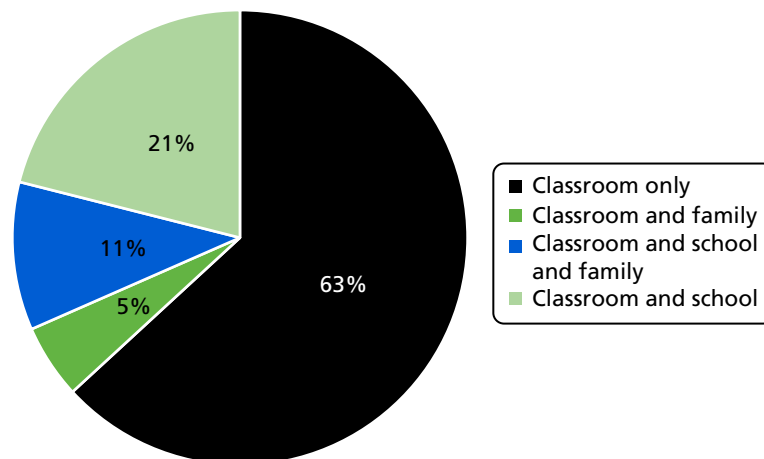


**FIGURE 5** Intervention components.

behaviour. Steps to Respect blended school climate with specific out-of-classroom resources. School-wide bullying policies were audited and developed via a targeted staff committee, and participants in bullying episodes (both aggressors and victimised students) received targeted coaching from a counsellor after an episode was observed or reported. PATHS[55,97,106,112,123] included school-level events and materials to support classroom-level learning.

Classroom and external domain strategies sought to influence young people's development both in and beyond the classroom. A frequent focus beyond the classroom was the family. KAT[126,127] included a parents' evening at the conclusion of the intervention, during which students could display the work they had completed on safe alcohol consumption. The goal of the parents' evening, as well as the goal of the materials (e.g. DVDs, leaflets) provided to parents, was to involve parents in conversations about how to time alcohol consumption. PATHS[55,97,106,112,123] similarly included 'take-home' materials in the form of homework on social–emotional learning to be completed with parents and newsletters on programme topics to facilitate parental involvement. LIFT[98–101,125] included several components to facilitate family involvement. Parents were able to call a classroom telephone line when the classroom teacher recorded details of the current intervention activities. They were also invited to attend a series of parents' evenings to discuss parenting skills and strategies. Although RHC[92,95] also included parent workshops, it was a 'wrap-around' intervention strategy that included extensive out-of-school enrichment and activities, including summer camps, to target out-of-school domains.

### Primary academic subject within which the intervention is delivered

English was the most common academic subject into which intervention curricula were integrated ($n = 7$) (*Figure 6*).[3,51–57,93,96,103,104,108–111,117,134,136–138] One intervention was integrated into drama classes,[93,119,120] one into an art class,[133] one into a chemistry class[135] and four into more than one subject.[55,84,94,97,105–107,112,123, 126–128,132] Five interventions did not specify which academic subject they should be integrated into.[55,60,83,85–92, 95,98–102,113–116,118,121,122,124,125,129–131,139–141] None of the included interventions was integrated into specific health classes although in some interventions (e.g. KAT and the Gatehouse Project) authors indicated that a health class would be suitable for the intervention to be integrated into.[60,88–91,121,122,126,127]

### Extent of integration of health and academic education

A key axis of variation in the interventions that included studies evaluated was the extent of integration; that is, the degree to which lessons were fully integrated (with activities addressing health and academic learning objectives simultaneously and seamlessly) or partially integrated (with interventions comprising separate learning activities for health and academic learning objectives within one overall package).
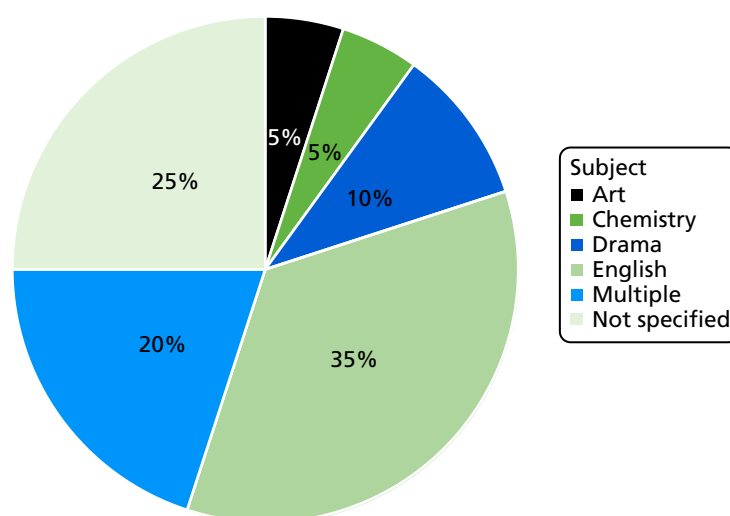


**FIGURE 6** Primary academic subject into which health education is integrated.

In 13 out of the 19 interventions (68%),[3,51–57,60,84,88–91,94,96,105,107–111,119–122,126–128,132–138] integration was full in the sense that interventions included activities aiming to address health and academic learning objectives seamlessly. Interventions involving full integration did not attempt to separate health-related and academic-related learning activities. The clearest example of this in the included interventions is I-LST,[84,128,133] in which the rationale for integration was to step away completely from delivering the LST content in a separate class. Instead, teachers presented and reflected LST content throughout academic subjects. In addition, the Gatehouse Project[88–91,122] targeted English classes for the delivery of content on social–emotional learning based on cognitive behavioural principles. Local integration was not the only way in which full integration was achieved. The children's stories delivered as part of 4Rs[3,51–57] and Youth Matters[109–111] were intended to support literacy, language and arts teaching in schools, and subsequent lessons on social–emotional learning reflected and extended literacy-based learning. In addition, student projects created as part of the Youth Matters interventions were shared with the school community.

In 4 out of the 19 interventions (32%),[55,83,85–87,92,93,95,97–104,106,112–118,123–125,129–131,139–141] integration was more partial, typically these interventions included activities separately addressing health and academic objectives but within one overall package. Interventions involving partial integration set aside time specifically for health and academic learning. For example, Positive Action[55,83,85–87,102,113–116,118,129,130] included a set of manualised lessons to be delivered on a weekly basis across the multiple years of the intervention. RHC,[92,95] LIFT[98–101,125] and PATHS[55,97,106,112,123] included discrete content on study skills. PATHS also provided materials for teachers to integrate content into their lessons and in RHC teachers were provided with professional development on linking academic success with behavioural development.

### Approach to integration
There were three overall approaches to integration, via the use of literature, local development of integration or linking to other developmental issues and concerns. In interventions using a literature-based pathway to integration, children's books were used as a starting point for the discussion of social–emotional learning skills. Of note is that all six interventions (seven evaluations) that primarily used a literature pathway to integration sought to address violence outcomes, including, in particular, bullying. For example, 4Rs[3,51–57] used weekly lessons divided into seven units in which each unit began with a children's book. Social–emotional learning was supported with lessons following on from the unit's book. The Youth Matters[109–111] intervention followed a similar format, with a story used to lead into modules relating to issues or skills (i.e. peer norms around bullying or how to respond to bullying). The Bullying Literature Project[96,134] used children's books over five sessions as a starting point to engage in a variety of role-playing, writing and behavioural modelling activities to practice the lessons offered by each book, each session included both literature and follow-up activities. The Steps to Respect[93,103,104,117] intervention sequenced literature integration differently; students began with social–emotional skills lessons and then participated in a classroom 'unit' based on children's literature to reinforce lessons on pro-social skills. Even when we judged that literature was not the primary strategy used by an intervention, this approach was present. For example, the Gatehouse Project targeted English classes to integrate cognitive–behavioural principles in the lessons delivered.

In interventions using a local development pathway to integration, teachers developed their own ways of connecting health and academic education.[88–91,122] In all cases, teachers were supported to do this by pre-existing materials and curricula on health and social–emotional well-being subjects. Three interventions (three evaluations) primarily used this approach to integrate academic and health education. I-LST[84,128,132] departed from the original LST programme in that lessons were not provided as part of discrete PSHE classes. Instead, teachers of academic subjects in participating schools developed a 'matrix' of the lessons and topics from the LST manual against academic subjects and developed a local plan for introducing relevant LST material across different classes. The underlying idea was to reinforce material across different academic subjects and potentially increase the intervention 'dose' that students would receive compared with a standard health education class. Although extensive detail regarding KAT[126,127] was not provided in the main trial report, classroom teachers delivered lessons relating to alcohol consumption and students participated in pairs and groups to produce work for a parents' evening. Thus, integration of academic and health education was specific to each classroom. Finally, PATHS[55,97,106,112,123] provides materials to teachers

(e.g. posters and homework) to support the integration of PATHS activities on social–emotional skills into daily instruction on academic subjects. This strategy was also present in other interventions, although we judged that it was not the primary strategy used. For example, Positive Action[55,83,85–87,102,113–116,118,129,130] encouraged teachers to locally integrate Positive Action content into academic classes, even though Positive Action was also delivered via discrete lessons.

Interventions that used a linking approach to the integration of academic and health education emphasised the inter-relationships between academic success and 'successful' child and adolescent development, specifically in terms of physical health and social–emotional well-being. This strategy was the primary approach used in four interventions (five evaluations). Positive Action,[55,83,85–87,102,113–116,118,129,130] throughout its extensive manualised curricula, links academic success with health and well-being. Thus, lessons taught to students on, for example, self-regulation and social skills, are linked to academic lessons, social–emotional development and positive health behaviours. Skills taught to students are also intended to influence both pro-social behaviours and academic success. The Gatehouse Project[88–91,122] took a similar approach, emphasising the basis of the intervention as a global understanding of child and adolescent development; thus, although drug use and bullying victimisation outcomes were measured, these were more distal impacts of the intervention, as opposed to the intervention's primary concern with social–emotional and behavioural development of adolescents. In LIFT,[98–101,125] the linking of academic and developmental concerns was achieved via delivering study skills content alongside social–emotional learning. RHC[92,95] focused on both study skills content and on co-delivering academic enrichment with behavioural management and substance use refusal, first by classroom intervention and subsequently by extracurricular opportunities. This strategy was also present in other interventions, although we judged that it was not the primary strategy used. For example, PATHS[55,97,106,112,123] linked to developmental concerns by presenting study skills content alongside social–emotional learning content in the last year of the intervention.

## Duration of interventions

Interventions integrating academic and health education could be placed at various time points in young people's social development as a single-year intervention or as a multiyear intervention. Single-year interventions located integration between health and academic education over the course of 1 year. For example, the Bullying Literature Project[96,134] was delivered over five sessions in one term. KAT[126,127] was similarly delivered over the course of an academic term. In contrast, LIFT[98–101,125] was delivered over the course of an academic year, but intervention activities were limited to that year.

Multiyear interventions located integration between academic and health education over the course of several academic years, generally by means of curricula targeted at students in a specific developmental phase. RHC,[92,95] Positive Action,[55,83,86,87,102,113–116,118,129,130] Steps to Respect,[92,103,104,117] Youth Matters,[109–111] PATHS,[55,97,106,112,123] Second Step[139–141] and I-LST[84,128,132] all included multiyear curricula. It should be noted that the second evaluation of Steps to Respect[93] tested only 1 year of implementation of Steps to Respect, even though the intervention itself was intended to last for 2 years. 4Rs[3,51–57] was also evaluated over a narrower range of intervention years (starting in year 4) than that in which the intervention was intended to be delivered. The evaluation of Learning to Read in a Healing Classroom[137,138] similarly evaluated only 1 year of a multiyear intervention.

# Chapter 4  Synthesis of theories of change

## About this chapter

Parts of this chapter are reproduced or adapted from Tancred et al.[150] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/. The text below includes minor additions and formatting changes to the original text.

This chapter describes and reports on the quality of the included study reports that describe theories of change for interventions included in this review. It then reports our thematic synthesis of these studies that aimed to produce an overall theory of change for this type of intervention to address research question 2. This chapter then uses this synthesis to develop some hypotheses about the factors that might influence the implementation, receipt or effectiveness of these interventions.

## Included studies

Appendix 16 summarises the 39 reports[3,52–57,83–88,92–100,102,103,105–107,110,112,113,118,119,121,122,124,125,127,133–135] that were included in this element of the review. Flay et al.[55] describes the theory of change of three interventions and is, therefore, repeated. Together, these reports provided the theoretical description of all 16 interventions included in the overall systematic review. A brief summary of the theories of change for each of the 16 interventions is included in Appendix 17.

## Quality of studies

Appendix 9 summarises the results of the quality appraisal of each study report. The combined appraisal for each study ranged from quite poor (20% of quality indicators judged to be present by reviewers) to quite good (70% of quality indicators judged to be present by reviewers). Only 16 of the 39 reports scored > 50%.[3,52–56,92,93,98,99,102,103,106,119,122,125]

## Synthesis of theories

Appendix 10 presents the overall coding framework that arose from our analysis, which informs the structure and content of this results section. The interventions within this review aimed to integrate and thus erode existing boundaries between health and academic education. The central theme in our analysis was that the theory of change of these interventions centred on the goal of eroding multiple boundaries within schools including, but not limited to, this erosion of boundaries between health and academic education. Many of these interventions' curriculum components aimed to train teachers to model positive behaviours and reinforce such behaviours among students, which we interpreted as aiming to erode boundaries between teachers and students. Classroom education was often supported by other components delivered at multiple levels (e.g. school and family), which aimed to ensure that learning and the reinforcement of positive behaviours occurred beyond the classroom. We interpreted these elements as aiming to erode boundaries between what occurs in classrooms and other settings, such as the wider school and students' families. Through these multiple but theoretically related mechanisms of erosion, interventions are thus intended

to promote experiential learning, the practising and reinforcement of positive behaviours, and enhanced relationships between teachers and students and between students and their pro-social peers. In turn, it is theorised that these will provide students with various attributes necessary to reduce engagement in substance use and violence, as well as to increase academic attainment. These attributes include the knowledge and skills necessary to avoid substance use and prevent conflict, the attitudes and social norms supportive of such behaviours, and students' sense of security within and positive connectedness to their classrooms and schools (*Figure 7*). The logic model depicts the intervention mechanism as a chain of proximal to distal impacts on participants and, hence, does not feature integration, which is a characteristic of the intervention itself.

The central organising theme of boundary erosion arose inductively from the data, but our ability to articulate this was informed by our knowledge of existing theory. The notion of boundary erosion recurred across our thematic analysis examining the mechanisms of each individual intervention's components. We interpreted it as resonating with an existing theory of how schools promote health: the theory of human functioning and school organisation.[77] This theory proposes the notion of boundary erosion as a multifaceted process by which schools can increase the commitment of students, particularly those of low SES, to school in terms of both learning and the school pastoral community. This is theorised to ensure the development of students' reasoning ability and a sense of positive affiliation with supportive peers, enabling them to make and enact more health-promoting decisions and avoid engagement with antischool peer groups and risk behaviours encouraged by such peers.

*Appendix 11*, *Table 14* was used and modified by the two reviewers to build the line of argument of the overall synthesis of the theory of change for the interventions included in our systematic review, expressed in the final column.

### Intervention inputs

Our review focused on school-based interventions and these overwhelmingly involved health curricula delivered in classrooms. Some interventions also included after-school or out-of-school components, whereas others had whole-school or family components. As an example of a multicomponent intervention, the Positive Action curriculum:

> . . . teaches specific positive actions for the whole self: the physical, intellectual, social, and emotional areas. Together, these make up the comprehensive set of skills for successful learning and living [. . .] The program trains teachers and parents to identify, teach, and reinforce positive thoughts, actions, and feelings about themselves by students and others in the school, leading to continual reinforcement of positive actions and enhanced student bonding with parents and school.
>
> *Flay and Allred[102]*

Some interventions required the training of teachers or other staff to deliver the curriculum. We have included it as an input (the dark green boxes in *Figure 7*) in our overall logic model for these interventions. Some curricula were externally facilitated and, therefore, no internal training of school staff was required.

Curricula addressed a wide variety of topic areas and desired skills. Of note, although our review outcomes were reductions of violence and substance use, only one intervention dealt directly with bullying (the Bullying Literature Project[96,134]). Overwhelmingly, the curricula concentrated on conflict resolution, problem-solving skills and social–emotional skills, as an example from the 4Rs intervention indicates:

> The approach of embedding social-emotional learning and conflict resolution lessons in a balanced literacy delivery strategy, and research tying together the social-emotional and academic domains, support our expectation for longer term effects on behavior and academic achievement.
>
> *Jones* et al.[3]
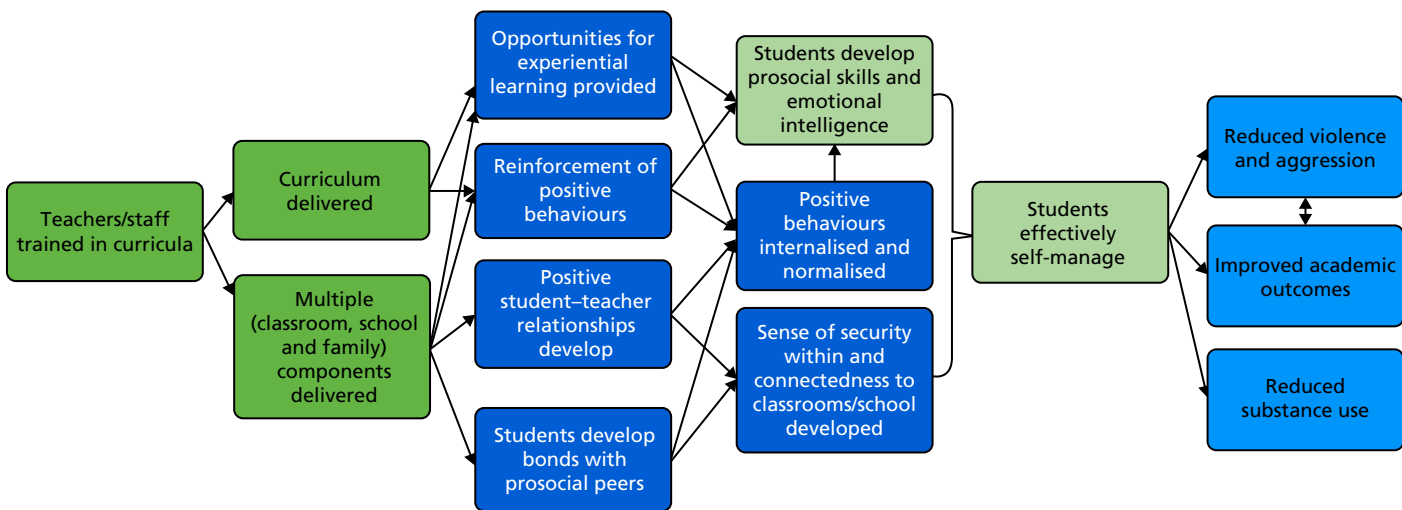
**FIGURE 7** Overall proposed logic model for integrated health and academic curricula for reduced violence or substance use. Dark green, intervention inputs; dark blue, mechanism of change expected to facilitate outcomes; light green, proximal outcomes; light blue, distal outcomes impacts.

### Mechanisms of change: improving health by eroding boundaries

As reported above, the central theme that emerged from our analysis was that eroding boundaries at multiple and mutually reinforcing levels (e.g. by integrating academic and health education, strengthening relationships between teachers and students, generalising learning from classrooms to the wider school environment and ensuring consistent messages from schools and families) will lead to the development of a community of engaged students oriented towards pro-social behaviour and away from substance use, violence and other risk behaviours.

### Eroding boundaries between health and academic education

A recurring theme was that interventions integrated and, thereby, eroded the boundaries between health and academic education. There were two types of integration, which we have defined here as 'full' integration and 'partial' integration. Authors described these approaches as well as the two key reasons for integration, detailed below.

In the full integration approach, health education was weaved seamlessly with academic subject classes, so that the same learning activities aimed to build both academic and health knowledge and skills seamlessly and simultaneously. An example of this integration is the DRACON intervention,[120] which used drama classes as a medium for building drama skills while also aiming to enable children to learn conflict resolution skills intended as a means of reducing aggression. Another example of a full integrated approach was 4Rs, which:

> . . . uses high-quality children's literature as a springboard for helping students gain skills and understanding in several areas including handling anger, listening, cooperation, assertiveness, and negotiation [. . .] ensuring students understand the primary themes of the story and allowing them to connect the themes to their own lives.
>
> *Aber* et al.[53]

In the second approach (i.e. partial integration), interventions would include separate learning activities for addressing substance use or violence and for addressing academic education, but the learning activities were not woven together. An example of this type of integration is PATHS which, in its grade 5 curriculum, included literacy components.[55,97,106,112] However, the PATHS curriculum is primarily focused on building non-academic skills, and although the curriculum is delivered during normal school hours within academic classes, these periods are distinct from the normal academic curriculum:

> Language arts (both reading and writing) are bridged to PATHS in most lessons by including supplementary suggestions for teachers to utilize such things as quality children's literature to reinforce lesson concepts. Further, one of the chapters in the Instructional Manual provides many ideas for how teachers can directly tie PATHS concepts to language arts, social studies/history, and other subject areas.
>
> *Greenberg and Kusché*[106]

Two common themes emerged in the synthesis as to the reasons why integration of health and academic curricula was attempted. The more pragmatic rationale referred to the decreasing time in the school day dedicated to teaching topics such as health, wellness or conflict resolution. This theme was particularly strong in the reports of some US interventions delivered after the 2001 No Child Left Behind Act.[151] The No Child Left Behind Act[151] emphasised academic performance on standardised tests, which, in practice, was unfortunately at the expense of other aspects of student life. Authors discussed how such interventions:

> . . . evolved in response to the tension between the movement to reform education between standards-based accountability with its focus on academic achievement, on the one hand (e.g. the policy and practice zeitgeist promoted by the No Child Left Behind Act of 2001), and social and character development, on the other.
>
> *Brown* et al.[54]

A second, more scientific, theme suggested that improving academic performance and health outcomes could be achieved simultaneously and that these outcomes were mutually reinforcing, which we have indicated in the outcomes component (see *Figure 7*) in our overall logic model. The extract below from the Roots of Empathy intervention[94] offers an example of this kind of discussion:

> *Schools can be reassured by scientific and pragmatic evidence that time spent implementing [social–emotional learning] will enhance the academic success of students [. . .] Current research shows that [social–emotional learning] programs contribute to the academic success of students as well as to their health and wellbeing.*
>
> *Cain and Carnellor[94]*

In discussions of the rationale for integration, a divergent opinion on the state of health education in schools was expressed by the evaluators of the Gatehouse Project, who suggested that there was, in fact, a new convergence of academic and health mandates in schools that may reflect the particular policy context of this Australian intervention:

> *There are signs that health and educational agendas are converging. There has, for example, been an increasing interest in the forms of schooling that might best meet the educational needs of early adolescents.*
>
> *Patton et al.[122]*

Two additional points concerning the rationale for integration were raised to a much lesser extent but are likely relevant to most of these interventions, these are: (1) an integrated curriculum provides opportunities for repetition and reinforcement of health messages and (2) when health messages are not overt (as they may not be if explored through literature or other subject matter) students may be less resistant to them.[108]

> *Optimal learning occurs when information is embedded in meaningful contexts, applications and multiple representations are provided, and there are opportunities for learners to generate personally relevant questions.*
>
> *Bechtel et al.[84]*

Several interventions aimed to integrate health and academic education by making learning more experiential, rooting learning in active participation and providing opportunities for practical skill development. For example, in the I-LST intervention:

> *. . . teachers were also encouraged to be certain that the activities were developmentally appropriate and that the lessons included facilitation, coaching and behavioural rehearsal of skills as well as student-centred learning strategies. . .*
>
> *Bechtel et al.[84]*

And in the DRACON intervention:

> *Through re-enactment or role-play the participants access a more meaningful experience of the conflict, including thoughts, feelings and body experiences. On the other hand, the participants can distance themselves from these experiences through the fictional character of role-playing. In this way they can explore alternative actions and their consequences.*
>
> *Malm and Löfgren[119]*

## Eroding boundaries between teachers and students

The role of the teacher was central to intervention-driven change. A recurring theme was that interventions aimed to transform relationships between students and teachers and, thus, achieve a better classroom climate. Although this theme resonates with the concept of boundary erosion, it also suggests that an elaboration of the latter is required, as has previously been argued.[152] Boundary erosion is not meant to

be an erosion of sources of difference between staff and students or a challenge to disparities of power, but rather a transformation of the quality and the strengthening of relationships so that these move beyond the instrumental task of academic education and embrace more affective dimensions and students' overall development. In the included studies, the positive relationships between teachers and students (and between pro-social peers) were often seen as providing fundamental stable relationships that would increase student connectedness to the classroom and school and invoke a sense of responsibility to act in a way that is beneficial for themselves, their peers and their teacher:

> Teacher–student relationships are a joint function of the unique characteristics of children [. . .] and teachers [. . .] and the cultural norms, values, and practices they bring to the relationship and to the classroom. Together these characteristics contribute to the climate of the classroom.
>
> *Brown* et al.[54]

Enhancing students' relationships with teachers was regarded as pivotal because it would enable teachers to act as role models for pro-social behaviours and to encourage these among students. The establishment of student–teacher bonds was regarded as important for normalising positive behaviours in students. Additionally, although not exclusive to student–teacher relationships, bonding with pro-social peers and adults both inside and outside the classroom was seen as important to ingraining these behaviours in students:

> Once strong bonds are established, individuals will tend to behave in a manner consistent with the norms and values of the individuals and groups with whom they associate. In turn, stronger pro-social bonds support positive belief formation against antisocial behaviors (e.g. adolescent substance use).
>
> *Brown* et al.[92]

Teachers' relationships with the curricula that they delivered, specifically their internalisation of the curriculum and the values promoted therein, were felt to be instrumental. Teachers invested in the integrated curriculum were expected to role-model desired behaviours expressed within the curriculum, further normalising these for students:

> When teachers embrace and practice the program's principles and implementation strategies, they establish a set of expectations and norms for behaviors in their classrooms, and children begin using those skills and behaviors.
>
> *Brown* et al.[54]

Steps to Respect highlights the role of teachers in changing social norms for young people by changing their attitudes to, in this instance, bullying:

> . . . the program seeks to change attitudes about the acceptability of bullying through clearly labelling bullying behaviour as unfair and wrong, increasing empathy for students who are bullied, and educating students about their responsibilities as bystanders to bullying.
>
> *Brown* et al.[93]

### Eroding boundaries between the classroom and the wider school

A recurring theme was the importance of interventions aiming to generalise learning from the classroom to the wider school environment. Engaging students outside the classroom enabled them to experience reinforcement for positive behaviours and to increase their broader sense of connectedness with the school.

Having opportunities at multiple levels for skill reinforcement was emphasised by several interventions. At the school level, providing an environment for students to use skills or behaviours developed through

the curriculum could be achieved by involving non-teaching members of staff in some of the intervention activities, as was the case, for example, in the PATHS and Positive Action interventions:

> *Ecologically oriented programs emphasize not only the teaching of skills, but also the creation of meaningful real-life opportunities to use skills and the establishment of structures to provide reinforcement for effective skill application.*
>
> *Greenberg and Kusché[106]*

> *The school climate component leads to changes in school-wide activities such as reinforcement and recognition of positive behavior and character attributes demonstrated by students.*
>
> *Flay and Allred[102]*

Positive reinforcement of pro-social skill demonstration could occur via explicit rewards or simply through students feeling good about themselves as a result of practising such behaviours. Positive reinforcement was often described as being strengthened through opportunities for demonstration in different areas of life, leading to further reinforcement. Demonstration of positive behaviours was then felt to contribute to a positive sense of self, leading to further positive behaviours. Such processes were regarded as enabling internalisation of curriculum content by students.

In the case, for example, of the PATHS intervention it was reported that:

> *. . . among other important uses, internalization is the primary process utilized in the development of an individual's conscience. When the outcome is positive, the conscience [. . .] works as a powerful system through which a person can 'take ownership' and achieve mastery over his or her own impulses and actions.*
>
> *Greenberg and Kusché[106]*

Many interventions, either directly (e.g. the Gatehouse Project[122]) or indirectly (e.g. the 4Rs intervention[3,51–57]), sought to improve students' sense of connectedness to the school environment, again, strongly resonating with the notion, in the theory of human functioning and school organisation, that eroding boundaries will encourage students to commit to school. A sense of connectedness and bonding with the school was seen to be linked to overall emotional well-being and security among students.

Connectedness was also seen to help foster better academic learning, as the school becomes a more positive environment that students are invested in.

## Eroding boundaries between schools and families

Some multicomponent interventions included family components, which were theorised to provide opportunities for students to apply their learning to engage in pro-social behaviours, in particular conflict resolution at home, and receive positive reinforcement for this. Some interventions also aimed to provide training to parents so that they might role-model the positive behaviours encouraged through the curriculum. For example, from the KAT intervention:

> *Where parents or other community members are actively involved in programmes, they are exposed to the same health-behaviour messages as younger participants and, if they accept those messages, can reinforce them through their own actions, behaviours and attitudes.*
>
> *Segrott* et al.[127]

For example, in the case of PATHS:

> *The extensive generalization procedures, teacher training, and focus on some level of parent participation used in PATHS have the goal of combining classroom instruction with efforts to create environmental support and reinforcement from peers, family members, school personnel, behavioural health professionals, and other concerned community members.*
>
> *Greenberg and Kusché[106]*

Interpreting this in the light of our overarching theme of boundary erosion, these interventions can be viewed as aiming to ensure consistent messaging and erosion of boundaries between learning and behaviour at school and at home.

## Additional mechanisms to reduce substance use

Although the notion of boundary erosion offered a useful overarching theme to structure our analysis, some mechanisms that emerged from our synthesis, although they did not contradict the existing thematic structure, could not be incorporated within it. As well as cultivating pro-social skills and relationships as outlined above, additional key mechanisms of change involved educating students about substance use and providing training in resistance skills (i.e. teaching students to resist peer pressure to use substances). For example, Positive Action:

> . . . targets the reduction of students' health-compromising and other negative behaviors (e.g., substance abuse, violence, disciplinary referrals, suspensions), while simultaneously enhancing health-enhancing and other positive behaviors (e.g., honesty, time management) and behavioral attributions directed at the self [. . .] and social relationships [. . .], with these leading to improved school-related performance.
>
> *Beets[85]*

For example, within RHC:

> . . . intervention components within [RHC] gradually shifted from early risk and protective factors in the social domains of school and family (e.g., academic performance, bonding, and parental monitoring) toward individual- and peer-related risk and protective factors (e.g., refusal skills, healthy beliefs, and associations with substance-using peers).
>
> *Brown et al.[92]*

## Hypothesised intervention outcomes

Most of the interventions aimed to reduce aggression, violence or bullying in young people through the mechanisms outlined above, usually in combination. Proximal outcomes, commonly highlighted across the interventions, that led to reduced violence included the development of pro-social skills, emotional intelligence (broadly understood as self-awareness) and self-management (encompassing managing emotions and responding constructively to bullying and conflict). These proximal outcomes were theorised as arising out of the mechanisms described above. Most authors made an overall assumption that antisocial behaviours occur because of deficits in these social and emotional skills.

For example, the Roots of Empathy intervention:

> . . . significantly improved the attitudes, knowledge and social emotional competencies of all participating teachers and children. A decrease in the frequency of bullying and an increase in pro-social behaviours were reported by all participants.
>
> *Cain and Carnellor[94]*

## Hypotheses arising from the theory synthesis

The discussion above suggests the following hypotheses:

1. Interventions that involve multiple components to erode multiple boundaries (e.g. health and academic education, staff–student, classroom–school and school–family) will be more effective than interventions that aim only to erode boundaries between health and academic education.
2. Interventions that aim for full integration of health education into academic subjects will be more acceptable to and less liable to resistance from students and, thus, more effective than interventions that deliver health and academic educational activities but without full integration.
3. Intervention effects on substance use and violence outcomes may be greater among students of low SES, as they are the most likely to benefit from boundary erosion.

It was not apparent from the synthesis of theories of change that there was a discrete subset of interventions that are of particular relevance to the UK. All of the theories of change seemed plausible within a UK context.

## Feedback from consultation on theory synthesis

Stakeholders were asked to reflect on our hypotheses (as above).

### On potential benefits

Adult policy stakeholders were positive about the theorised mechanisms of the potential for reductions in violence and substance use, as were some young people. Other young people wondered whether or not interventions might be effective in reducing in-school but not out-of-school risk behaviours.

### On multilevel programmes

Overwhelmingly, multilevel programmes with both a whole-school and a family component were seen as likely to be the most effective. Stakeholders suggested that a school ethos supportive of the programme's goals and approaches would facilitate whole-school programme components. Stakeholders commenting on the Welsh experience highlighted that the Welsh Assembly Government has promoted a health-promoting schools approach that would be conducive to multilevel programming. Young people generally agreed that developing stronger relationships between staff and students would be an important element of multilevel programmes.

It was generally perceived that a whole-school approach may be more feasible in primary schools, given that there is more flexibility. In primary schools, there may be dedicated time to health themes, for example a antibullying week, during which there may be approaches beyond the classroom to extend an integrated, health-promoting curricula.

The inclusion of parents, or having a family component, was universally agreed by stakeholders to be desirable. It was suggested that parents and families play an even stronger role in influencing student behaviour than school curricula or the school environment and, as such, parents and families performing a reinforcing role around health messaging was regarded as highly valuable, if possible. However, it was also agreed that engaging parents could be very challenging, particularly parents of students in secondary school. Parents of higher-risk students who may already be working with school officials were viewed as being easier to engage.

### On full integration into academic subjects

It was generally perceived that a fully integrated approach would be more likely to be feasible in primary schools, given that there may be more flexibility in the curriculum. In primary school, teachers engage with students for a full day and may teach one health topic across many academic subjects; whereas in secondary school, teachers typically see students only for specific academic subjects. Participants generally felt that some subjects (e.g. drama or art) may offer particularly appropriate platforms for practical skill development. There was also clear recognition that, if integrated programmes were able to support the achievement of academic learning objectives, receptiveness would generally be high. However, it was also pointed out that pay raises and teacher job security can hinge on their ability to support their students in achieving specific learning objectives, so it may be perceived as a personal risk to adopt a programme that may compromise core academic learning objectives. Where school leaders are judged partly on the basis of objectives concerning health and well-being, integrated programmes would be more likely to be taken up. If a school is judged largely based on academic performance then these programmes may be less likely to be taken up, despite their attempt to bridge health and academic learning objectives. Some stakeholders also suggested that some schools may prefer to implement health programmes that are clearly labelled as such, rather than engage in the more complex task of integrating health and academic education.

Stakeholders agreed that all of the intervention theories of change seemed plausible within a UK context.

# Chapter 5 Synthesis of process evaluations

## About this chapter

Parts of this chapter are reproduced or adapted from Tancred *et al.*[150] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/. The text below includes minor additions and formatting changes to the original text.

This chapter describes and reports on the quality of process evaluations that met our inclusion criteria. It then reports on our thematic synthesis of these studies, which focused on the factors that affected implementation and receipt in order to address research question 3.

## Included studies

Sixteen empirical studies of process were included, reporting on 12 different interventions.[57,84–86,89,94,107,108,117, 118,120,123,126,127,133,135] There were two studies of KAT,[126,127] three studies of Positive Action[85,86,118] and two studies of Roots of Empathy.[94,107] The remaining interventions (i.e. 4Rs,[57] DRACON,[120] English classes,[108] Hashish and Marijuana,[135] I-LST,[84] Peaceful Panels,[133] PATHS,[123] Steps to Respect[117] and the Gatehouse Project[89]) are reported on by one empirical study. The included studies reported quantitative and qualitative data. Some reported on standalone process evaluations, whereas other studies also included outcome evaluations. Of the 12 interventions evaluated, four focused on primary schools,[57,117,123,126,127] five on secondary schools[84,89,108,133,135] and three on both primary and secondary schools.[85,86,94,107,118,120] A summary of all included studies of process and interventions is given in *Appendix 18*.

## Quality of studies

The quality of study reports is detailed in *Appendix 19*, *Table 24*. This table presents a consolidated overview of quality after two reviewers (TT and CB) reached consensus. The original agreement rates on indicators of quality was 74%. Study reliability and usefulness varied. Only five reports were judged highly reliable and trustworthy,[85,86,107,117,123] and five reports provided insights of a high value in answering our research questions.[57,85,86,89,118] Six[57,89,118,126,127,133] and five[84,94,108,120,135] reports were judged 'medium' and 'low', respectively, in terms of reliability and trustworthiness.

## Characteristics of interventions, deliverers, participants and contexts that might facilitate successful implementation and receipt of interventions

Five overarching themes emerged, within which there were one or more subthemes related to facilitators of, or barriers to, implementation. These areas were support from school senior managers and the broader policy environment, teachers' immediate working environment, teachers' attitudes towards intervention characteristics, students' attitudes towards intervention characteristics, and parental support. These themes and the subthemes within them are introduced and described below. Although these themes arose inductively from the data, as with the synthesis of the theories of change, nonetheless certain subthemes resonated with the theory of human functioning and school organisation, as will be highlighted below.[77]

### Support from school senior managers and the broader political environment

### Support from senior managers

Support from school managers and other senior staff, including administrators, was cited as a key driver of successful programme implementation in eight study reports.[57,86,89,94,107,117,120,123] Among these, only two[94,120] were of low reliability owing to limited detail provided by authors on methods. Otherwise, these studies contributed useful findings, even though they were supported by limited qualitative data, which was a shortcoming of many of the process studies included in our review.

A supportive school climate was presented as not only one in which the ethos of the school aligns with the aims of the intervention but also one in which school managers significantly invest in the intervention and are willing to contribute to its success. In the case of the latter, providing mentoring and coaching to teachers involved directly in intervention delivery and committing dedicated time and resources to the curricula were viewed as important. In such cases, teachers who were typically engaged as intervention implementers were more likely to feel a sense of support and connectedness to the school.

Although deemed to be of medium reliability, findings from the Gatehouse Project process evaluation were deemed highly useful. The research team suggested that:

> . . . ongoing practical support from leadership has been acknowledged as important for mainstreaming the promotion of emotional well-being . . . [One teacher recalls] 'getting the time on the timetable, setting up a team, can't happen unless you've got someone in administration that thinks it's a great idea'.
>
> *Bond* et al.[89]

Furthermore, wider support across the school helped encourage teachers' beliefs in and acceptance of their responsibility to teach the respective curriculum (more on this subtheme below). For example, from one of the three highly reliable and useful studies of the Positive Action intervention in the USA, Beets *et al.*[86] reported that:

> . . . school leadership should develop a culture that encourages a shared [. . .] vision among staff and administration, is supportive of new innovations, and is aligned with the core values and concepts a given program is promoting . . .
>
> *Beets* et al.[86]

### Supportive policy environment

Broader support in terms of an intervention aligning with district-level priorities that leads to dedicated policies and funding to facilitate its implementation could be essential for successful implementation. Despite our judging it to be a low-quality study in terms of the reliability of its findings, the process evaluation of the Roots of Empathy intervention reported that:

> . . . the support of the Department of Education and Training [. . .] was crucial to the implementation and sustainability of the program in Western Australia. Because of the financial support of [the department's] Behaviour Standards and Wellbeing Directorate and its co-ordination of training, the program was successfully implemented.
>
> *Cain and Carnellor*[94]

Together, the factors identified within these two subthemes were presented as the ones promoting greater implementation fidelity for interventions that integrated health and academic education.

### Teachers' immediate working environment

Teachers' perceptions of their school as possessing a generally supportive culture for them and other teachers was viewed as enabling higher-quality implementation. This mindset was crucially linked to

teachers' confidence in their teaching environment as one that could support the intervention, which was in turn linked to teachers' own motivation for intervention delivery. This could have a direct impact on implementation. Three subthemes emerged from 10 studies.[57,84,89,94,108,118,120,123,127,135] Unfortunately, only one study among these[123] was designated as a high-quality study with very reliable findings, because of insufficient reporting on methods in the other studies. However, in terms of usefulness of findings, these studies explored important characteristics of implementation and, in some instances, had an explicit focus on the processes of integrating health and academic curricula.[84,108]

### Teachers working collaboratively and learning from one another

Within the two trials of Positive Action, one in Chicago[118] and one in Hawaii,[85,86] successful implementation was associated with teachers' perceptions of their schools having an innovative culture and strong relationships between teaching staff. These studies were of high reliability and usefulness because of their clarity of reporting and exploration of key implementation characteristics directly from the perspectives of implementers. This finding resonates with the notion of the erosion of boundaries, in this case among different staff members, that was first suggested in the synthesis of intervention theory but with the focus here on intervention delivery rather than the mechanism of effect. Various authors suggested that these findings were due to schools with a capacity for innovation being perceived as more open to change and more open to the introduction of approaches that give teachers more freedom to explore new areas.[86,118] Strong relationships between teachers were linked in several studies to a sense of mutual support and connectedness that teachers felt would help them to optimise intervention delivery.[84,89,108,118] All of these studies were deemed to have medium to high usefulness because of the comprehensiveness of the findings. However, none was deemed to be of high reliability because of poor reporting of methods.

### Teachers feeling well-prepared to deliver the curriculum

This subtheme was raised more than any other, across six different studies. It concerns teachers' attitudes to feeling prepared to deliver the curriculum, and to what aspects of interventions might help with this.[57,84,94,120,123,127] These studies varied in reliability, being of low,[84,94,120] medium[57,127] and high reliability,[123] largely based on the transparency of reporting on methods and findings. Findings presented alongside primary qualitative data were lacking in most cases. However, all but one study[120] were found to have medium or high usefulness because of the characteristics of implementation that were reported on and the level of detail included in these reports. The consistency with which this subtheme recurred suggests that this is essential to successful intervention delivery. This subtheme is linked to the first theme concerning supportive schools because much of this feeling of preparedness and confidence among teachers reflected support from senior managers within a school. More practically, the availability of resources, such as an easy-to-follow curriculum, adequate training and pre-prepared materials to use, was highlighted as being very useful to teachers.

For example, although deemed a lower-quality study, a process evaluation of the I-LST intervention in the USA compared the implementation of a non-integrated health curriculum with an integrated health curriculum, which was very useful.

> Teachers reported that the training was critical in adequately preparing them to integrate [life skills] components into their curriculum. They indicated that the training was especially effective in their development and implementation of infused lessons . . .
>
> *Bechtel* et al.[84]

### Teachers' workloads, burnout and administrative support

As most of these interventions were facilitated by teachers and required additional preparation from them to be effectively delivered, it is perhaps unsurprising that five studies identified teacher workload and/or burnout as a barrier to intervention implementation.[57,84,108,120,123] This may be partly addressed via higher-level

school management support, but other specific actions could also help prevent burnout. Some of these studies lacked methodological clarity, which reduced their reliability, but nonetheless provided useful information on implementation in their findings and generally rated as having medium to high usefulness.

Limited methodological detail was provided by Holcomb and Denk[108] in their study of the 'English classes' intervention. However, theirs was the only study to explore characteristics of integration in detail, as this programme required that health curricula be fully embedded within English classes.

> One noted program weakness related to the occasional need for teachers to locate resources other than those provided by project staff [. . .] Teachers' lack of time or access to information, in some cases, may have limited the amount of health content applied to individual lessons.
>
> *Holcomb and Denk[108]*

However, a collaborative culture among teachers, as well as support from school administrators, was regarded as helpful not only in ensuring preparedness for intervention delivery but also in coping with burnout and a heavy workload.

This subtheme links clearly to *Teachers feeling well-prepared to deliver the curriculum*, showing that teacher preparation for an additional curriculum responsibility could contribute to burnout. However, with adequate training and administrative support, teachers appeared to experience less burnout and were more likely to implement the intervention successfully.

In their high-quality study of PATHS, Ransford *et al.*[123] reported that:

> . . . teachers who perceived their school administration as more supportive reported higher implementation quality [. . .] Teachers who reported the highest levels of burnout and the most negative perceptions of curriculum supports reported the lowest levels of implementation dosage and quality.
>
> *Ransford* et al.[123]

### Teachers' attitudes towards intervention characteristics

A key theme in six studies[86,89,94,118,120,135] concerned the acceptability to teachers of the interventions themselves. This theme linked to the earlier theme about teachers' views on how supportive the school climate was of the implementation of these interventions. Save for two studies[120,135] that were poorly reported, the findings from these studies were designated as having medium to high reliability and usefulness. This subtheme was emphasised across studies as a factor enabling successful implementation to a greater extent than the acceptability of the intervention to students (see *Teachers' belief in and ownership of the curriculum*), probably because teachers were typically the primary deliverers of the interventions.

### Teachers' belief in and ownership of the curriculum

Coming to accept ownership of intervention objectives was found to be viewed as affecting the teachers' attitudes towards the curriculum and their beliefs in their responsibility to teach the new curricula, as well as giving the teachers a sense of ownership of the integrated curriculum.

For example, a methodologically rigorous and reliable study, Beets *et al.,*[86] reported from a highly useful evaluation of the Positive Action intervention in US primary schools that:

> . . . attitudes towards [Positive Action] were positively related to the amount of the Positive Action curriculum delivered.
>
> *Beets* et al.[86]

### Positive teacher attitude towards intervention potential

Teacher perceptions of the role of social and emotional learning, which was a part of the curriculum in all but two interventions included in this review, influenced their internalisation and subsequent role-modelling

of the behaviours promoted within the curriculum, a mechanism that was found to be important in our review of intervention theory. This was identified by a number of studies as being a critical mechanism of action through which outcomes could be achieved.[57,86,89,94,118,123] These studies were all found to be of medium to high usefulness and were generally presented with good methodological clarity. For example, this was a theme in an evaluation of the Roots of Empathy intervention in Western Australia that, although a low-quality study, was found to be of medium usefulness as a result of a detailed exploration of teacher perspectives.

> *All participants were committed to the importance of [social and emotional learning] in their teaching . . . they considered [it] essential to the academic learning that underpinned the teaching philosophy of all participants. The pedagogical understandings in the [. . .] program were consistent with each participant's philosophy of learning and teaching.*
>
> *Cain and Carnellor[94]*

Conversely, teachers' initial scepticism about the new intervention or teachers feeling that these interventions were a distraction from academic learning could be barriers to implementation. Although judged as of lower reliability, the process evaluation of the drama-based DRACON intervention[120] in Australia provides useful evidence that such a barrier operated in the case of this intervention:

> *A few [teachers] start with stronger reservations [. . .] and some [. . .] have chosen to withdraw from the project. These reservations are usually expressed as: not trusting drama to achieve its purpose, sometimes because it is perceived to potentially disrupt an orderly classroom, or to be too time-consuming in a full syllabus . . .*
>
> *O'Toole and Burton[120]*

### Teachers' freedom to be innovative and flexible
Curricula that were perceived by teachers to be adaptable to their classroom settings were generally implemented to a greater extent.[108,118,135] These studies were generally of medium to low reliability and usefulness but, importantly, two did provide explicit reflection on the nature of fully integrated curricula.[108,135] Some curricula were designed to be flexible and offered teachers scope to adapt components of the intervention in line with the academic goals for their classes and the interests of their students.

These findings resonate with our theory synthesis that interventions aiming to erode traditional barriers between health and academic education might be facilitated by a school culture among staff as well as an approach that is positively inclined towards challenging rigidity.

### *Students' attitudes towards intervention characteristics*

### Students' perception of the curriculum
As reported in *Teachers' attitudes towards intervention characteristics*, there was a sense that intervention acceptability to students was not as important to successful delivery as the attitudes of teachers. Nonetheless, the acceptability of the intervention to students was important in its own terms and could sometimes facilitate successful implementation. This was reported to be the case in six studies, particularly in relation to student views as to whether or not curriculum messages were relevant to their own lives.[85,108,126,127,133,135] However, only one of these studies was determined to be of high reliability and usefulness.[85] Holcomb and Denk[108] suggest that within 'English classes':

> *. . . it was generally believed that the high level of interest among students was generated by the relevance of the health topics, especially issues like use of drugs and alcohol, disease prevention, gun control, AIDS [acquired immune deficiency syndrome], and the environment to students' lives and experiences.*
>
> *Holcomb and Denk[108]*

### Students' pre-existing attitudes aligning with intervention ideals

Not unlike teachers' views about social and emotional learning, students' pre-existing views on intervention messaging, if positive, could facilitate good implementation and receipt of interventions. In Low et al.'s[117] high-quality study of the Steps to Respect intervention in the USA:

> . . . significant positive associations with students' engagement in the [Steps to Respect] lessons were found for classroom average levels of student support [of the intervention], [pre-existing] student attitudes against bullying, student climate and school connectedness.
>
> *Low* et al.[117]

### Interventions involved learning activities viewed positively by students

Similarly, intervention acceptability was greater when the learning activities that the interventions required were perceived to be relevant to students and fun to learn. For example, Wales[133] reported from a generally comprehensive and methodologically rigorous evaluation of the Peaceful Panels intervention in secondary schools in the USA that:

> . . . although the students were not unanimous in positive feelings about the program, the great majority of them stated that they enjoyed it [. . .] The students' positive feelings implied that students enjoyed learning through comics and it is possible that this helped them retain what they learned.
>
> *Wales*[133]

Students were particularly positive when an intervention integrating health and academic education encouraged teachers to focus on topics that were judged more relevant to students than traditional academic content, or to use more participative learning methods than would traditionally be the case. Students were particularly positive where an intervention integrating health and academic education encouraged teachers to focus on topics that were judged more relevant to students than traditional academic content, or to use more participative learning methods than would traditionally be the case. For example, Bechtel et al.[84] report that students liked the integration of education on substance use with academic learning.

### *Parental support*

### Lack of parental support as a barrier

Direct parental involvement was, in some cases, a feature of intervention activities,[126,127] and in such cases a lack of enthusiasm could detract from intervention feasibility. Additionally, when interventions included parental involvement through reinforcement and role-modelling of curriculum messaging,[57,126,127] parental engagement could vary with consequences for the success of the intervention. For example, in Sung's[57] evaluation of the 4Rs intervention in primary schools in the USA (which was rated of medium reliability but of high usefulness):

> . . . [an implementing teacher] viewed inconsistency between the way students are taught at school and at home as an impediment [. . .] [W]hereas she taught children to 'talk things out' without using violence in a conflict, some parents encouraged their children to use violence as means of solving social conflicts.
>
> *Sung*[57]

The synthesis of process evaluations identified important facilitators of, or barriers to, implementation or receipt but did not find evidence that would enable us to identify which interventions were or were not relevant to the UK. This finding confirmed the findings from the theory synthesis: that it is not possible to identify a discrete subset of interventions that are particularly appropriate to the UK context.

## Feedback from consultation on the process evaluation synthesis

After being presented with our preliminary findings on implementation, stakeholders acknowledged that receptiveness to and uptake of these types of integrated health and academic curricula hinged on several factors. The first is political will, in that there must be an external government mandate to support such programmes. The second, relatedly, is that there must be alignment of an integrated health and academic education programme within measured standards that reflect school performance. The third is internal political will, in that there must be support from management for the implementation of the programme as well as adequate budgetary allocation. Finally, there is a need for a programme curriculum that implementers (e.g. teachers) could be easily trained in, with ready-to-use materials. These factors are described in greater detail below.

Stakeholders agreed with our finding from the synthesis of process evaluations that it is not possible to identify a discrete subset of interventions that were most relevant to the UK. All interventions were potentially relevant to the UK with adaptation. The adaptations required would concern the detail of the intervention materials rather than the overall intervention approaches and theories of change.

### External political will

Each of the four countries within the UK have different policies for health and wellness in schools. Unfortunately, we did not speak with stakeholders from Scotland or Northern Ireland, but the differences between England and Wales are stark. In England, PSHE is not a statutory subject. In Wales, Personal and Social Education is part of the basic curriculum for pupils aged 7–16 years at maintained schools.[153] However, as of 2016 in the UK, Ofsted has introduced evaluation criteria linked to student health and welfare and personal development.[40] Integrated health and academic curricula may therefore help to address these wellness-based objectives as well as academic ones. As such, there may be varying levels of receptiveness to integrated health and academic curricula depending on the country context and the emphasis placed on health and wellness alongside academic attainment. The bottom line, according to stakeholders, was simply that there must be political prioritisation of the content within an integrated curriculum for it to have a chance at being taken up.

### Alignment with evaluated learning objectives

As we already introduced, in the previous stakeholder consultation feedback on the findings from our synthesis of intervention theory, integrated health and academic curricula were generally viewed positively, but there was consensus that they would be far more likely to be taken up if they helped to meet academic learning objectives, such as the standards assessed by Ofsted. Strongly related to the previous consultation feedback, stakeholders consistently reiterated the need for political will and for this to be communicated to schools, for example via school metrics and inspection standards.

Young people on the ALPHA group could readily see the interconnections between reduced substance use and violence and improved academic attainment. Violence and other antisocial behaviour in schools could hamper learning. Student substance use could impede ability to learn. Adult policy stakeholders acknowledged that it was widely accepted that improved health among students leads to better academic outcomes, but that these relationships are difficult to evaluate.

### Internal political will

There were comments from stakeholders on the necessity of buy-in, from both administrators, implementers (usually teachers) and students. The first point raised was that schools must see available programmes as having the capacity to address an issue that the school already recognises as a problem. For example, substance use may not be perceived by all schools as an issue that needs to be addressed and, therefore, programmes emphasising substance use, regardless of whether or not they are integrated, may not be taken up.

It was reiterated on a number of occasions that there must be champions throughout the school who are willing to take up and support an integrated curriculum. A strong evidence-base for the effectiveness of these programmes to achieve outcomes was mentioned by most stakeholders as being a necessary and compelling factor that would influence uptake. Additionally, it was conveyed that students must also see the programme as being relevant, which was a point made by young people on the ALPHA group as well as by the adult policy stakeholders.

Beyond having enthusiasm for a programme, from a very practical point of view, adequate budget allocation by the school to integrated health and academic curricula was highlighted, as well as sufficient time to train teachers (see *Effective teacher training*). Young people made the point that programmes would not be well delivered when teachers lacked the time and resources to prepare themselves for this form of teaching. Additionally, having trained staff within the school who could mentor implementers and address any problems was seen as highly useful, but often lacking in many health programmes currently offered.

### Effective teacher training

Teacher training was a point of interest. Many of the stakeholders had either led or been participants in continuing teacher training. There was a general agreement that it would be very challenging to get full days of teacher time to do training on integrated curricula. However, it was agreed that training in person, with lots of interactive activities and opportunities for role play, would generally be more effective than simply providing teachers with written materials. It was generally agreed that online training could be useful to supplement learning, but that it should not replace face-to-face training. The need to train the trainers was seen as being very important and also relates to having the necessary supports within schools. Furthermore, this type of model would allow a programme to be scaled up for delivery to a wider population.

One important consideration regarding training for delivering integrated health and academic curricula in general is that programmes adopting this platform must be flexible to tailor materials to each school's individual context. However, given the importance of an evidence base, it was also felt that programmes should have core components delivered with high fidelity to try to reproduce positive outcomes achieved elsewhere.

### Differences between primary and secondary schools

In addition to the factors outlined earlier in this chapter, there were a number of reflections on the feasibility of integrated health and academic education in primary schools versus secondary schools. Although it was generally agreed that such interventions could be taken up by secondary schools, it was viewed as considerably less likely. In secondary schools, a narrower focus on academic attainment often consumes the seemingly constantly reducing space for health and wellness curricula. Teachers in secondary schools face enormous pressure for their students to perform well academically and may be less likely to risk academic attainment by giving space to health education, especially within core academic subjects. With fewer classes such as art and drama, which may provide a more natural platform for these integrated curricula, especially when practical skill development is emphasised, delivery in secondary schools may be further challenged. Other comments from both policy stakeholders and young people included the difficulty in discussing substance use and violence in secondary schools where students may already be engaged in these behaviours. Furthermore, it was felt that primary schools, which often use topic-based learning across several different academic lessons, may be much more appropriate platforms for integrated curricula than secondary schools.

### Comparison with alternative personal, social and health education delivery models

One stakeholder drew our attention to McWhirter *et al.*'s[154] *Understanding Personal, Social, Health and Economic Education in Secondary Schools*, in which different approaches to delivering PSHE are described, with an indication of the advantages and disadvantages of each approach. This source acknowledges teaching PSHE across the curriculum within other subjects as one potential approach. The authors highlight some key advantages, such as avoiding the need for discrete time slots for PSHE and increasing ownership

of the curriculum across the school. Among the disadvantages highlighted, there are some that resonate with our own findings. Related to 'internal political will', McWhirter *et al.*[154] emphasise the need for 'absolute' commitment from the school leadership team. There are also several other key points raised that were not necessarily reflected in our findings, nor in our consultations with stakeholders. First, when health and academic objectives are brought together, there is a risk that the academic objectives may be seen as having lesser importance and, indeed, emphasis on health topics may become tokenistic. Second, if health education is integrated across a number of different subject lessons, then there is a risk that it will be difficult to ensure consistency and continuity. Third, the risk that evaluating learning objectives related to both health and academic education may prove challenging for teachers when monitoring student progress.[154]

# Chapter 6 Synthesis of outcome evaluations

## About this chapter

Parts of this chapter are reproduced or adapted from Melendez-Torres et al.[155] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/. The text below includes minor additions and formatting changes to the original text.

Parts of this chapter are reproduced or adapted from Melendez-Torres et al.[156] This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/. The text below includes minor additions and formatting changes to the original text.

This chapter introduces the included outcome evaluations that met our inclusion criteria. It describes the quality of the evaluations, followed by our narrative and meta-analytic syntheses of these studies to address research question 4.

## Included studies

We included 16 outcome evaluations of 14 interventions reported across 41 papers,[3,51–53,60,83,87,88,90–93,95–98,100,101,103,104,109–111,113–116,125,127–132,134,136–141] of which 26 papers contributed evidence to our synthesis of outcome evaluations.[3,51,52,60,83,87,90–92,95–98,101,104,109–111,114–116,125,127–129,132] Included outcome evaluations and their characteristics and quality appraisals are included in *Appendix 20*.

## Evaluation design

All included studies were designed as RCTs. In each trial, the comparator was treatment as usual. All trials were randomised at the school level except the Bullying Literature Project[96] and the Bullying Literature Project–Moral Disengagement,[136] which were randomised at the classroom level in two schools. Of the 13 evaluations, seven collected substance use outcomes,[60,83,87,88,90,91,93,97,98,100,101,103,104,109–111,113–116,125,127–132] 11 collected violence perpetration outcomes[3,51–53,83,87,93,96–98,100,101,103,104,109–111,113–116,125,129–131,134,136,139–141] and nine collected violence victimisation outcomes.[60,88,90,91,93,96,97,103,104,109–111,134,136–141] Five of the included evaluations also presented academic attainment outcomes,[3,51–53,83,87,92,93,95–116,129,130] although this was not an inclusion criterion. One evaluation was conducted in the UK (KAT[127]), one was conducted in Australia (the Gatehouse Project[90]) and one was conducted in the Democratic Republic of Congo;[137,138] all others were conducted in the USA.

Analyses varied in the length of follow-up time. In this chapter we report phases of education converted to UK school years. Only three interventions (i.e. the Bullying Literature Project,[96] Bullying Literature Project–Moral Disengagement[136] and KAT)[127] evaluated outcomes in the same term as when the intervention was administered. Two additional evaluations, Learning to Read in a Healing Classroom[137,138] and the second evaluation of Steps to Respect,[93] evaluated outcomes at the end of 1 year of the intervention. The first evaluation of Steps to Respect[104] included end-point data at the end of 1 intervention year, in addition to trajectory-based analyses over 2 years. The evaluation of 4Rs[52] included outcomes over 2 intervention years, results after 3 years were not available for this analysis. An additional five evaluations followed up participants

for 3 years (the Gatehouse Project,[90] I-LST,[128] PATHS,[97] Second Step[139–141] and Youth Matters[110]). Finally, four evaluations involved longer-term follow-up. The Hawaii evaluation of Positive Action[87] tested outcomes after the fifth year of the intervention, and the Chicago evaluation of Positive Action[115,116] followed up children over the 6 years of the intervention. RHC[92,95] followed up children over 8 intervention years, with follow-up extending for 1 or 2 years past the intervention; thus, participants could be followed up for up to 9 years from baseline. In addition, the evaluation of LIFT[100] recruited two cohorts [one in year 2 at baseline (for which results are only available for 1 year) and one in year 6 at baseline], which were followed up for 8 years.

Study populations in most evaluations drew from different year cohorts at baseline. Students receiving the Bullying Literature Project[96] intervention were drawn from year 4 and year 5 classrooms, and students in the trial of KAT[126,127] were largely drawn from years 5 and 6, with some students from year 4 in mixed-year classrooms. Learning to Read in a Healing Classroom was implemented in years 3 to 5.[137,138] As previously noted, LIFT[100] enrolled a cohort in year 2 and a cohort in year 6 at baseline, but only the year 6 cohort was followed up beyond a year and is relevant to this synthesis. The Hawaii evaluation of Positive Action[87] included students in year 2 or 3 at baseline, as did the evaluation of RHC.[95] Both evaluations of Steps to Respect drew on cohorts across multiple years. In the first evaluation,[104] students were drawn from years 4 to 7, whereas in the second evaluation,[93] students were drawn from years 4 to 6. Unusually, the evaluation of the Gatehouse Project[60,90,91] included students in one cohort from year 9 through to year 11 and also surveyed three consecutive cohorts of year 9 students at the end of 1 intervention year. Of the remaining six evaluations, three (4Rs,[52] PATHS,[97] and Positive Action Chicago[115,116]) recruited students in year 4, one (Youth Matters[111]) recruited students in year 5, one recruited students in year 7 (Second Step[139–141]) and one (I-LST[128]) recruited students in year 8.

## Risk of bias and quality of evidence

Quality was assessed using a modified version of The Cochrane Collaboration's tool for assessing risk of bias in randomised trials.[157] For many of the items, studies did not report sufficient information to render a judgement. Findings are presented below separated by the level of the evaluation.

### Random generation of allocation sequence

Of the 16 included evaluations, 10 evaluations, reported on in 21 publications, did not report enough information to judge the risk of bias arising from allocation sequence generation.[60,87,88,90–93,95–97,103,104,109–111, 128–130,132,134,136] The remaining six evaluations (4Rs,[3,52,53] KAT,[127] LIFT,[98,100,125] Second Step,[139–141] Learning to Read in a Healing Classroom[137,138] and Positive Action Chicago[83,113–116]) all presented enough information to appraise these evaluations as having a low risk of bias in this domain. Both 4Rs[52] and Positive Action Chicago[115,116] used random number generators, whereas LIFT drew allocations from a hat.[125] KAT used optimal allocation to determine the randomisation sequence.[127]

### Allocation concealment

In total, 15 out of 16 included evaluations did not state if or how allocation was concealed, thus preventing the determination of the risk of bias in this domain. Learning to Read in a Healing Classroom[137,138] was assessed as having a low risk of bias because of reported methods of concealment.

### Blinding

Although blinding is often difficult in trials of school interventions, blinding of outcome assessors may often be possible and is useful for reducing ascertainment bias. Of the 16 included evaluations, only one provided enough information to judge whether or not any blinding occurred. In LIFT,[98,125] outcome assessors were blind to allocation. This evaluation was therefore scored as being at a low risk of bias.

### Complete outcome data

In total, 15 out of 16 evaluations were judged as having a relatively low or balanced attrition when this was relevant to outcome assessment. In some evaluations, such as the evaluations of Positive Action,[87,113–116] whole-school intervention implementation meant that students were followed up at a certain grade level regardless of the duration of their time in the school or whether or not baseline data were collected. Learning to Read in a Healing Classroom[137,138] was assessed as having a high risk of bias because of the high levels of missing data.

### Complete reporting

Of the 16 evaluations, 12 were judged as being at an unclear risk of selective reporting. This was generally due to a lack of a trial protocol. Two evaluations, 4Rs[51] and Second Step,[139–141] were judged as being at a high risk of bias from selective reporting because of unclear and incomplete reporting of findings from the 3-year results. Bullying Literature Project–Moral Disengagement and Learning to Read in a Healing Classroom were both judged as being at a low risk of bias because all expected outcomes were reported.

### Confounding addressed

Although randomisation should on average produce groups balanced for known and unknown confounders, one of the issues arising in school-randomised or classroom-randomised trials is that the relatively small number of higher-level units (i.e. schools) can produce imbalances between arms. Of the 16 evaluations, 10 were at a low risk of bias from confounding. In these evaluations, confounding was addressed by including a vector of person-level confounders in multivariate regression models. The remaining six evaluations were at an unclear risk of confounding because inadequate information was presented in outcome evaluations.

### Accounting for clustering

Evaluations generally were at a low risk of bias owing to inadequate assessment of clustering because clustering was generally well assessed. Of the 16 evaluations, 13 (reported in 35 publications) used appropriate analytical methods to account for clustering;[3,51–53,60,83,87,88,90,91,93,97,98,100,101,103,104,109–111,113–116,125,127–130,132,137–141] for example, robust standard errors in the Gatehouse Project[90,91] or generalised linear mixed-effects models to nest students within schools (e.g. 4Rs, Positive Action Hawaii and Chicago, PATHS and Youth Matters[3,51–53,83,87,109–111,113,114,116,129,130]). One evaluation, I-LST,[97,115,128,132] presented results that did not account for clustering but observed that results accounting for clustering were not different. One evaluation had an unclear risk of bias (RHC[92,95]) and two had a high risk of bias in this domain (Bullying Literature Project[96,134] and Bullying Literature Project–Moral Disengagement[136]).

### Additional sources of bias

Most evaluations (14/16, reported on in 35 publications)[3,51–53,60,83,87,88,90–93,95,97,98,100,101,103,104,113–116,125,127–130,132] used a variety of methods to reduce other sources of bias; for example, additional methods of addressing missing data or sensitivity analyses for different model assumptions. Two evaluations, Youth Matters[110,111] and Bullying Literature Project,[96,134] did not present the relevant information to permit a clear determination of other sources of bias.

### Suitable control group

Of the 16 evaluations, 10 (reported in 29 publications) reported enough information to judge that the intervention and control groups were similar on key characteristics or accounted for differences in the analysis.[3,51–53,60,83,87,88,90–92,95,109–111,113–116,128–130,132,136–141] However, six evaluations (reported in 12 publications) did not provide enough information to judge whether or not the intervention and control groups were sufficiently similar; the evaluations did provide evidence of baseline differences between the groups but did not demonstrate how these differences were accounted for.[93,96–98,100,101,103,104,125,127,131,134]

A visual summary of possible sources of bias by intervention is presented in *Figure 8*.
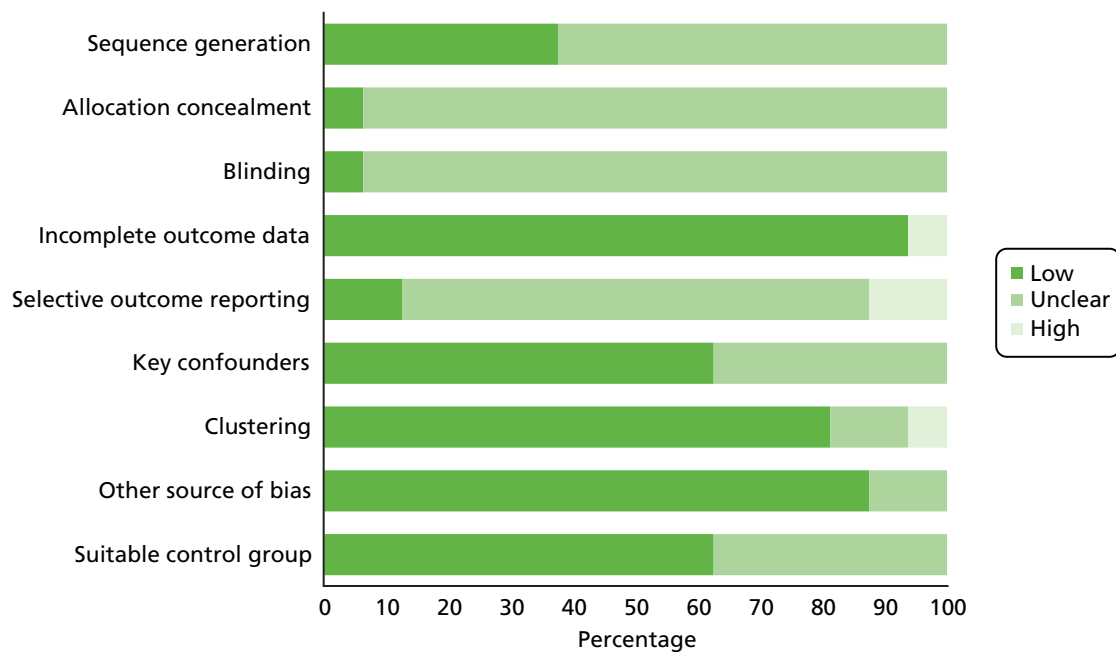
**FIGURE 8** Possible sources of bias within eight dimensions by intervention.

### *Grading of Recommendations, Assessment, Development and Evaluations assessment of primary outcomes*

The quality of evidence for outcomes by KS is presented in *Table 3*. For most combinations of outcome and KS, the quality of evidence was low or very low. However, for substance use outcomes in KS3 and 4 and for violence victimisation outcomes in KS2, quality of evidence was moderate.

## Effects of interventions on substance use

Seven of the included intervention evaluations presented substance use outcomes (*Tables 4* and *5*). Three evaluations presented substance use outcomes measured in KS2: KAT,[127] Positive Action Chicago[116] and Positive Action Hawaii.[87] Five evaluations presented substance use outcomes measured in KS3: the Gatehouse Project,[60,90] LIFT,[101] I-LST,[128] Positive Action Chicago[114] and RHC.[92] Three evaluations presented substance use outcomes measured in KS4: the Gatehouse Project,[90,91] I-LST[132] and RHC.[92] Finally, only LIFT[98] presented substance use outcomes in KS5.

### *Key stage 2 substance use outcomes*

Of the three evaluations presenting substance use outcomes for KS2, one evaluation, Positive Action Chicago,[116] presented an omnibus outcome comparing counts of substances used. Both the KAT[127] and Positive Action Hawaii[87] interventions presented outcomes relating to alcohol use. In addition, Positive Action Hawaii[87] presented outcomes relating to smoking and illicit drug use.

#### Omnibus substance use outcomes

At the end of the third intervention year (corresponding to year 6), intervention students in the Positive Action Chicago trial[116] had a lower count of types of substance use than control students [incidence rate ratio (IRR) 0.69, 95% CI 0.50 to 0.97]. In the Positive Action Hawaii trial,[87] intervention students reported fewer substances used (IRR 0.41, 90% CI 0.25 to 0.66), but differences between groups were not significant for teacher report (IRR 0.66, 90% CI 0.30 to 1.45).

**TABLE 3** The GRADE assessment summary by outcome

| Quality assessment | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Number of studies | Study design | Risk of bias | Inconsistency | Indirectness | Imprecision | Other considerations | Summary of effects | Quality |
| *Substance use in KS2* | | | | | | | | |
| 3 | Randomised trials | Serious[a] | Serious[b] | Not serious | Very serious[c] | None | Interventions were inconsistent in their impact on substance use, although two large trials[87,116] suggested some impact across different substance use outcomes | ⊕◯◯◯ Very low |
| *Substance use in KS3* | | | | | | | | |
| 5 | Randomised trials | Serious[a] | Not serious | Not serious | Not serious | None | In a meta-analysis based on 55 standardised MDs from five evaluations,[60,90,92,114,128] interventions globally reduced substance use ($d = -0.09$, 95% CI $-0.17$ to $-0.01$) | ⊕⊕⊕◯ Moderate |
| *Substance use in KS4* | | | | | | | | |
| 3 | Randomised trials | Serious[a] | Not serious | Not serious | Not serious | None | In a meta-analysis based on 34 standardised MDs from three evaluations,[90–92,132] interventions globally reduced substance use ($d = -0.06$, 95% CI $-0.09$ to $-0.02$) | ⊕⊕⊕◯ Moderate |
| *Substance use in KS5* | | | | | | | | |
| 1 | Randomised trials | Serious[a] | Not serious | Not serious | Serious[d] | None | Findings from one trial[98] suggested some effect in reducing alcohol use, smoking and, to a less certain degree, illicit drug use | ⊕⊕◯◯ Low |
| *Violence perpetration in KS2* | | | | | | | | |
| 9 | Randomised trials | Serious[a] | Serious[e] | Serious[f] | Serious[g] | None | Evaluations reporting this outcome[52,87,96,97,104,110,116,125,136] were inconsistent in their effectiveness and in their reporting of outcome measures | ⊕◯◯◯ Very low |

**TABLE 3** The GRADE assessment summary by outcome (*continued*)

| Quality assessment | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Number of studies | Study design | Risk of bias | Inconsistency | Indirectness | Imprecision | Other considerations | Summary of effects | Quality |
| *Violence perpetration in KS3* | | | | | | | | |
| 2 | Randomised trials | Very serious[h] | Very serious[e] | Serious[f] | Serious[i] | None | Findings from three trials[109,115,139–141] suggested dissimilar effects on bullying perpetration in KS3 | ⊕◯◯◯ Very low |
| *Violence victimisation in KS2* | | | | | | | | |
| 5 | Randomised trials | Serious[a] | Not serious | Not serious | Not serious | None | Findings from seven evaluations[93,96,97,104,136,137] contributed to a meta-analysis that did not suggest a significant effect of these interventions on reducing violence victimisation ($d = -0.04$, 95% CI −0.14 to 0.07) | ⊕⊕⊕◯ Moderate |
| *Violence victimisation in KS3* | | | | | | | | |
| 2 | Randomised trials | Serious[a] | Not serious | Not serious | Serious[i] | None | Findings from three trials[90,109,111,139–141] did not suggest a convincing effect of interventions on violence victimisation, specifically bullying, in KS3 | ⊕⊕◯◯ Low |
| *Violence victimisation in KS4* | | | | | | | | |
| 1 | Randomised trials | Serious[a] | Not serious | Not serious | Very serious[j] | None | Findings from one trial[90] did not suggest that the intervention was associated with less violence victimisation, specifically bullying victimisation, in KS4 | ⊕◯◯◯ Very low |

a  Evidence was downgraded because many appraisal items could not be assessed.
b  Evidence was downgraded because of the wide spread of effect estimates in findings relating to alcohol.
c  Evidence was downgraded because of the wide spread of effect estimates, the lack of evidence within substance use types and the heterogeneity of trials and follow-up times.
d  Evidence was downgraded because only one trial contributed effect estimates to this analysis.
e  Evidence was downgraded because of the wide spread in effect estimates.
f  Evidence was downgraded because several outcomes measuring violence perpetration captured 'aggression', including non-physical aggression, as a construct.
g  Evidence was downgraded because CIs for effect estimates in several included evaluations suggested the possibility of appreciable harm.
h  Evidence was downgraded because one trial randomised schools and then consented them.
i  Evidence was downgraded because of the few trials considered as part of this outcome.
j  Evidence was downgraded because one trial contributed effect estimates to this analysis and the CIs included appreciable harm.

**TABLE 4** Measures used in studies reporting substance use outcomes

| Evaluation | Measure | Notes |
|---|---|---|
| The Gatehouse Project | Substance use (multicategorical) | No recent use of alcohol, tobacco or cannabis. Recent use of at least one substance. Recent heavy use of one substance or more |
| | Any drinking | Any drinking in the last month |
| | Binge drinking | $\geq 5$ drinks in a row in the last month |
| | Regular drinking | $\geq 3$ drinking days in the past week |
| | Any smoking in the last month | Any smoking in the last month |
| | Regular smoking | $\geq 6$ smoking days in the past week |
| | Any cannabis use | Cannabis use in the last 6 months |
| | Weekly cannabis use | Weekly or more frequent cannabis use in the last 6 months |
| KAT | Drunk in the last 30 days | |
| | Ever been drunk | |
| | Ever had an alcoholic drink | |
| | Had a drink in the last 30 days | |
| LIFT | Initiation of alcohol | |
| | Patterned alcohol use | Use of alcohol at least once every 2–3 months in the past year |
| | Ever used tobacco | Exchangeable with initiation of tobacco |
| | Ever used marijuana | Exchangeable with initiation of marijuana |
| | Initiation of illegal drugs | Includes marijuana, amphetamines, heroin and cocaine |
| I-LST | Alcohol | Based on mean frequency count: never/a few times but not in the last year, a few times per year, once per month, a few times per month, once per week, a few times per week, once per day, more than once per day |
| | | All outcomes were presented stratified by sex, but data on smokeless tobacco use was presented for males only |
| | Binge drinking | |
| | Drunkenness | |
| | Cigarettes | |
| | Smokeless tobacco (males only) | |
| | Marijuana | |
| | Inhalants | |

**TABLE 4** Measures used in studies reporting substance use outcomes (*continued*)

| Evaluation | Measure | Notes |
|---|---|---|
| Positive Action Chicago | Substance use count | Based on a count of lifetime substance use: smoking, drinking, drunkenness, marijuana and 'other more serious drugs' |
| | Substance use frequency | Average of frequency scores for five substances: smoking, drinking, drunkenness, marijuana and 'other more serious drugs' |
| | Lifetime use of alcohol | |
| | Lifetime use of alcohol: more than once | |
| | Lifetime drunkenness | |
| | Lifetime drunkenness: more than once | |
| | Lifetime use of cigarettes | |
| | Lifetime use of cigarettes: more than once | |
| | Lifetime use of marijuana | |
| | Lifetime use of marijuana: more than once | |
| Positive Action Hawaii | Lifetime use of alcohol | Assessed by both student and teacher report |
| | Lifetime drunkenness | Assessed by student report only |
| | Lifetime tobacco use | Assessed by both student and teacher report |
| | Lifetime use of an illegal drug | Assessed by both student and teacher report |
| | Lifetime 'high' on illegal drug(s) | Assessed by student report only |
| RHC | Any alcohol use | Use in the past year |
| | Frequency of alcohol use | Six-point scale: no use in the previous year to ≥ 20 times in the last month |
| | Any cigarette use | Use in the past year |
| | Frequency of cigarette use | Six-point scale: no use in the previous year to > 40 cigarettes per day |
| | Any marijuana use | Use in the past year |
| | Frequency of marijuana use | Six-point scale: no use in the previous year to ≥ 20 times in the last month |

## Alcohol use outcomes

Findings for the effectiveness of included interventions on alcohol use outcomes drew on two evaluations, which were significantly different in follow-up period, design and execution. Outcomes for alcohol use in KAT, which were measured during the first intervention year and included students in years 5 and 6, were inconsistent and had wide CIs.[127] Intervention students were more likely, but not significantly more likely, to have been drunk in the last 30 days (OR 1.5, 95% CI 0.4 to 5.8) and to have ever been drunk (OR 1.7, 95% CI 0.5 to 6.8), and significantly more likely to have ever had an alcoholic drink (OR 5.3, 95% CI 1.2 to 23.9). However, intervention students were less likely, but not significantly less likely, to have had a drink in the last 30 days (OR 0.7, 95% CI 0.2 to 2.5). In the Positive Action Hawaii trial,[87] students and teachers were queried about students' alcohol use in year 6 (corresponding to either the fourth or fifth intervention year, depending on when students began the trial). Intervention students were less likely to report ever having used alcohol than control students (OR 0.48, 90% CI 0.34 to 0.68) and less likely to report having ever been drunk (OR 0.30, 90% CI 0.15 to 0.57) but teacher reports on student alcohol use were not significant (OR 0.81, 90% CI 0.41 to 1.58).

**TABLE 5** Analysis schematic for substance use

| Intervention | Cohort | Year | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| The Gatehouse Project | | | | | | | | | ▨ | ▨ | ▨ | | |
| KAT | Year 5 (some from year 4) | | | | ▨ | | | | | | | | |
| | Year 6 | | | | | ▨ | | | | | | | |
| I-LST | | | | | | | ▨ | | ▨ | ▨ | | | |
| LIFT | Fifth grade | | | | | | | ▨ | | | | | ▨ |
| | | | | | | ▦ | ▦ | ▦ | ▦ | ▦ | ▦ | ▦ | ▦ |
| Positive Action Chicago | | | | | | ▨ | | ▨ | | | | | |
| Positive Action Hawaii | First grade | | | | | ▨ | ▨ | | | | | | |
| | Second grade | | | | | ▨ | ▨ | | | | | | |
| RHC | First grade | | | | | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | |
| | | | | | | ▦ | ▦ | ▦ | ▦ | ▦ | ▦ | ▦ | |
| | Second grade | | | | | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | |
| | | | | | | ▦ | ▦ | ▦ | ▦ | ▦ | ▦ | ▦ | |

Green cells indicate intervention delivery. Diagonally striped cells indicate an end-point measurement. Dotted cells indicate a longitudinal, trajectory-based analysis.

## Smoking outcomes

As with alcohol, intervention students in year 6 in the Positive Action Hawaii trial[87] reported being less likely to have ever smoked than control students (OR 0.52, 90% CI 0.31 to 0.88) but teacher report for the same was not significant (OR 0.54, 90% CI 0.28 to 1.02).

## Illicit drug use outcomes

Intervention students in year 6 in the Positive Action Hawaii trial[87] were less likely to report ever having been high on illegal drugs (OR 0.20, 90% CI 0.09 to 0.44) or ever having used illegal drugs (0.28, 90% CI 0.14 to 0.54). Teacher report on student use of illegal drugs matched these findings closely (OR 0.27, 90% CI 0.10 to 0.72).

## Meta-analyses

Because of the sparse number of studies in each category of substance use and the difference between interventions and follow-up periods, we elected not to undertake a meta-analysis of substance use outcomes in KS2.

### *Key stage 3 substance use outcomes*

Five evaluations reported outcomes relating to substance use in KS3: the Gatehouse Project,[60,90] LIFT, I-LST,[128] Positive Action Chicago[114] and Raising Health Children.[92] Of these, the Gatehouse Project[60] and Positive Action Chicago[114] reported omnibus substance use outcomes. Alcohol, smoking and illicit drug use outcomes were reported by the same five evaluations: the Gatehouse Project,[60,90,91] LIFT,[101] I-LST,[128] Positive Action Chicago[114] and RHC.[92]

## Omnibus substance use outcomes

The two evaluations using omnibus measures of substance use outcomes in KS3 related to different interventions and follow-up periods. In the Gatehouse Project,[60] intervention recipients in year 9 were not different in substance use patterns than control recipients at the end of the first intervention year. This was the case for each of the first (OR 0.92, 95% CI 0.7 to 1.2), second (OR 0.84, 95% CI 0.61 to 1.18) and third (OR 0.85, 95% CI 0.65 to 1.12) cohorts reported on. However, in Positive Action Chicago,[114] intervention participants in year 9 used fewer substances ($g = -0.29$, 95% CI $-0.51$ to $-0.08$) and used substances less ($g = -0.27$, 95% CI $-0.48$ to $-0.05$) than control participants at the end of the sixth intervention year.

## Alcohol use outcomes

Within alcohol use outcomes reported in KS3, two evaluations (the Gatehouse Project[60] and I-LST[128]) reported follow-up measurements at the end of the first or second intervention year. Evidence from these two evaluations were not suggestive of effectiveness. Compared with control participants and at the end of the first intervention year, the Gatehouse Project[60] intervention participants in year 9 were not less likely to report any drinking (OR 0.93, 95% CI 0.71 to 1.21), binge drinking (OR 0.95, 95% CI 0.69 to 1.32) or regular drinking (OR 1.09, 95% CI 0.77 to 1.57). At the end of the first intervention year (year 8) in I-LST,[128] neither males nor females in the intervention group were less likely to report different frequency of alcohol use (males: MD $-0.07$, $t = -0.61$; females: MD $-0.23$, $t = -1.54$) or frequency of drunkenness (males: MD $-0.08$, $t = -0.82$; females: MD $-0.08$, $t = -0.72$). However, whereas male students were not significantly different on binge drinking (0.00, $t = -0.01$), female students in the intervention group reported a lower frequency of this outcome at the end of the first intervention year (MD $-0.25$, $t = -2.11$). At the end of the second intervention year (year 9), there were no differences between groups on any of these outcomes: frequency of alcohol use (males: MD 0.03, $t = 0.20$; females: MD $-0.13$, $t = -0.94$), frequency of binge drinking (males: MD 0.04, $t = 0.32$; females: MD $-0.09$, $t = -1.01$) or frequency of drunkenness (males: MD 0.08, $t = 0.79$; females: MD $-0.07$, $t = -0.69$).

However, the three evaluations reporting longer follow-up periods in KS3 presented a stronger picture of effectiveness. Intervention participants in the evaluation of LIFT surveyed in year 9 (the fourth intervention year) were less likely to initiate patterned alcohol use than control participants (OR 0.67; $p < 0.01$).[98–101,125]

Similarly, at the end of the sixth intervention year (year 9), intervention participants in the Positive Action Chicago trial[114] were less likely than their control counterparts to report ever having used alcohol ($g = -0.35$, 95% CI –0.57 to –0.14), having used alcohol more than once ($g = -0.35$, 95% CI –0.57 to –0.14), ever having been drunk ($g = -0.29$, 95% CI –0.51 to –0.07) or having been drunk more than once ($g = -0.22$, 95% CI –0.43 to –0.002). In RHC, proportions using alcohol were not different between intervention ($n = 466$) and control ($n = 493$) participants in year 7, corresponding to the sixth intervention year (29% vs. 30%, respectively), year 8, corresponding to the sixth or seventh intervention year depending on the cohort in which students began the intervention (33% vs. 29%, respectively) and year 9, corresponding to the seventh or eighth intervention year (37% vs. 40%, respectively). However, frequency of use estimates were not available by group for this intervention at these follow-up periods.

## Smoking outcomes

Across evaluations, evidence did not consistently suggest that interventions were effective in preventing smoking. Although the Gatehouse Project intervention participants were less likely than their control counterparts in year 9 (end of first intervention year) to report being regular smokers (OR 0.66, 95% CI 0.46 to 0.95), groups were not significantly different in terms of the odds of ever smoking (OR 0.89, 95% CI 0.72 to 1.12).[90] At the end of the first intervention year (year 8) in I-LST,[128] male intervention and control participants were not significantly different in frequency of use of either cigarettes (MD 0.16, $t = 1.08$) or smokeless tobacco (MD 0.13, $t = 0.73$). However, young women in the intervention group reported lower frequency of cigarette use (MD –0.40, $t = -2.01$). These patterns were similar at the end of the second intervention year (year 9) for both young men (cigarettes: MD –0.17, $t = -0.72$; smokeless tobacco: MD 0.07, $t = 0.35$) and women (cigarettes: MD –0.61, $t = -2.39$).

Longer-term follow-up measurements presented a similarly inconsistent picture. After the fourth intervention year in LIFT (year 9), odds of tobacco initiation were not significantly less in intervention recipients (OR 0.90, not significant).[101] After the sixth intervention year of the Positive Action Chicago trial,[114] intervention participants were less likely than control participants at marginal significance to report ever using cigarettes ($g = -0.21$, 95% CI –0.43 to 0.01) but not less likely to have used cigarettes more than once ($g = 0.03$, 95% CI –0.19 to 0.25). Proportions trying cigarettes in the evaluation of RHC[92] were not meaningfully different in year 8 students (during the sixth or seventh intervention year; 9% in the intervention group vs. 8% in the control group) or in year 9 students (during the seventh or eighth intervention year; 14% vs. 13%), although estimates relating to frequency of use were not presented by group.

## Illicit drug use outcomes

### *Marijuana*

As for smoking outcomes, evidence on the whole did not consistently suggest that interventions were effective in preventing illicit drug use. Reported illicit drug use primarily focused on cannabis use, although one evaluation, I-LST,[128] reported on inhalant use, for which outcomes are presented separately below.

In the Gatehouse Project,[90] intervention participants were not less likely than control participants to report cannabis use in the last 6 months at the end of the first intervention year, corresponding to year 9 (OR 0.98, 95% CI 0.69 to 1.4). Male students receiving I-LST[128] were not different from their control group peers in frequency of marijuana use at the end of either the first intervention year (MD 0.00, $t = 0.00$) or the second intervention year (MD 0.05, $t = 0.31$). Although young women receiving I-LST reported less frequent marijuana use at the end of the first intervention year (MD –0.26, $t = -2.18$), differences were not significant at the end of the second intervention year (MD –0.17, $t = -1.71$).

Longer-term follow-up measurements were inconsistent across evaluations. Participants in LIFT[101] were less likely to report initiation of marijuana use than their peers after the fourth intervention year (year 9), but this difference was not significant (OR 0.81, not significant). However, participants in Positive Action Chicago[114] were less likely than their control peers at the end of the sixth intervention year (year 9) to report having ever used marijuana ($g = -0.23$, 95% CI –0.44 to –0.01) or having used marijuana more

than once ($g = -0.17$, 95% CI –0.39 to 0.05), although the last finding was not statistically significant. As with alcohol and smoking outcomes, proportions trying marijuana in the evaluation of RHC[92] were not meaningfully different in year 8 students (during the sixth or seventh intervention year; 8% in the intervention group vs. 9% in the control group) or in year 9 students (during the seventh or eighth intervention year; 16% vs. 18%), although estimates relating to frequency of use were not presented by group.

### Inhalants

Young men receiving I-LST[128] were not different from their control condition peers in frequency of inhalant use at the end of either the first intervention year (year 8) (MD –0.09, $t = -1.28$) or the second intervention year (year 9) (MD 0.09, $t = 1.86$). Young women had a similar pattern of results (end of first intervention year: MD –0.05, $t = -0.75$; end of second intervention year: MD 0.08, $t = 1.88$).

### Meta-analyses

As specified in the protocol,[1] we undertook separate meta-analyses for alcohol use, smoking and illicit drug use. Because the preponderance of outcomes relating to illicit drug use measured marijuana use, we also meta-analysed marijuana outcomes separately as an exploratory analysis. We did not undertake meta-analysis for omnibus substance use outcomes alone given that only two evaluations would have been included. Finally, as specified in the protocol,[1] we undertook an overall meta-analysis of substance use outcomes in KS3.

Interventions integrating academic and health education had a marginally non-significant effect in reducing alcohol use among students in KS3 ($d = -0.11$, 95% CI –0.23 to 0.004) (*Table 6* and *Figure 9*). However, this analysis, which included 23 effect sizes across five evaluations, had substantial between-studies heterogeneity ($I^2 = 54\%$). Meta-analyses were not indicative of an effect of these interventions for smoking specifically ($d = -0.05$, 95% CI –0.12 to 0.02) (*Figure 10*; see *Table 6*), based on 13 effect sizes from five evaluations. However, these interventions were effective in reducing illicit drug use ($d = -0.07$, 95% CI –0.14 to –0.01) and marijuana use specifically ($d = -0.10$, 95% CI –0.16 to –0.04) (*Figure 11*; see *Table 6*). Unlike the meta-analysis of alcohol outcomes, meta-analyses for smoking and illicit drug use all had negligible between-studies heterogeneity ($I^2 = 0\%$). A pooled analysis across all substance use outcomes, including 55 effect sizes from five evaluations, suggested a small but statistically significant global effect of these interventions ($d = -0.09$, 95% CI –0.17 to –0.01), albeit with some between-studies heterogeneity ($I^2 = 35\%$) (*Figure 12*; see *Table 6*).

### *Key stage 4 substance use outcomes*

Three evaluations reported outcomes relating to substance use in KS4: the Gatehouse Project,[90,91] I-LST[132] and RHC.[92] All outcomes reported related to alcohol use, smoking and illicit drug use; that is, no included evaluations reported omnibus substance use outcomes in KS4. All evaluations reported each of these three outcome types.

**TABLE 6** Summary of KS3 substance use outcomes

| Outcome | $k$ ($n$) | Standardised MD (95% CI) | $I^2$, study level (%) |
|---|---|---|---|
| Alcohol | 5 (23) | –0.11 (–0.23 to 0.004) | 54 |
| Smoking | 5 (13) | –0.05 (–0.12 to 0.02) | 0 |
| Illicit drug use: marijuana | 5 (10) | –0.10 (–0.16 to –0.04) | 0 |
| Illicit drug use | 5 (14) | –0.07 (–0.14 to –0.01) | 0 |
| All drug use outcomes | 5 (55) | –0.09 (–0.17 to –0.01) | 35 |

| Study | Paper | Outcome | Year | Effect size (95% CI) |
|-------|-------|---------|------|----------------------|
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any drinking in last 6 months | 9 | −0.04 (−0.19 to 0.11) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Binge drinking in last month | 9 | −0.03 (−0.21 to 0.15) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular drinking | 9 | 0.05 (−0.15 to 0.24) |
| LIFT | Eddy 2003[101] | Patterned alcohol use | 9 | −0.22 (−0.39 to −0.06) |
| I-LST | Smith 2004[128] | Alcohol, males | 8 | −0.08 (−0.32 to 0.17) |
| I-LST | Smith 2004[128] | Alcohol, males | 9 | 0.03 (−0.22 to 0.27) |
| I-LST | Smith 2004[128] | Alcohol, females | 8 | −0.21 (−0.47 to 0.06) |
| I-LST | Smith 2004[128] | Alcohol, females | 9 | −0.13 (−0.39 to 0.14) |
| I-LST | Smith 2004[128] | Binge drinking, males | 8 | −0.00 (−0.25 to 0.24) |
| I-LST | Smith 2004[128] | Binge drinking, males | 9 | 0.04 (−0.20 to 0.29) |
| I-LST | Smith 2004[128] | Binge drinking, females | 8 | −0.28 (−0.55 to −0.02) |
| I-LST | Smith 2004[128] | Binge drinking, females | 9 | −0.14 (−0.40 to 0.13) |
| I-LST | Smith 2004[128] | Drunkenness, males | 8 | −0.10 (−0.35 to 0.14) |
| I-LST | Smith 2004[128] | Drunkenness, males | 9 | 0.10 (−0.15 to 0.34) |
| I-LST | Smith 2004[128] | Drunkenness, females | 8 | −0.10 (−0.36 to 0.17) |
| I-LST | Smith 2004[128] | Drunkenness, females | 9 | −0.09 (−0.36 to 0.17) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime drunkenness | 9 | −0.29 (−0.51 to −0.07) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime drunkenness more than once | 9 | −0.22 (−0.43 to −0.00) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of alcohol | 9 | −0.35 (−0.57 to −0.14) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of alcohol more than once | 9 | −0.35 (−0.57 to −0.14) |
| RHC | Brown 2005[92] | Any alcohol use in last year | 9 | −0.07 (−0.20 to 0.06) |
| RHC | Brown 2005[92] | Any alcohol use in last year | 7 | −0.00 (−0.20 to 0.15) |
| RHC | Brown 2005[92] | Any alcohol use in last year | 8 | 0.10 (−0.02 to 0.23) |
| Alcohol use outcome overall | | | | −0.11 (−0.23 to 0.00) |

−0.6 −0.5 −0.4 −0.3 −0.2 −0.1 0 0.1 0.2 0.3 0.4

**FIGURE 9** Key stage 3 substance use outcomes (alcohol): effect size.

| Study | Paper | Outcome | Year | | Effect size (95% CI) |
|---|---|---|---|---|---|
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any smoking in the last month | 9 | | −0.06 (−0.19 to 0.06) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular smoking | 9 | | −0.23 (−0.43 to −0.03) |
| LIFT | Eddy 2003[101] | Ever used tobacco | 9 | | −0.06 (−0.22 to 0.11) |
| I-LST | Smith 2004[128] | Cigarettes, males | 8 | | 0.14 (−0.11 to 0.38) |
| I-LST | Smith 2004[128] | Cigarettes, males | 9 | | −0.09 (−0.34 to 0.15) |
| I-LST | Smith 2004[128] | Cigarettes, females | 8 | | −0.27 (−0.54 to −0.01) |
| I-LST | Smith 2004[128] | Cigarettes, females | 9 | | −0.32 (−0.59 to −0.06) |
| I-LST | Smith 2004[128] | Smokeless tobacco, males | 8 | | 0.09 (−0.15 to 0.34) |
| I-LST | Smith 2004[128] | Smokeless tobacco, males | 9 | | 0.04 (−0.20 to 0.29) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of cigarettes | 9 | | −0.21 (−0.43 to 0.01) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of cigarettes more than once | 9 | | 0.03 (−0.19 to −0.25) |
| RHC | Brown 2005[92] | Any cigarette use in last year | 9 | | 0.05 (−0.08 to 0.17) |
| RHC | Brown 2005[92] | Any cigarette use in last year | 8 | | −0.07 (−0.06 to 0.20) |
| Smoking outcomes overall | | | | | −0.05 (−0.12 to 0.02) |

−0.6  −0.5  −0.4  −0.3  −0.2  −0.1  0  0.1  0.2  0.3  0.4

**FIGURE 10** Key stage 3 substance use outcomes (smoking): effect size.

| Study | Paper | Outcome | Year | Effect size (95% CI) |
|---|---|---|---|---|
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any cannabis use in last 6 months | 9 | −0.01 (−0.21 to 0.18) |
| LIFT | Eddy 2003[101] | Ever used marijuana | 9 | −0.12 (−0.28 to 0.05) |
| I-LST | Smith 2003[128] | Inhalants, males | 8 | −0.16 (−0.41 to 0.08) |
| I-LST | Smith 2004[128] | Inhalants, males | 9 | 0.23 (−0.01 to 0.48) |
| I-LST | Smith 2004[128] | Inhalants, females | 8 | −0.10 (−0.36 to 0.16) |
| I-LST | Smith 2004[128] | Inhalants, females | 9 | 0.25 (−0.01 to 0.52) |
| I-LST | Smith 2004[128] | Marijuana, males | 8 | 0.00 (−0.25 to 0.25) |
| I-LST | Smith 2004[128] | Marijuana, males | 9 | 0.04 (−0.21 to 0.28) |
| I-LST | Smith 2004[128] | Marijuana, females | 8 | −0.29 (−0.56 to −0.03) |
| I-LST | Smith 2004[128] | Marijuana, females | 9 | −0.23 (−0.49 to 0.03) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of marijuana | 9 | −0.23 (−0.44 to −0.01) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of marijuana more than once | 9 | −0.17 (−0.39 to 0.05) |
| RHC | Brown 2005[92] | Any marijuana use in last year | 9 | −0.08 (−0.20 to 0.05) |
| RHC | Brown 2005[92] | Any marijuana use in last year | 8 | −0.07 (−0.20 to 0.06) |
| Illicit drug use: marijuana overall | | | | −0.10 (−0.16 to −0.04) |
| Illicit drug use: all overall | | | | −0.07 (−0.14 to −0.01) |

FIGURE 11 Key stage 3 substance use outcomes (illicit substance use): effect size.

| Study | Paper | Outcome | Year | Effect size (95% CI) |
|---|---|---|---|---|
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any drinking in last 6 months | 9 | −0.04 (−0.19 to 0.11) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Binge drinking in last month | 9 | −0.03 (−0.21 to 0.15) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular drinking | 9 | 0.05 (−0.15 to 0.24) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any cannabis use in last 6 months | 9 | −0.01 (−0.21 to 0.18) |
| The Gatehouse Project, Melbourne | Patton 2006[60] | Substance use (multicategorical) | 9 | −0.05 (−0.19 to 0.10) |
| The Gatehouse Project, Melbourne | Patton 2006[60] | Substance use (multicategorical) | 9 | −0.10 (−0.28 to 0.09) |
| The Gatehouse Project, Melbourne | Patton 2006[60] | Substance use (multicategorical) | 9 | −0.09 (−0.24 to 0.06) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any smoking in last month | 9 | −0.06 (−0.19 to 0.06) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular smoking | 9 | −0.23 (−0.43 to −0.03) |
| LIFT | Eddy 2003[101] | Patterned alcohol use | 9 | −0.22 (−0.39 to −0.06) |
| LIFT | Eddy 2003[101] | Ever used marijuana | 9 | −0.12 (−0.28 to 0.05) |
| LIFT | Eddy 2003[101] | Ever used tobacco | 9 | −0.06 (−0.22 to 0.11) |
| I-LST | Smith 2004[128] | Alcohol, males | 8 | −0.08 (−0.32 to 0.17) |
| I-LST | Smith 2004[128] | Alcohol, males | 9 | 0.03 (−0.22 to 0.27) |
| I-LST | Smith 2004[128] | Alcohol, females | 8 | −0.21 (0.47 to 0.06) |
| I-LST | Smith 2004[128] | Alcohol, females | 9 | −0.13 (−0.39 to 0.14) |
| I-LST | Smith 2004[128] | Binge drinking, males | 8 | −0.00 (−0.25 to 0.24) |
| I-LST | Smith 2004[128] | Binge drinking, males | 9 | 0.04 (−0.20 to 0.29) |
| I-LST | Smith 2004[128] | Binge drinking, females | 8 | −0.28 (−0.55 to −0.02) |
| I-LST | Smith 2004[128] | Binge drinking, females | 9 | −0.14 (−0.40 to 0.13) |
| I-LST | Smith 2004[128] | Drunkenness, males | 8 | −0.10 (−0.35 to 0.14) |
| I-LST | Smith 2004[128] | Drunkenness, males | 9 | 0.10 (−0.15 to 0.34) |
| I-LST | Smith 2004[128] | Drunkenness, females | 8 | −0.10 (−0.36 to 0.17) |
| I-LST | Smith 2004[128] | Drunkenness, females | 9 | −0.09 (−0.36 to 0.17) |
| I-LST | Smith 2004[128] | Inhalants, males | 8 | −0.16 (−0.41 to 0.08) |
| I-LST | Smith 2004[128] | Inhalants, males | 9 | 0.23 (−0.01 to 0.48) |
| I-LST | Smith 2004[128] | Inhalants, females | 8 | −0.10 (−0.36 to 0.16) |
| I-LST | Smith 2004[128] | Inhalants, females | 9 | 0.25 (−0.01 to 0.52) |
| I-LST | Smith 2004[128] | Marijuana, males | 8 | 0.00 (−0.25 to 0.25) |
| I-LST | Smith 2004[128] | Marijuana, males | 9 | 0.04 (−0.21 to 0.28) |
| I-LST | Smith 2004[128] | Marijuana, females | 8 | −0.29 (−0.56 to −0.03) |
| I-LST | Smith 2004[128] | Marijuana, females | 9 | −0.23 (−0.49 to 0.03) |
| I-LST | Smith 2004[128] | Cigarettes, males | 8 | 0.14 (−0.11 to 0.38) |
| I-LST | Smith 2004[128] | Cigarettes, males | 9 | −0.09 (−0.34 to 0.15) |
| I-LST | Smith 2004[128] | Cigarettes, females | 8 | −0.27 (−0.54 to −0.01) |
| I-LST | Smith 2004[128] | Cigarettes, females | 9 | −0.32 (−0.59 to −0.06) |
| I-LST | Smith 2004[128] | Smokeless tobacco, males | 8 | 0.09 (−0.15 to 0.34) |
| I-LST | Smith 2004[128] | Smokeless tobacco, males | 9 | 0.04 (−0.20 to 0.29) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime drunkenness | 9 | −0.29 (−0.51 to −0.07) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime drunkenness more than once | 9 | −0.22 (−0.43 to −0.00) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of alcohol | 9 | −0.35 (−0.57 to −0.14) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of alcohol more than once | 9 | −0.35 (−0.57 to −0.14) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of marijuana | 9 | −0.23 (−0.44 to −0.01) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of marijuana more than once | 9 | −0.17 (−0.39 to 0.05) |
| Positive Action Chicago | Lewis 2012[114] | Substance use count | 9 | −0.29 (−0.51 to −0.08) |
| Positive Action Chicago | Lewis 2012[114] | Substance use frequency | 9 | −0.27 (−0.48 to −0.05) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of cigarettes | 9 | −0.21 (−0.43 to 0.01) |
| Positive Action Chicago | Lewis 2012[114] | Lifetime use of cigarettes more than once | 9 | 0.03 (−0.19 to 0.25) |
| RHC | Brown 2005[92] | Any alcohol use in last year | 9 | −0.07 (−0.20 to 0.06) |
| RHC | Brown 2005[92] | Any alcohol use in last year | 7 | −0.03 (−0.20 to 0.15) |
| RHC | Brown 2005[92] | Any alcohol use in last year | 8 | 0.10 (−0.02 to 0.23) |
| RHC | Brown 2005[92] | Any marijuana use in last year | 9 | −0.08 (−0.20 to 0.05) |
| RHC | Brown 2005[92] | Any marijuana use in last year | 8 | −0.07 (−0.20 to 0.06) |
| RHC | Brown 2005[92] | Any cigarette use in last year | 9 | 0.05 (−0.08 to 0.17) |
| RHC | Brown 2005[92] | Any cigarette use in last year | 8 | 0.07 (−0.06 to 0.20) |
| All drug use outcomes overall | | | | −0.09 (−0.17 to −0.01) |

−0.6  −0.5  −0.4  −0.3  −0.2  −0.1  0  0.1  0.2  0.3  0.4

**FIGURE 12** Key stage 3 substance use outcomes (all): effect size.

## Alcohol use outcomes

Findings from the Gatehouse Project[90] did not suggest differences between intervention and control recipients at the end of the second intervention year (year 10) on probability of any drinking (OR 1.00, 95% CI 0.78 to 1.28), binge drinking (OR 0.99, 95% CI 0.70 to 1.38), or regular drinking (OR 1.05, 95% CI 0.70 to 1.57). Findings were similar at the end of the third intervention year: any drinking (OR 0.96, 95% CI 0.69 to 1.33), binge drinking (OR 1.02, 95% CI 0.71 to 1.46) and regular drinking (OR 1.13, 95% CI 0.77 to 1.66). Findings from the I-LST evaluation[132] found similar patterns after the third intervention year (year 10). Males in the intervention group were not different from their control peers on frequency of alcohol use [MD 0.28, standard error (SE) 0.22, not significant], binge drinking (MD 0.08, SE 0.18, not significant) or drunkenness (MD 0.07, SE 0.18, not significant), and a similar pattern held for female students [frequency of alcohol use (MD 0.02, SE 0.22, not significant), binge drinking (MD 0.08, SE 0.16, not significant) and drunkenness (MD 0.08, SE 0.18, not significant)].

Long-term follow-up for RHC in KS4[92] did not suggest a difference between groups in the proportion reporting alcohol use at either year 10 (corresponding to the eighth or ninth intervention year; 46% in the intervention group vs. 48% in the control group) or at year 11 (corresponding only to students in the ninth intervention year; 52% vs. 50%). However, groups had significantly different frequency of alcohol use in the ninth intervention year (year 11), with a standardised MD of −0.40 ($p < 0.05$).

## Smoking outcomes

At the end of the second intervention year (year 10), recipients of the Gatehouse Project were not significantly less likely to report any cigarette use than their control peers (OR 0.92, 95% CI 0.63 to 1.33), with similar findings at the end of the third intervention year (OR 0.91, 95% CI 0.67 to 1.24).[88–91,122] However, findings for regular smoking were more suggestive of a positive effect of the Gatehouse Project at the end of both the second (OR 0.72, 95% CI 0.47 to 1.09) and the third (OR 0.79, 95% CI 0.58 to 1.07) intervention years. I-LST did not reduce frequency of cigarette use in young men (MD 0.20, SE 0.38, not significant) or women (MD −0.18, SE 0.47, not significant) at the end of the third intervention year.[84,128,132] Finally, long-term follow-up for RHC in KS4[92] did not suggest a difference between groups in the proportion reporting smoking at either year 10 (corresponding to the eighth or ninth intervention year; 16% in the intervention group vs. 17% in the control group) or at year 11 (corresponding only to students in the ninth intervention year; 16% vs. 20%). Frequency estimates were not presented for this outcome as the difference in trajectories was not significant.

## Illicit drug use outcomes: marijuana

All three evaluations reporting illicit drug use outcomes reported on marijuana use. The evaluation of the Gatehouse Project included a diversity of estimates relating to cannabis use. Although effect estimates generally supported a reduction in cannabis use, none of the corresponding CIs excluded unity. At the end of the second intervention year,[90] recipients of the Gatehouse Project were not different in terms of the odds of cannabis use in the last 6 months (OR 1.06, 95% CI 0.75 to 1.49), although effects at the end of the third intervention year suggested decreased, albeit statistically non-significant, odds in the intervention group (OR 0.81, 95% CI 0.57 to 1.16). Findings at the end of the third intervention year[91] were similar for incidence (OR 0.81, 95% CI 0.56 to 1.15) of any cannabis use in the last 6 months, as well as for incidence (OR 0.73, 95% CI 0.43 to 1.25) and prevalence (OR 0.74, 95% CI 0.45 to 1.20) of weekly cannabis use. As for alcohol use and smoking, I-LST[132] did not reduce frequency of marijuana use in male (MD 0.04, SE 0.23, not significant) or female (MD 0.08, SE 0.20, not significant) students at the end of the third intervention year. Finally, patterns in marijuana use from the long-term follow-up for RHC in KS4[92] reflected patterns in alcohol use outcomes from this evaluation. Although findings did not suggest a difference between groups in the proportion reporting marijuana use at either year 10 (corresponding to the eighth or ninth intervention year; 25% in the intervention group vs. 28% in the control group) or at year 11 (corresponding only to students in the ninth intervention year; 30% vs. 31%), intervention recipients had a lower frequency of marijuana use in the ninth intervention year (year 11), with a standardised MD of −0.57 ($p < 0.05$).

## Meta-analyses

As specified in the protocol, we undertook separate meta-analyses for alcohol use, smoking and illicit drug use (namely, marijuana use). We also undertook an overall meta-analysis of substance use outcomes in KS4. Meta-analyses for alcohol use integrating 15 effect sizes from three evaluations did not suggest that interventions integrating health and academic education had a statistically significant impact on alcohol use ($d = -0.01$, 95% CI $-0.09$ to 0.06) (*Table 7* and *Figure 13*). This analysis had some minor statistical heterogeneity, with a between-studies $I^2$ of 2%. In contrast, meta-analyses for smoking suggested a small but statistically significant effect of these interventions in reducing smoking ($d = -0.08$, 95% CI $-0.15$ to $-0.01$) (*Figure 14*; see *Table 7*). In this analysis, which included nine effect sizes from three evaluations, between-studies heterogeneity was negligible ($I^2 = 0\%$). Interventions also appeared to have a positive effect in reducing marijuana use ($d = -0.10$, 95% CI $-0.17$ to $-0.03$) (*Figure 15*; see *Table 7*, based on 10 effect sizes from three evaluations. A pooled analysis integrating all 34 effect sizes from the three evaluations suggested a small but significant overall effect of these interventions in reducing substance use ($d = -0.06$, 95% CI $-0.09$ to $-0.02$) (*Figure 16*; see *Table 7*). Between-studies heterogeneity was negligible in this analysis as well ($I^2 = 0\%$).

### Key stage 5 substance use outcomes

At the end of the eighth year from the start of the intervention, participants in the original year 6 cohort from LIFT[98] were significantly less likely to have initiated alcohol use (OR 0.93; $p < 0.05$) or smoking (OR 0.90; $p < 0.01$) and less likely to have initiated illicit substance use at marginal non-significance (OR 0.91; $p < 0.10$).

### Trajectories of substance use

Two studies presented analyses of the trajectories of substance use behaviours: LIFT[98] and RHC.[92] The interpretation of trajectory-based analyses (also described as longitudinal or growth curve models) is different from end point-based analyses. In trajectory-based analyses, the effectiveness of the intervention is measured in terms of how different the intervention group and the control group are in change over time in the outcome of interest. Thus, in these analyses, the effectiveness of the intervention as compared with the control is estimated via a regression model, where the interaction term of intervention condition and time is the parameter of interest.

In LIFT,[98] students were followed up over 8 years to examine changes in frequency of tobacco use, alcohol use and illicit drug use. The intervention did not affect change over time in substance use behaviours in the whole population, but intervention recipients were less likely to use tobacco ($\beta = -0.10$; $p < 0.05$), alcohol ($\beta = -0.15$; $p < 0.001$) or illicit drugs ($\beta = -0.12$; $p < 0.05$). When authors tested a model including the moderating effect of sex on intervention effectiveness, females receiving the intervention had lower levels of average illicit drug use ($\beta = -0.11$; $p < 0.05$), and slower linear ($\beta = -0.14$; $p < 0.05$) and quadratic ($\beta = -0.11$; $p < 0.05$) growth over time. Females also had slower linear ($\beta = 09.14$; $p < 0.001$) and quadratic ($\beta = -0.13$; $p < 0.05$) growth over time in tobacco use.

In RHC,[92] change in substance use over time was examined over 4 years. Authors estimated frequency of use and prevalence of use jointly in their model, and examined a diversity of model fits. The intervention did not have a significant effect on change in the prevalence of alcohol use over time (years 7 to 9: $\beta = -0.005$, SE 0.198;

**TABLE 7** Summary of KS4 substance use outcomes

| Outcome | $k$ ($n$) | Standardised MD (95% CI) | $I^2$, study level (%) |
| --- | --- | --- | --- |
| Alcohol | 3 (15) | $-0.01$ ($-0.09$ to $-0.06$) | 2 |
| Smoking | 3 (9) | $-0.08$ ($-0.15$ to $-0.01$) | 0 |
| Illicit drug use: marijuana | 3 (10) | $-0.10$ ($-0.17$ to $-0.03$) | 0 |
| All drug use outcomes | 3 (34) | $-0.06$ ($-0.09$ to $-0.02$) | 0 |

| Study | Paper | Outcome | Year | Effect size (95% CI) |
|---|---|---|---|---|
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any drinking in last 6 months | 10 | 0.00 (–0.14 to 0.14) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Binge drinking in last month | 10 | –0.01 (–0.19 to 0.18) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular drinking | 10 | 0.03 (–0.20 to 0.25) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any drinking in last 6 months | 11 | –0.02 (–0.20 to 0.16) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Binge drinking in last month | 11 | 0.01 (–0.19 to 0.21) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular drinking | 11 | 0.07 (–0.14 to 0.28) |
| I-LST | Vicary 2006[132] | Alcohol, males | 10 | 0.43 (–0.23 to 1.09) |
| I-LST | Vicary 2006[132] | Alcohol, females | 10 | 0.03 (–0.63 to 0.69) |
| I-LST | Vicary 2006[132] | Binge drinking, males | 10 | 0.21 (–0.70 to 1.11) |
| I-LST | Vicary 2006[132] | Binge drinking, females | 10 | 0.21 (–0.60 to 1.01) |
| I-LST | Vicary 2006[132] | Drunkenness, males | 10 | 0.17 (–0.71 to 1.06) |
| I-LST | Vicary 2006[132] | Drunkenness, females | 10 | 0.20 (–0.68 to 1.08) |
| RHC | Brown 2005[92] | Any alcohol use in last year | 10 | –0.04 (–0.17 to 0.08) |
| RHC | Brown 2005[92] | Frequency of alcohol use | 11 | –0.40 (–0.66 to –0.14) |
| RHC | Brown 2005[92] | Any alcohol use in last year | 11 | 0.04 (–0.14 to 0.23) |
| Alcohol use outcomes overall | | | | –0.01 (–0.09 to 0.06) |

FIGURE 13 Key stage 4 substance use outcomes (alcohol): effect size.

| Study | Paper | Outcome | Year | | Effect size (95% CI) |
|---|---|---|---|---|---|
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any smoking in the last month | 10 | | −0.05 (−0.25 to 0.16) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular smoking | 10 | | −0.18 (0.41 to 0.05) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular smoking | 11 | | −0.13 (−0.30 to 0.04) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any smoking in the last month | 11 | | −0.05 (−0.22 to 0.12) |
| I-LST | Vicary 2006[132] | Cigarettes, males | 10 | | 0.14 (−0.38 to 0.66) |
| I-LST | Vicary 2006[132] | Cigarettes, females | 10 | | −0.13 (−0.77 to 0.52) |
| RHC | Brown 2005[92] | Any cigarette use in last year | 10 | | −0.04 (−0.17 to 0.09) |
| RHC | Brown 2005[92] | Any cigarette use in last year | 11 | | −0.15 (−0.34 to 0.04) |
| RHC | Brown 2005[92] | Frequency of cigarette use | 11 | | 0.00 (−0.44 to 0.44) |
| Smoking outcomes overall | | | | | −0.08 (−0.15 to −0.01) |

**FIGURE 14** Key stage 4 substance use outcomes (smoking): effect size.

| Study | Paper | Outcome | Year | Effect size (95% CI) |
|---|---|---|---|---|
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any cannabis use in last 6 months | 10 | 0.03 (–0.16 to 0.22) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any cannabis use in last 6 months | 11 | –0.12 (–0.31 to 0.08) |
| The Gatehouse Project, Melbourne | Bond 2004[91] | Weekly cannabis use prevalence | 11 | –0.17 (–0.44 to 0.10) |
| The Gatehouse Project, Melbourne | Bond 2004[91] | Weekly cannabis use incidence | 11 | –0.17 (–0.47 to 0.12) |
| The Gatehouse Project, Melbourne | Bond 2004[91] | Any cannabis use incidence | 11 | –0.12 (–0.31 to 0.08) |
| I-LST | Vicary 2006[132] | Marijuana, males | 10 | 0.09 (–0.89 to 1.07) |
| I-LST | Vicary 2006[132] | Marijuana, females | 10 | 0.17 (–0.68 to 1.03) |
| RHC | Brown 2005[92] | Any marijuana use in last year | 10 | –0.08 (–0.21 to 0.04) |
| RHC | Brown 2005[92] | Frequency of marijuana use | 11 | –0.57 (–0.91 to –0.23) |
| RHC | Brown 2005[92] | Any marijuana use in last year | 11 | –0.03 (–0.21 to 0.16) |
| Illicit drug use: marijuana overall | | | | –0.10 (–0.17 to –0.03) |



**FIGURE 15** Key stage 4 substance use outcomes (illicit drug use): effect size.

| Study | Paper | Outcome | Year | Effect size (95% CI) |
|---|---|---|---|---|
| The Gatehouse Project, Melbourne | Bond 2004[90] | Binge drinking in last month | 10 | −0.01 (−0.19 to 0.18) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any drinking in last 6 months | 10 | 0.00 (−0.14 to 0.14) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular drinking | 10 | 0.03 (−0.20 to 0.25) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular drinking | 11 | 0.07 (−0.14 to 0.28) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Binge drinking in last month | 11 | 0.01 (−0.19 to 021) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any drinking in last 6 months | 11 | −0.02 (−0.20 to 0.16) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any cannabis use in last 6 months | 10 | 0.03 (−0.16 to 0.22) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any cannabis use in last 6 months | 11 | −0.12 (−0.31 to 0.08) |
| The Gatehouse Project, Melbourne | Bond 2004[91] | Weekly cannabis use prevalence | 11 | −0.17 (−0.44 to 0.10) |
| The Gatehouse Project, Melbourne | Bond 2004[91] | Weekly cannabis use incidence | 11 | −0.17 (−0.47 to 0.12) |
| The Gatehouse Project, Melbourne | Bond 2004[91] | Any cannabis use incidence | 11 | −0.12 (−0.31 to 0.08) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any smoking in the last month | 10 | −0.05 (−0.25 to 0.16) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular smoking | 10 | −0.18 (−0.41 to 0.05) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Regular smoking | 11 | −0.13 (−0.30 to 0.04) |
| The Gatehouse Project, Melbourne | Bond 2004[90] | Any smoking in the last month | 11 | −0.05 (−0.22 to 0.12) |
| I-LST | Vicary 2006[132] | Alcohol, females | 10 | 0.03 (−0.63 to 0.69) |
| I-LST | Vicary 2006[132] | Drunkenness, males | 10 | 0.17 (−0.71 to 1.06) |
| I-LST | Vicary 2006[132] | Drunkenness, females | 10 | 0.20 (−0.68 to 1.08) |
| I-LST | Vicary 2006[132] | Alcohol, males | 10 | 0.43 (−0.23 to 1.09) |
| I-LST | Vicary 2006[132] | Binge drinking, males | 10 | 0.21 (−0.70 to 1.11) |
| I-LST | Vicary 2006[132] | Binge drinking, females | 10 | 0.21 (−0.60 to 1.01) |
| I-LST | Vicary 2006[132] | Marijuana, males | 10 | 0.09 (−0.99 to 1.07) |
| I-LST | Vicary 2006[132] | Marijuana, females | 10 | 0.17 (−0.68 to 1.03) |
| I-LST | Vicary 2006[132] | Cigarettes, males | 10 | 0.14 (−0.38 to 0.66) |
| I-LST | Vicary 2006[132] | Cigarettes, females | 10 | −0.13 (−0.77 to 0.52) |
| RHC | Brown 2005[92] | Any alcohol use in last year | 10 | −0.04 (−0.17 to 0.08) |
| RHC | Brown 2005[92] | Any alcohol use in last year | 11 | 0.04 (−0.14 to 0.23) |
| RHC | Brown 2005[92] | Frequency of alcohol use | 11 | −0.40 (−0.66 to −0.14) |
| RHC | Brown 2005[92] | Any marijuana use in last year | 10 | −0.08 (−0.21 to 0.04) |
| RHC | Brown 2005[92] | Frequency of marijuana use | 11 | −0.57 (−0.91 to −0.23) |
| RHC | Brown 2005[92] | Any marijuana use in last year | 11 | −0.03 (−0.21 to 0.16) |
| RHC | Brown 2005[92] | Any cigarette use in last year | 10 | −0.04 (−0.17 to 0.09) |
| RHC | Brown 2005[92] | Any cigarette use in last year | 11 | −0.15 (−0.34 to 0.04) |
| RHC | Brown 2005[92] | Frequency of cigarette use | 11 | 0.00 (−0.44 to 0.44) |
| All drug use outcomes overall | | | | −0.08 (−0.09 to −0.02) |

−1.0  −0.8  −0.6  −0.4  −0.2  0  0.2  0.4  0.6  0.8  1.0

FIGURE 16 Key stage 4 substance use outcomes (all): effect size.

years 9 to 11: $\beta = 0.047$; $p = 0.190$). In contrast, although the intervention did not result in a significant reduction in frequency of use over time in school years 7 to 9 ($\beta = -0.29$, SE 0.095), the intervention did result in a significant decrease in change in frequency of use over time in school years 9 to 11 ($\beta = -0.199$, SE 0.096; $p < 0.05$). That is, after both groups increased frequency of alcohol use between school years 7 to 9, the intervention group reduced frequency of alcohol use faster than the control group in school years 9 to 11. Similarly, although the intervention did not significantly impact change over time in prevalence of marijuana use (linear: $\beta = 0.055$; $p = 0.104$; quadratic: $\beta = -0.008$, SE 0.143), students receiving the intervention had a greater decrease over time in frequency of use of marijuana ($\beta = -0.223$, SE 0.052; $p < 0.001$). That is, intervention students reduced frequency of marijuana use more quickly than control students. The intervention group was not associated with linear or quadratic changes in prevalence (linear: $\beta = -0.153$, SE 0.105; quadratic: $\beta = -0.123$, SE 0.155) or frequency (linear: $\beta = -0.008$, SE 0.042; quadratic: $\beta = -0.033$, SE 0.092) of cigarette use over time.

## Effects of interventions on violence

Eleven intervention evaluations reported outcomes relating to violence perpetration (*Tables 8* and *9*). These outcomes were reported across 15 papers. Ten of the evaluations, including 12 papers, reported outcomes in KS2. Three evaluations, Positive Action Chicago,[115] Second Step[139–141] and Youth Matters,[109] reported outcomes in KS3.

### *Key stage 2 violence perpetration outcomes*
Findings for violence perpetration in KS2 were inconsistent across evaluations. Unlike findings for the year 2 cohort in LIFT,[125] effects of the intervention in the year 6 cohort at the end of the first intervention year on observed physical aggression in the playground were constant over values of the pre-intervention score [$d = -0.14$ at mean, 1 standard deviation (SD) and 2 SDs above the pre-intervention mean]; these findings were also described as 'statistically significant'.

After the first intervention year (year 4) of 4Rs,[52] intervention participants were not different from control participants on teacher-reported aggression ($b = 0.02$, SE 0.05, based on a 1–4 scale). In contrast, after the second intervention year,[2] students who received 4Rs had lower levels of teacher-reported aggression than their control peers ($d = -0.21$; $p < 0.05$).

The Bullying Literature Project, implemented in year 4 and year 5 classrooms, did not generate a change in physical aggression by either teacher report on individual students (intervention group: mean 1.12, SD 0.47, $n = 95$; control group: mean 1.19, SD 0.47, $n = 55$; $p = 0.67$) or student report (intervention group: mean 1.20, SE 0.44, $n = 90$; control group: mean 1.14, SD 0.36, $n = 42$; $p = 0.84$) when measured at 1 week post intervention.[96] This finding was the same in the Bullying Literature Project–Moral Disengagement version ($F_{1,80} = 0.83$; $p = .431$), implemented in year 4 classrooms only, although only combined student-reported physical and emotional bullying estimates were available.[136]

In PATHS, which was implemented starting in year 4 classrooms,[97] student-reported and teacher-reported aggression were measured using multilevel models incorporating all waves of measurement and effect sizes estimated corresponding to each wave of measurement; thus, standard errors for effect sizes are not provided directly. Small positive effects of the intervention on student-reported aggression at the end of the first intervention year ($d = -0.048$, 95% CI $-0.189$ to 0.092) and at the start ($d = -0.064$, 95% CI $-0.205$ to 0.076) and end ($d = -0.048$, 95% CI $-0.188$ to 0.093) of the second intervention year gave way to a small deleterious effect of the intervention at the end of the third intervention year ($d = 0.082$, 95% CI $-0.060$ to 0.224). The intervention had an opposite effect on teacher-reported aggression, with initially small but negative intervention effects at the end of the first intervention year ($d = 0.036$, 95% CI $-0.105$ to 0.178) and at the start of the second intervention year ($d = 0.035$, 95% CI $-0.107$ to 0.178), suggesting progressively greater effects of the intervention at the end of the second ($d = -0.005$, 95% CI $-0.146$ to 0.136) and the third ($d = -0.199$, 95% CI $-0.338$ to $-0.060$) intervention years.

**TABLE 8** Measures used in studies reporting violence perpetration outcomes

| Evaluation | Measure | Notes |
|---|---|---|
| 4Rs | Aggression | Frequency score on 13 aggressive behaviours assessed by teacher report in the last month, including physical aggression and threatening of others |
| Bullying Literature Project | Physical bullying | Assessed by teacher and student report; mean of frequency scores relating to reports of violence |
| Bullying Literature Project–Moral Disengagement | Bullying | Assessed by student report; mean of frequency scores relating to physical and emotional bullying |
| LIFT | Change in child physical playground aggression | Measured by observation, including physical bullying observed by children |
| PATHS | Aggression | Assessed by teacher and student report; mean of frequency scores relating to verbal and physical aggression |
| Positive Action Chicago | Bullying | Student report: mean of frequency scores relating to verbal or physical aggression behaviours in the past 2 weeks |
| | | Parent report: count of observed verbal or physical aggression behaviours in the past 30 days |
| | Violence-related behaviours | Count of lifetime behaviours: carried a knife, threatened to cut or stab someone, cut or stabbed someone on purpose, has been asked to join a gang, 'hung out' with gang members, has been a member of a gang |
| Positive Action Hawaii | Cut or stabbed others | Student report; lifetime prevalence |
| | Shot another person | Student report; lifetime prevalence |
| | Physically hurts other people | Teacher report |
| | Gets into a lot of fights | Teacher report |
| Second Step | Physical aggression perpetration | Student report; any fighting behaviours in the last 30 days |
| | Sexual harassment and violence perpetration | Student report; any verbal sexual violence, groping behaviours or forced sexual contact |
| Steps to Respect I | Bullying | Playground observation of students |
| | Direct aggression | Mean of student-reported frequency scores of direct bullying |
| Steps to Respect II | Bullying perpetration | Measured by student report; proportion of students with at least one bullying behaviour |
| | Physical bullying perpetration | Measured by teacher report; proportion of students with at least one physical bullying behaviour |
| Youth Matters | Bullying | At least two or three times per month on at least one bullying behaviour |
| | Bully, victim or bully–victim | Classification of students into one of three categories based on questionnaire responses |

At the end of the third intervention year (year 6) in Positive Action Chicago,[116] students reported both lower counts of bullying behaviours (IRR 0.59, 95% CI 0.37 to 0.92) and of serious violence-related behaviours, including 'cut[ting] or stab[bing] someone on purpose' (IRR 0.63, 95% CI 0.45 to 0.88). Findings from Positive Action Hawaii[87] were similar for both counts of student-reported violent behaviours (IRR 0.42, 90% CI 0.24 to 0.73) and teacher-reported violent behaviours (IRR 0.54, 90% CI 0.30 to 0.77). For students in year 6 (the fourth or fifth intervention year), intervention recipients were less likely to report cutting or stabbing someone (OR 0.29, 90% CI 0.16 to 0.52) or shooting someone (OR 0.24, 90% CI 0.14 to 0.40). Teachers were less likely to report that students hurt others (OR 0.61, 90% CI 0.38 to 0.97) or got into a lot of fights (OR 0.63, 90% CI 0.47 to 0.84).

**TABLE 9** Analysis schematic for violence perpetration

| Intervention | Cohort | Year | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 4Rs | | | | | | | | | | | | | |
| Bullying Literature Project | | | | | | | | | | | | | |
| LIFT | First grade | | | | | | | | | | | | |
| | Fifth grade | | | | | | | | | | | | |
| PATHS | | | | | | | | | | | | | |
| Positive Action Chicago | | | | | | | | | | | | | |
| Positive Action Hawaii | First grade | | | | | | | | | | | | |
| | Second grade | | | | | | | | | | | | |

**TABLE 9** Analysis schematic for violence perpetration (*continued*)



Green cells indicate intervention delivery. Diagonally striped cells indicate an end-point measurement. Dotted cells indicate a longitudinal, trajectory-based analysis.

In the first evaluation of Steps to Respect,[104] the intervention, implemented among students ranging from year 4 to year 7, was associated with decreased bullying based on playground observation ($F_{91.3} = 5.02$; $p < 0.01$) at the end of the first intervention year, but not direct aggression based on student report ($F_{68.7} = 2.05$; $p > 0.05$). The second evaluation of Steps to Respect,[93] implemented among students ranging from years 4 to 6, revealed a similar pattern. Although teacher reporting of physical bullying perpetration was less in intervention schools than in control schools at the end of the first intervention year (OR 0.61, $t_{29} = -3.12$; $p < 0.01$), student reports did not suggest a difference between schools on bullying perpetration ($t_{29} = -1.06$).

Finally, at the end of the second intervention year (year 6) of Youth Matters,[110] students in intervention schools were not less likely to report violence perpetration (OR 0.85, 95% CI 0.29 to 1.47; $p = 0.585$). Evaluators of Youth Matters also explored the use of a latent class analysis method to classify intervention recipients as either victims, bullies or bully–victims. Proportions of intervention and control recipients classified as bullies or bully–victims were not significantly different at the end of the first (intervention group: 21%, $n = 356$; control group: 22%, $n = 392$) or second (intervention group: 19%, $n = 244$; control group: 23%, $n = 293$) intervention years.[109]

### Meta-analyses

Because of the heterogeneity of constructs, the diversity of analysis methods and the lack of available data to estimate consistent effect sizes, we did not undertake a meta-analysis for violence perpetration outcomes in KS2.

### *Key stage 3 violence perpetration outcomes*

The three evaluations examining violence perpetration outcomes in KS3 had dissimilar results. At the end of the sixth intervention year (year 9) in Positive Action Chicago,[115] students receiving the intervention reported a lower count of violence-related behaviours than their peers receiving the control intervention (IRR 0.38, 95% CI 0.18 to 0.81; equivalent to $d = -0.54$). Students also reported fewer bullying behaviours ($d = -0.39$), and parents reported that their children engaged in fewer bullying behaviours ($d = -0.31$). Significance values for these estimates were not presented, but both were supported by significant condition-by-time interactions in multilevel models, indicating that the intervention group showed an improved trajectory over time compared with the control group. In contrast, after the third intervention year (year 7) in Youth Matters,[109] proportions of students were not different in the collective bully and bully–victim groups (both groups, 16%; intervention group, $n = 283$; control group, $n = 289$). Finally, findings for Second Step were reported at the end of the first, second and third years of intervention. At the end of the first school year (year 7), students in intervention schools had decreased odds of physical aggression (OR 0.70; $p < 0.05$) but not sexual harassment and sexual violence perpetration (OR 1.04; $p > 0.05$).[139] These findings did not hold to the end of the second school year for physical aggression (OR 0.80, 95% CI 0.59 to 1.08); however, sexual harassment and sexual violence perpetration was significantly reduced in intervention schools in Illinois (OR 0.72, 95% CI 0.54 to 0.95) but not Kansas (OR 0.99, 95% CI 0.71 to 1.48).[140,141] At the end of the third school year, there were no direct effects of Second Step on sexual harassment perpetration ($\beta = 0.005$, SE 0.012); findings for physical aggression were not available.

### *Trajectories of violence perpetration*

Five evaluations presented analyses of trajectories of violence perpetration behaviours: 4Rs,[3] PATHS,[97] Positive Action Chicago,[115] Steps to Respect[103] and Youth Matters.[110,111] Evaluations were inconsistent in their findings, and comparison of findings was made challenging by the different time operators used in analyses.

In 4Rs,[3] teacher-reported aggression was examined in four waves of data collection over 2 years. The intervention was associated with a decrease in the rate of change in aggression over time ($\beta = -0.03$, SE 0.01; $p < 0.05$). This difference was standardised as a difference in growth parameters of 0.05 SDs, indicating a small intervention effect. In PATHS,[97] data on teacher-reported aggression were collected up to five times over 3 years. Neither the linear interaction between intervention and time ($\beta = 0.400$, 90% CI

–0.246 to 1.047) nor the quadratic interaction between intervention and time (β = –0.137, 90% CI –0.263 to 0.011) suggested a significant intervention effect in change of aggression over time. However, the general pattern was an increase in aggression in both groups followed by a faster rate of decline in the intervention group starting during year 5. Similarly, findings for student-reported aggression did not reveal a significantly different linear (β = –0.195, 90% CI –0.517 to 0.127) or quadratic (β = 0.049, 90% CI –0.013 to 0.111) difference between intervention and control groups in change over time.

In Positive Action Chicago,[115] data on bullying were collected eight times by student reports and six times by parent reports over the duration of the trial. Both student and parent reports suggested that the intervention slowed the rate of increase in bullying behaviours. In models examining student reports, the linear interaction between intervention and time suggested a slower rate in the increase of count of bullying behaviours reported (IRR 0.85, 95% CI 0.76 to 0.95), although this was slightly offset by a quadratic interaction term (IRR 1.02, 95% CI 1.00 to 1.04). In models examining parent report, the linear interaction between intervention and time suggested a similarly slower rate in increase of bullying behaviours (IRR 0.92, 95% CI 0.89 to 0.97).

In the first evaluation of Steps to Respect, bullying behaviours were assessed by observation and by student report four times over 2 years.[103] In models analysing observation data, the interaction between intervention assignment and grade was statistically significant ($F_{1,202}$ = 12.86; $p < 0.01$). This translated into an increase per year in terms of bullying occasions per hour of 0.035 ($p < 0.01$) in the control group compared with a non-significant increase of 0.04 in the intervention group. However, student reports of direct aggression did not suggest a difference in rate of change over time ($F_{1,1670} < 1$), with both groups increasing in average bullying behaviours reported over time (intervention group: 0.10/year; $p < 0.01$; control group: 0.13/year; $p < 0.01$).

In Youth Matters,[109–111] bullying perpetration was assessed four times over 2 years. The log odds of bully status did not change to a significantly different degree in students receiving Youth Matters compared with those in the control group (β = 0.079, SE 0.090; $p = 0.378$).[110] This result was confirmed after 3 years of data collection.[111] Follow-up models, including latent transition analyses, suggested that more intervention students transitioned from the bully latent class to the uninvolved latent class between baseline and the end of the first intervention year (54% of bullies in the intervention group transitioned to uninvolved vs. 37% of bullies in the control group; $p = 0.009$) and between the end of the second intervention year and the end of the third intervention year (44% of bullies in the intervention group transitioned to uninvolved vs. 34% of bullies in the control group; $p = 0.058$).[109]

### Violence victimisation

Nine intervention evaluations reported outcomes relating to violence victimisation. These outcomes were reported across 13 papers. Seven intervention evaluations included violence victimisation outcomes measured in KS2, three intervention evaluations included violence victimisation outcomes measured in KS3 and one intervention evaluation included outcomes measured in KS4 (*Tables 10* and *11*).

### Key stage 2 violence victimisation outcomes

Overall, although evaluations were similar in the follow-up period, they did not have a clear effect on the basis of reports alone on violence victimisation as measured in KS2. Intervention evaluations reporting victimisation outcomes in this period were the Bullying Literature Project,[96] Bullying Literature Project–Moral Disengagement[136], Learning to Read in a Healing Classroom[137], PATHS,[97] both the first[104] and second evaluations of Steps to Respect,[93] and Youth Matters.[109,110]

The Bullying Literature Project, implemented in year-4 and year-5 classrooms, did not generate a change in physical victimisation by either teacher report on individual students (intervention group: mean 1.04, SD 0.23, $n = 95$; control group: mean 1.04, SD 0.21, $n = 55$; $p = 0.39$) or student report (intervention group: mean 1.35, SD 0.54, $n = 90$; control group: mean 1.43, SD 0.66, $n = 42$; $p = 0.57$) when measured at 1 week post intervention.[96] However, students receiving the Bullying Literature Project–Moral Disengagement[136] version did report a decrease in victimisation (both physical and emotional combined) after the intervention

**TABLE 10** Measures used in studies reporting violence victimisation outcomes

| Evaluation | Measure | Notes |
|---|---|---|
| Bullying Literature Project | Physical bullying | Assessed by teacher and student report; mean of frequency scores relating to reports of violence |
| Bullying Literature Project–Moral Disengagement | Bullying victimisation | Assessed by student report; mean of frequency scores relating to physical and emotional bullying |
| The Gatehouse Project | Bullying victimisation | Assessed by student report; any of the following with regard to the student in question: being teased, having rumours spread about them, deliberate exclusion or experience of threats or violence |
| Learning to Read in a Healing Classroom | Victimisation | Assessed by student report; average of frequency scores of peer verbal and physical bullying |
| PATHS | Victimisation | Assessed by student report; sum of frequency scores of victimisation in the last 2 weeks |
| Second Step | Peer victimisation | Student report; any physical or verbal victimisation in last 30 days |
| | Sexual harassment and violence perpetration | Student report; any victimisation by verbal sexual violence or groping behaviours or forced sexual contact |
| Steps to Respect I | Target of bullying | Playground observation of students |
| | Victimisation | Assessed by student report; mean of frequency scores for physical and verbal victimisation items |
| Steps to Respect II | Victimisation | Assessed by student report; mean of frequency scores for physical and verbal victimisation items |
| Youth Matters | Victimisation | Assessed by student report; mean of frequency scores for physical and verbal victimisation items, and also at least two or three times per month victimisation, at least one bullying behaviour |
| | Bully, victim or bully–victim | Classification of students based on questionnaire responses into one of three categories |

(intervention group: mean 1.76, SD 0.81 to mean 1.60, SD 0.66, $n = 42$; control group: mean 1.23, SD 0.38 to mean 1.38, SD 0.53, $n = 42$), with a significant time-by-treatment interaction in an analysis of variance (ANOVA) ($F_{1,80} = 7.42$; $p = 0.047$).

Learning to Read in a Healing Classroom[137] examined relational and physical victimisation after 1 year of intervention implementation and found no significant effect of the intervention (weighted $d = -0.01$, SE 0.06). This intervention was implemented in years 3 to 5.

Promoting Alternative Thinking Strategies, which was implemented starting in year 4 classrooms, measured student-reported victimisation using standardised MDs and found small but non-significant increases at the end of the first intervention year ($d = 0.044$, 95% CI –0.098 to 0.185), at the start ($d = 0.074$, 95% CI –0.067 to 0.216) and end ($d = 0.092$, 95% CI –0.050 to 0.234) of the second year and at the end of the third year ($d = 0.089$, 95% CI –0.053 to 0.231) following the start of intervention implementation.[97] As for violence perpetration outcomes, standard errors were obtained via correspondence with authors.

Steps to Respect, which was evaluated and implemented across the range of years in the equivalent of KS2 in two different trials, did not find differences in student-reported bullying victimisation at the end of the first intervention year in either the first trial (intervention group: mean 0.80, SD 1.51; control group: mean 0.86, SD 1.44; $F < 1$)[104] or the second trial (intervention group: mean 2.11, SD 1.03; control group: mean 2.18, SD 1.06; $t_{29} = -1.15$).[93] Frey *et al.*[104] also undertook playground observation at the end of the first intervention year, which was suggestive of a marginally non-significant decrease in bullying victimisation (intervention group: mean 0.9, SD 0.82; control group: mean 1.01, SD 0.83; $F_{72.4} = 3.74$; $p < 0.10$).

**TABLE 11** Analysis schematic for violence victimisation

| Intervention | Cohort | Year | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| Steps to Respect II | Third grade | | | | | | | | | | | | |
| | Fourth grade | | | | | | | | | | | | |
| | Fifth grade | | | | | | | | | | | | |
| Youth Matters | | | | | | | | | | | | | |

Green cells indicate intervention delivery. Diagonally striped cells indicate an end-point measurement. Dotted cells indicate a longitudinal, trajectory-based analysis.
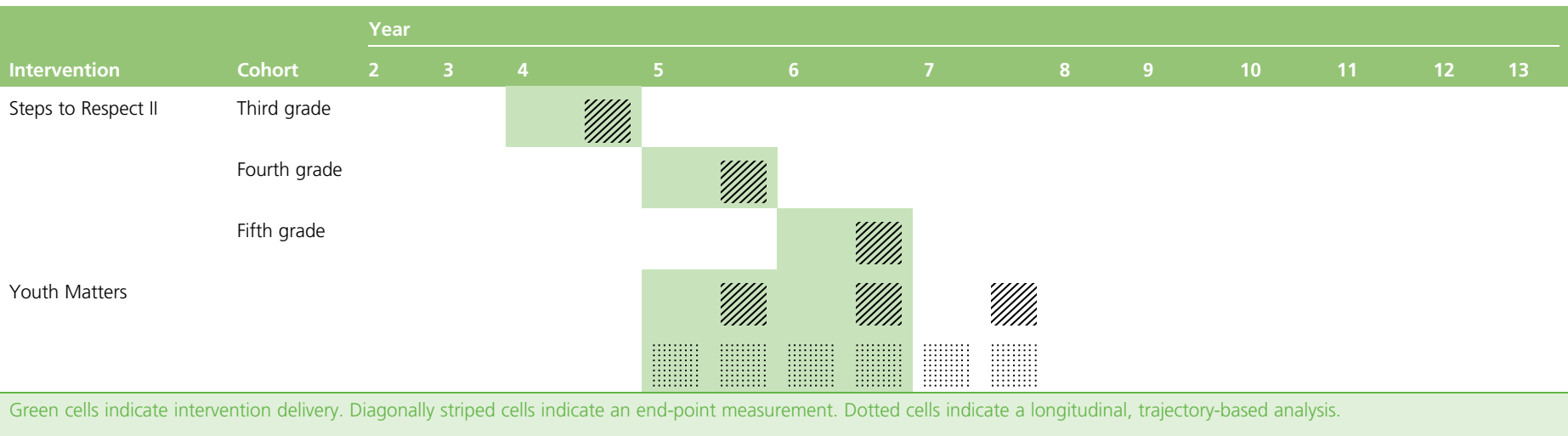
Finally, Youth Matters, which was implemented starting in Year 5, examined bullying victimisation through both continuous and dichotomous measures. Because of non-normality of continuous data, scores were log transformed. At the end of the second intervention year, the difference in log-transformed scores was suggestive of a decrease in bullying victimisation measured continuously (difference –0.171, SE 0.083; $p = 0.049$).[110] At the end of the second year, intervention recipients were less likely to have reported bullying victimisation as measured dichotomously, but this difference was not significant (OR 0.61; $p = 0.098$). Based on the latent class analysis, proportions of intervention and control recipients classified as victims or bully–victims were not significantly different at the end of the first (intervention group: 47%, $n = 356$; control group: 46%, $n = 392$) or second (intervention group: 45%, $n = 244$; control group: 47%, $n = 293$) intervention years.[109]

## Meta-analyses

As stated in our protocol, we undertook a meta-analysis for violence victimisation in KS2. The multilevel meta-analysis included 12 effect sizes from seven evaluations. Interventions integrating health and academic education did not have a statistically significant impact in reducing violence victimisation, with a standardised MD of –0.02 (95% CI –0.10 to 0.05) (*Figure 17*). Statistical heterogeneity was not large in this meta-analysis, with a between-studies $I^2$-value of 1.7%.

### *Key stages 3 and 4 violence victimisation outcomes*

Intervention evaluations presenting violence victimisation outcomes in the equivalent of KS3 and KS4 did not suggest evidence of effectiveness. Outcomes in KS3 were presented by Second Step,[139–141] Youth Matters[109,111] and the Gatehouse Project.[90] The Gatehouse Project also presented outcomes for KS4.

In Second Step, neither peer victimisation (OR 1.01; $p > 0.5$) nor sexual harassment and violence victimisation (OR 1.01; $p > 0.05$) were different between students in intervention schools and control schools after the first intervention year;[139] this remained the case at the end of the second intervention year (peer victimisation: OR 0.94, 95% CI 0.75 to 1.18; sexual victimisation: 0.91 95% CI 0.72 to 1.15).[141]

In Youth Matters, differences in the log-transformed scores for bullying victimisation were not suggestive of a decrease in victimisation in intervention recipients compared with control recipients, but this difference was no longer significant (difference –0.123, SE 0.068; $p = 0.08$).[111] However, at the end of the third intervention year (equivalent to year 7), fewer students in the intervention group were members of the victim or bully–victim classes than in the control group (intervention group: 36%, $n = 283$; control group: 45%, $n = 289$).[109] Based on our own chi-squared test, this difference was significant ($p = 0.029$).

The Gatehouse Project, which was implemented starting in year 9, did not find evidence of a change in bullying victimisation at the end of the first intervention year (OR 1.03, 95% CI 0.86 to 1.26). Evaluators also did not find a significant difference in the prevalence of bullying victimisation at the end of the second (OR 1.03, 95% CI 0.78 to 1.34) or third (OR 0.88, 95% CI 0.68 to 1.13) intervention years, which corresponded to the first 2 years of KS4.

### *Trajectories of violence victimisation*

Three evaluations presented analyses of trajectories of violence victimisation over time: PATHS,[97] Steps to Respect [103] and Youth Matters.[110,111] None of the evaluations consistently suggested that interventions were associated with decreases in rates of change in bullying victimisation over time. In PATHS,[97] data on student-reported victimisation were collected up to five times over 3 years. Neither the linear interaction between intervention and time ($\beta = 0.245$, 90% CI –0.230 to 0.720) nor the quadratic interaction between intervention and time ($\beta = -0.032$, 90% CI –0.406 to 0.342) suggested a significant intervention effect in change in experiences of bullying over time. In the first evaluation of Steps to Respect,[103] bullying victimisation was assessed four times over 2 years. Both groups reported non-significant decreases over time in the average number of bullying behaviours experienced (intervention group: –0.08/year; control group: –0.07/year), but these slopes were not significantly different over time ($F_{1,1672} < 1$). Finally, in Youth Matters, bullying victimisation was assessed as both a continuous and a binary variable. Binary outcome

| Study | Paper | Outcome | Year | Effect size (95% CI) |
|---|---|---|---|---|
| Bullying Literature Project | Couch 2015[96] | Physical victimisation (teacher report) | 4 to 5 | 0.00 (−6.71 to 6.71) |
| Bullying Literature Project | Couch 2015[96] | Physical victimisation (student report) | 4 to 5 | −0.14 (−6.02 to 5.75) |
| PATHS | Crean 2013[97] | Victimisation (student report) | 4 | 0.04 (−0.10 to 0.19) |
| PATHS | Crean 2013[97] | Victimisation (student report) | 5 | 0.09 (−0.05 to 0.23) |
| PATHS | Crean 2013[97] | Victimisation (student report) | 6 | 0.09 (−0.05 to 0.23) |
| PATHS | Crean 2013[97] | Victimisation (student report) | 5 | 0.07 (−0.07 to 0.22) |
| Steps to Respect | Frey 2005[104] | Victimisation (student report) | 4 to 7 | −0.45 (−0.91 to 0.01) |
| Steps to Respect | Frey 2005[104] | Victimisation (observation) | 4 to 7 | −0.04 (−0.15 to 0.07) |
| Steps to Respect II | Brown 2011[93] | Bullying victimisation (student report) | 4 to 6 | −0.07 (−0.18 to 0.05) |
| Youth Matters | Jenson 2007[110] | Victimisation (student report) | 6 | −0.27 (−0.58 to 0.04) |
| Violence victimisation outcomes overall | | | | −0.04 (−0.14 to 0.07) |

**FIGURE 17** Violence victimisation outcomes: effect size.

analyses after 2 years suggested a faster decrease in the log odds of bully victim status over time, but this did not approach significance ($\beta = -0.146$, SE 0.078; $p = 0.061$).[110] Continuous outcome analyses after 3 years of follow-up[111] did not suggest a significant difference between groups in the change in the degree of bullying victimisation over time (intervention by time: $\beta = -0.017$, SE 0.014; $p = 0.236$). Follow-up analyses using latent transition analysis suggested that more intervention students than control students transitioned from bullied to uninvolved between each measurement occasion [between baseline and end of first intervention year: 26% in intervention group vs. 19% in control group ($p = 0.04$); end of first intervention year to end of second intervention year: 36% in intervention group vs. 20% in control group ($p = 0.01$); end of second intervention year to end of third intervention year: 45% in intervention group vs. 38% in control group ($p = 0.007$)].

## Effects of interventions on academic attainment

Five included evaluations reported academic attainment outcomes.[2,51–53,83,87,92,93,95,113–116,129,130] Because many of these outcomes were reported at the school-level based on standardised test achievement, these outcomes are reported separately. We elected not to undertake meta-analysis on academic attainment given the diversity in outcome measurement, intervention type, KS and follow-up period.

### Key stages 1 and 2 academic attainment outcomes

Three evaluations examined academic attainment in KS1 and KS2: RHC,[95] 4Rs[3,52] and the second evaluation of Steps to Respect.[93] Findings were not collectively suggestive of effectiveness, although meta-analysis was judged inappropriate. Although Learning to Read in a Healing Classroom[138] presented effects on academic attainment, these were not presented as estimates of the intervention's direct effects and thus are not presented here.

In RHC, students were recruited in years 2 and 3 at the start of the intervention (thus straddling KS1 and KS2). Academic performance was measured by both teacher and parent report, and reporting was stratified by sex. Because the range of the outcomes was not clear, we calculated standardised MDs (Cohen's $d$) to facilitate interpretation. Based on teacher report, males in the intervention group ($n = 273$) performed better than males in the control group ($n = 227$) at the end of the first intervention year (intervention group: mean 3.25, SD 1.18; control group: mean 3.06, SD 1.21; $d = 0.16$, 95% CI –0.02 to 0.34) and significantly at the start of the second intervention year (intervention group: mean 3.20, SD 1.15; control group: mean 2.81, SD 1.16; $d = 0.34$, 95% CI 0.16 to 0.52) and the end of the second intervention year (intervention group: mean 2.21, SD 1.16; control group: mean 2.98, SD 1.24; $d = 0.28$, 95% CI 0.10 to 0.45). Females in the intervention group ($n = 224$) were similarly better performing than females in the control group ($n = 214$) at the end of the first intervention year (intervention group: mean 3.35, SD 1.16; control group: mean 3.18, SD 1.25; $d = 0.14$, 95% CI –0.05 to, 0.33), the start of the second intervention year (intervention group: mean 3.28, SD 1.21; control group: mean 3.12, SD 1.17; $d = 0.13$, 95% CI –0.05 to 0.32) and the end of the second intervention year (intervention group: mean 3.34, SD 1.18; control group: mean 3.26, SD 1.18; $d = 0.07$, 95% CI –0.12 to 0.26), but differences were not significant at any point. Differences at the end of the second intervention year in teacher-reported outcomes were tested in a multilevel model and intervention recipients had significantly improved academic performance relative to their control peers ($b = 0.188$; $p = 0.019$). Parent-reported academic achievement reflected an improvement for males at the end of the first year ($d = 0.26$, 95% CI 0.09 to 0.44) and at the end of the second year ($d = 0.23$, 95% CI 0.05 to 0.41), but not for females (end of first intervention year: $d = 0.09$, 95% CI –0.10 to 0.27; end of second intervention year: $d = 0.10$, 95% CI –0.08 to 0.29). However, differences at the end of the second intervention year in parent-reported outcomes were tested in a multilevel model and intervention recipients overall were not significantly better in terms of academic performance ($b = 0.082$; $p > 0.05$).

In the second evaluation of Steps to Respect,[93] including students in years 4 to 6, teacher-reported academic achievement was not significantly better in intervention schools at the end of the first intervention year ($t_{29} = -0.15$). In the evaluation of 4Rs, which included students starting in year 4, intervention recipients were not different from control recipients at the end of the first academic year on academic skill ($b = -0.02$, SE 0.08),

maths ($b = -7.40$, SE 5.69) or reading ($b = -5.35$, SE 4.96) (36-point scale).[52] Findings at the end of the second intervention year for maths ($b = -4.35$, SE 10.25) and reading ($b = -0.99$, SE 10.73) suggested a similar pattern.[3]

### Key stage 3 academic attainment outcomes

The evaluation of Positive Action Chicago[83] included academic attainment outcomes at the end of the sixth intervention year, when students were in year 9. Although maths outcomes from standardised tests were not reported specifically (aside from an observation that differences were not statistically significant), intervention students transitioning from year 8 to year 9 were significantly better on reading ($d = 0.83$; $p = 0.026$). However, there were only small differences in year 9 students on self-reported grades ($d = 0.02$) or teacher-reported academic ability ($d = 0.14$). CIs were not presented for these estimates, but condition-by-time interactions in multilevel models were not statistically significant for self-reported grades or for the better-fitting model in teacher-reported academic ability.

### School-level academic attainment

Both evaluations of Positive Action presented school-level academic attainment outcomes. In Positive Action Chicago,[83] at the end of the sixth intervention year, schools receiving the intervention were better than control schools on maths ($d = 0.38$) and reading ($d = 0.22$). Although CIs for these findings were not provided, findings from two-tailed significance tests of condition-by-time interactions in multilevel models were not significant for reading or maths.

In Positive Action Hawaii, differences between schools in percentage at or above the average for reading and for maths were examined at the end of the fourth and fifth intervention years. For each outcome and follow-up point, estimates were provided for both nationally standardised tests and Hawaii-specific standardised tests. At the end of the fourth intervention year, intervention schools were better, but not significantly so, on nationally standardised maths ($g = 0.50$; $p = 0.495$) and reading ($g = 0.58$; $p = 0.108$), and significantly so on Hawaii-specific maths ($g = 0.69$; $p = 0.04$) and reading ($g = 0.72$; $p = 0.029$) tests. Findings were similar at the end of the fifth intervention year for nationally standardised maths ($g = 0.52$; $p = 0.291$) and reading ($g = 0.54$; $p = 0.028$) and Hawaii-specific maths ($g = 1.1$; $p = 0.006$) and reading ($g = 0.65$; $p = 0.043$).

## Feedback from consultation on outcome synthesis

In September/October 2017, we revisited our stakeholders to reflect broadly on three primary questions: are these findings of relevance to you/your organisation and why? What about these findings is of relevance (e.g. effect size? Rigour of the studies from which these findings are derived?)? How could these findings best be communicated to the right people at your organisation in the right way?

### Relevance of findings

Generally, there seemed to be interest among adult policy and youth stakeholders, the former representing schools, non-governmental organisations and public health institutions. However, perhaps unsurprisingly, there were differing levels of interest. Given that our primary positive finding was a reduction in substance use in KS2 and KS3, if stakeholders were concerned with substance use, these findings were of interest. Stakeholders from schools or organisations that did not emphasise substance use were less interested. More than half of stakeholders were also very interested in outcomes related to academic achievement, and suggested that these should be given emphasis before schools might be willing to consider them. This comment resonates well with some of our findings in previous consultations when we were reflecting on process findings, as alignment with learning objectives and school targets was seen as essential.

Young people on the ALPHA group were surprised to see an effect on substance use but not on violence when the latter is much more visible in schools and, they theorised, open to reductions from this sort of school-based intervention.

Many stakeholders agreed that the findings from outcome evaluations, although necessary, should not be presented in isolation, and argued that they should be presented in the context of our findings about proposed mechanisms of action and suggested barriers and facilitators of implementation.

There was a discussion about effect sizes and whether or not our findings were compelling. Most stakeholders are conversant in appreciating public health research, and as such were not expecting enormous effect sizes. Others, however, felt that the effect sizes might need to be larger to compel administrators to take on an intervention in their school, especially if that intervention involved teacher time and resources and could potentially compromise academic learning.

Importantly, there were various comments raised about the importance of geographical context, with an appreciation that these findings are limited insofar as they mostly do not emerge from the UK and this might dampen the potential impact of such findings. Using these findings to design and implement a UK-based intervention was thought to be considerably more useful in terms of creating an evidence base that UK stakeholders would be more likely to subscribe to.

### Communicating findings

There was considerable variation between stakeholders on this point. However, overwhelmingly, the suggestion that information must come in a digestible manner through a trusted channel resonated among stakeholders. 'Digestible' may take on many forms. Teachers felt that findings should be communicated as if for a lay audience, removing jargon and 'academese'. Other stakeholders who are more engaged with research directly wanted full scientific explanations and lengthier reports alongside summaries. Channels like conferences, meetings and websites (or webinars to communicate research findings) were suggested as potential means of communicating findings. Mass e-mails or e-mails in general were dismissed as being unhelpful.

There was an underlying sentiment from all stakeholders that schools, teachers and administrators are bombarded with information about prospective interventions; to stand out, it must come with a 'seal of approval' from a trusted channel that has worked closely with the school. 'Trusted channels' were perceived differently, however. It was suggested that simply originating from academic institutions may be enough to make the work eye-catching, whereas others suggested that working with known non-governmental organisations or government institutions would be essential. The presentation of findings in terms of content and platforms of dissemination needs to be tailored to the audience.

Finally, stakeholders were unanimous in their understanding of the importance of engagement of an administrative team in schools in order for such interventions to be taken up. Subject heads or heads of pastoral care were suggested as key audiences, although it was noted that, with the goal of school-wide effects or possibly cross-curricular work, the head teacher or assistant head teachers would also have to be championing the intervention. As such, communicating to people occupying these types of roles was felt to be essential.

# Chapter 7 Discussion and conclusion

## Review deviations and limitations

A summary of our minor deviations from the protocol is found in *Appendix 1*. There were no major deviations. However, we did encounter several challenges in conducting the review.

### Searching and inclusion criteria

Our focus was interventions that integrate health and academic education. However, most interventions that included studies evaluated were seeking a practical way to introduce health education into the school timetable, and integration was not always included in report titles or abstracts or at the centre of the theory of change. Searching for these interventions was therefore difficult and it is possible that we missed some potentially relevant studies. Screening was also challenging. All study reports of included studies explicitly described some degree of integration between health and academic education but the extent and clarity of this description varied enormously. It is possible that we excluded studies of interventions where in practice health education was sometimes or always integrated with academic lessons but where this was not mentioned as part of the description of the intervention in study reports. Despite this risk of excluding potentially relevant studies, we think that restricting inclusion to studies of interventions where integration was explicitly referred to as part of the description of the intervention was the only way to ensure that the review had a clear focus and avoided a situation in which screening decisions required detailed examination of intervention materials or lengthy discussion with authors.

We excluded a number of quasi-experimental studies that may have provided useful insights. However, we think that our strategy of excluding these, based on the fact that this class of intervention is amenable to evaluation via RCTs that provide more accurate estimates of intervention effects, was correct.

### Theory reports

There were some limitations both in our review methods and in the literature synthesised. In terms of review methods, the main challenge was in appraising the quality of reports of theory. There was little guidance available from other studies as to how to do this. As is the case in most syntheses of theory,[158] no reports were excluded on the basis of quality and all contributed, albeit to different degrees, to the synthesis on the basis that they provided useful insights despite some areas of weakness. The detail provided on intervention theory varied enormously between reports. Reports providing more detail or theories that were reported on by multiple reports tended to be assessed as of higher quality. There was quite a low rate (65%) of agreement between the two reviewers assessing quality assessment. This reflected several challenges. First, with regards to generalisability, many theories of change were intervention-specific and thus it was not obvious how generalisable they were intended to be. Second, when assessing whether or not theories were based on existing evidence, we found that intervention theories of change often drew from multiple existing theories, each with a different evidence base, rendering overall assessment challenging. Finally, when judging clarity, different authors assumed different levels of prior understanding of constructs among readers, and so provided variable levels of explanations for constructs and their relationships. Reviewers thus found themselves making quite subjective decisions about whether or not authors' assumptions about readers' prior knowledge was reasonable. In some cases, authors providing more text describing constructs and pathways led to a higher quality score, whereas in others authors' assumptions about tacit knowledge were considered reasonable. Because of these challenges, our assessment of quality did not greatly inform how we weighted each report when synthesising. Instead, their contribution to the synthesis was determined by the extent to which they contributed insights into theory that could be reciprocally translated with those from other reports or contribute to line-of-argument syntheses alongside other reports.

In terms of limitations of the literature synthesised, there was a general assumption in the literature that risk behaviours arise from deficits in individual assets such as knowledge, social and emotional skills and attitudes, albeit within a broader social and institutional context. However, it has been argued that behaviours such as violence or substance use emerge as symbolic practices among young people who are, for complex structural and institutional reasons, disengaged from schools in terms of learning and the social community.[75,159] Thus, behaviours such as smoking, drug use or violence might occur not merely through individual deficits but as a result of rationally chosen and socially sanctioned decisions, albeit perhaps in a context of limited choices. The lack of consideration of such mechanisms in intervention theories in this review might lead such interventions to fail to address important influences on young people's substance use and violence because of insufficient dose or traction to transform the engagement and sense of connectedness, particularly of socially disadvantaged students. Moreover, despite some mention of the importance of relationships outside the classroom, the classroom environment was overwhelmingly the focus of theories of change for these interventions, which may fail to address the broader context in which young people develop behaviours. Furthermore, only a few study reports discussed the theorised interaction between educational and non-educational assets and how this would contribute to reductions in risk behaviours.[92,95,102,111] Theoretical models such as developmental cascades[61] and positive youth development[160] have the capacity to address this relationship, but they are not emphasised across our included studies.

Finally, most theories did not consider how contextual factors might interact with mechanisms of change to generate outcomes, as would be the case with theories informed by realist ontologies.[161] Haegerich and Metz[162] extrapolated a summary of moderating factors for some of the same interventions in our synthesis, which included student (e.g. SES, ethnicity, sex, intervention dose) and contextual (e.g. school, family, cultural, policy) characteristics. However, discussions of the role of these factors were consistently underdeveloped in most theories in our synthesis, despite potential implications for the impact of interventions and the extent to which they might translate between different populations and settings.

### Process evaluations

Our synthesis of evidence from process evaluations was somewhat limited by the paucity of in-depth evaluation of the factors that promote or hinder implementation of this category of intervention. Most process evaluations limited themselves to quantitative assessment of implementation fidelity and acceptability, and many failed to report on how implementation was affected by characteristics of interventions, deliverers, participants or school contexts. Evaluations were also very diverse in focus and methods so that thematic synthesis was challenging. Nevertheless, our synthesis did succeed in identifying recurrent and relevant themes from this diverse and rather impoverished body of literature.

Our review overall included 35 outcome evaluations but only 16 process evaluations. This imbalance is common in the literature evaluating social interventions. For example, a review of implementation data by Michie *et al.*[163] found that only 5–30% of publications of experimental studies had detailed intervention descriptions at all. This lack of reporting of intervention and process presents challenges to those aiming to use existing evidence to translate interventions to new settings, a problem noted by several authors across disciplines.[163–168]

### Outcome evaluations

One challenge that arose in our meta-analysis of outcome evaluations was the limitation of statistical methods in accounting for trajectory-based evidence. For example, the effectiveness of RHC[92,95] was most clearly evidenced in the examination of differences in trajectories of frequency of substance use. Examination of trajectories is appropriate in complex interventions aiming to target longitudinal processes of child and adolescent social development, especially as the goal of some interventions was explicitly to target developmental cascades to interrupt the evolution of negative behaviour patterns. This contrasts with 'standard' trials, in which effects are measured at prescribed end points. Although a trajectory-based analysis is an intuitive choice for interventions integrating health and academic education, especially when those interventions are hypothesised to have temporally distal effects, meta-analytic methods are not available to synthesise evidence arising from these

analyses meaningfully. In addition, trajectory analyses may be most appropriate when considering the uptake or frequency, rather than the initiation, of harmful behaviours. Differences in RHC[92,95] in last-year use of substances were not in evidence, but intervention and control groups clearly diverged in the frequency of both alcohol and marijuana use over time.

Our 'matrix'-based meta-analysis method had both strengths and limitations. One strength is that it could examine effects in specific developmental phases of relevance to educators and intervention implementers. However, one limitation is that it involved combining estimates across a diversity of follow-up times. In the narrative synthesis, we attended to the follow-up times attendant to each outcome measurement; however, the few studies included in each meta-analysis precluded examination of heterogeneity by follow-up time.

A consistent problem across assessments of violence perpetration was that measures were frequently indirect, that is, they measured broader constructs such as 'aggression', despite our attempts to locate the most proximal estimate of violent behaviours possible. Measures also drew on a diversity of informants, including student reports, teacher reports and observation, but findings were rarely consistent across informants.

## Summary of key findings

A total of 78,451 unique references were identified from the original searches. Of these, 76,917 (99%) were excluded. 702 references remained, and we were able to obtain the full-text reports of 690 (98%). Of these, 628 studies were further excluded, resulting in 62 reports included in the review. Updated searches retrieved 2355 and 1945 references and resulted in the inclusion of an additional six outcome evaluation reports. Overall, the 68 included reports arose from 30 distinct studies reporting on 19 distinct interventions. Of these reports, 39 reported on theory, 16 reported on processes (from 15 distinct studies) and 41 reported on outcomes (from 16 distinct studies).

### Underlying theories of change for interventions integrating academic and health curricula

An overarching theory that emerged was that eroding boundaries at multiple and mutually reinforcing levels – by integrating health and academic education, by transforming relationships between teachers and students, by generalising learning from classrooms to the wider school environment and by ensuring consistent messages from schools and families – will lead to the development of a community of engaged students oriented towards pro-social behaviour and away from substance use, violence and other risk behaviours.

The interventions in this review aimed to integrate and thus erode boundaries between health and academic education. They also generally aimed to train teachers to model positive behaviours, reinforce such behaviours among students and enhance their relationships with students. We characterised these as fitting within an overarching theme of boundary erosion, but in doing so clarified that this means transforming relationships from being instrumentally focused on academic learning to being more affective and focused on students' broader development. This work in the classroom was supported by other components delivered at multiple levels (e.g. the overall school environment and the family), which aimed to ensure that learning and reinforcement of positive behaviours occurred beyond the classroom. We interpreted this mode of delivery as a mechanism that aimed to erode boundaries between what occurred in classrooms and other settings such as the wider school and students' families. Through these multiple but theoretically related mechanisms, interventions were thus intended to promote experiential learning, the practising and reinforcement of positive behaviours and enhanced relationships between teachers and students and between students and their pro-social peers. In turn, it is theorised that these will provide students with various attributes necessary to reduce engagement in substance use and violence as well as to increase academic attainment. These attributes include the knowledge and skills necessary to avoid substance use and prevent conflict, the

attitudes and social norms supportive of such behaviours and students' sense of security in and positive connectedness to their classrooms and schools.

Ultimately, eroding boundaries between health and academic education, teachers and students, classrooms and the wider school, and schools and families was theorised as being critical to establishing new frameworks of family, classroom or school organisation that are conducive to promoting both academic and social–emotional outcomes.

### Characteristics affecting implementation of interventions integrating health and academic curricula

To answer our original research question, 'what characteristics of interventions, deliverers, participants and school contexts facilitate or limit successful implementation and receipt of such interventions?', we summarised facilitators of and barriers to the implementation of our interventions of interest. Although contributing factors varied widely depending on the intervention, several – often linked – themes did emerge from our synthesis, namely concerning the necessity of support from senior school managers and a conducive broader policy environment, a positive teaching environment, positive pre-existing teacher and student attitudes towards integrated health and academic interventions, favourable staff opinions about the autonomy and innovation that the interventions enabled and parental support of interventions.

It is worth noting that many of the themes above relate to factors affecting implementation that might apply generally to school-based health promotion or social and emotional learning interventions. Here, we aim to draw out what our synthesis suggests about factors affecting the implementation of our specific category of interventions that integrate health and academic education. First, this category of intervention particularly benefits from consistent cross-school support from administrators and colleagues in integrating health across the curriculum. Strong networks, continuous training and shared understanding about the overall aims of integration take time to build and efforts to sustain; thus, ongoing support from administrators, both practically and in terms of morale, is crucial. Second, interventions need to be flexible and locally adaptable if they are to mesh with the existing teaching environment and curriculum. Third, such interventions are innovative and challenging and so require teachers and staff belief in, and commitment to, integration as a longer-term aim to improve students' health and social and emotional learning. But such support appears to be promoted both because teachers value the scope that they provide for local adaptation and professional autonomy and because students value the chance to engage in learning methods that are more participative and topics that appear more relevant to students' lives than might normally be the case in traditional academic subjects.

To our knowledge, there are no existing reviews of interventions that integrate health and academic education. However, reviews of related interventions can help in contextualising our findings. In their review of health promotion interventions in schools, Chilton et al.[169] similarly noted that school and teacher cultural norms concerning substance use affected the extent to which interventions addressing this were successfully implemented. Staff investment overall was critical, including support from administrators. Likewise, Pearson et al.[48] echoed the importance of engaging staff and suggested that 'implementation hinges on negotiation and programme delivery and the acceptability (or otherwise) of the programme to those who deliver it'. They further commented on the importance of deliverers' enthusiasm for the intervention and the need to root it in their perceived responsibility for its success. Bonell et al.'s review[64] of process evaluations of interventions aiming to increase the healthiness of school environments reported, based on five studies, on the importance of a health intervention's alignment with school ethos as a predictor of its success, as well as the importance to good implementation of the broad participation of all staff and support from administrators. Rimm-Kaufman and Hulleman[170] noted similar factors in a review of social and emotional learning interventions in primary schools, emphasising teachers' enthusiasm for interventions as being pivoted on their overall culture of education on these subjects and of these skills. Coupled with school-wide support and ongoing mentoring from higher-level staff, a supporting ethos enhanced teachers' commitment to interventions and was thus crucial to their success.

Indeed, our review provides evidence that teachers' perceptions of their school's teaching culture is a key determinant of successful implementation, something which did not emerge as a key theme in the other reviews cited above. This factor may reflect the importance of genuine integration between health and mainstream academic elements when delivering this particular category of intervention.

### Outcomes of interventions integrating health and academic education

In our synthesis of outcome evaluations, quality of evidence was highly variable and often related to whether or not meta-analysis was possible. The strongest evidence for the effectiveness of interventions integrating academic and health education was found in the reduction of substance use in KS2 and KS3, and a meta-analysis of the effectiveness of these interventions in reducing violence victimisation in KS2 did not find an effect. Included estimates of academic attainment as a secondary outcome did not permit a conclusion as to the impact of these interventions on this domain.

Our findings extend and reflect the findings from prior related systematic reviews on substance use. Although a previous landmark review[41] on alcohol use prevention in schools did not pool study estimates, authors nonetheless concluded that interventions had the potential to be effective. Our meta-analyses found specific evidence of effectiveness in KS3, but not KS4, for interventions integrating health and academic education. Although effects were small in magnitude, at a population level and with universal implementation the impact could be substantial. Similarly, a systematic review of school-based programmes to reduce smoking found evidence for interventions on preventing initiation of smoking.[42] Although our analysis combined all smoking outcomes, we were able to find evidence for effectiveness in reducing smoking in KS4. Again, the magnitude of effects was small but of potential public health significance at a population level. Finally, a major systematic review on illicit drug use prevention in schools found 'small but consistent' effects.[43] As was the case in our meta-analyses, the authors of this existing review noted that most included studies focused on marijuana use outcomes. Our findings reflect theirs in sum and substance, but we were able to locate effects more specifically as occurring in KS3 and KS4, with weaker evidence supporting effectiveness in KS2.

Because we were largely unable to undertake meta-analysis in respect of violence outcomes, our ability to compare with previous systematic reviews is limited. Our findings suggested a mixed pattern of results, and we were hampered by heterogeneous definitions of violence. In contrast, a systematic review of school-based programmes to reduce bullying and victimisation found significant reductions for each.[44] However, this review included non-randomised evidence and a diversity of intervention models.

Across outcome domains, few evaluations showed a consistent pattern of results. Both evaluations of Positive Action[83,87,113–116,129,130] presented clear and compelling evidence of consistent effects in both substance use and violence perpetration. In substance use alone, the evaluation of LIFT[98–101,125] suggested small but sustained effects in substance use initiation. RHC[92,95] did demonstrate differences in frequency, but not prevalence, of substance use. In violence perpetration alone, evaluations presented mixed findings within evaluations. For example, findings from the first evaluation of Steps to Respect[93,103,104] suggested decreases in observed, but not student-reported, bullying. The evaluation of Youth Matters[109–111] had inconsistent findings by analysis method, that is, whereas the main analysis did not present evidence of an effect in bullying perpetration or victimisation at longest follow-up, an exploratory analysis that focused on classifying students into uninvolved students, bullies, victims or bully–victims presented some evidence of increased transition to uninvolved in the intervention group.

This emergent distinction between clearly effective and less clearly effective interventions suggests that examination of these differences is in order. That is, what sets apart Positive Action[83,87,113–116,129,130] and LIFT[98–101,125] (and, to a lesser degree, RHC)[92,95] from other interventions? One possible answer is that trials of all three interventions had substantial developer involvement, which is associated with improved effectiveness in a variety of intervention domains.[171] An equally important, and equally plausible, answer is in the developmental period in which these interventions were initiated. All three interventions began in the equivalent of English and Welsh schools' year 6 or earlier, and all three interventions incorporated 'wraparound' techniques beyond

integration of academic and health education. For example, Positive Action emphasised a whole-school approach,[55,83,85–87,102,113–116,118,129,130] LIFT emphasised home–school interactions[98–101,125] and RHC[92,95] included substantial programming beyond academic and health integration over several years. Moreover, what sets apart these interventions from the Gatehouse Project, which incorporated similar wraparound in the school setting, is the relatively earlier timing of intervention initiation. Positive Action in particular is a highly manualised intervention including lessons over multiple school years and concomitant efforts to change school climate. Echoing findings from our synthesis of process and theory studies, Positive Action was the intervention most likely of the evaluations included in our synthesis of outcome evaluations to involve erosion of boundaries between students and teachers.

## Implications for research

Our review of theoretical literature provides a clearer definition of this relatively novel but important category of intervention and an insight into the common mechanisms through which diverse real-life examples of integration are intended to work. Our challenges in appraising the quality of theoretical reports indicate that, although quality assessment was certainly necessary, there continues to be a need to refine quality assessment criteria for theoretical literature. Our suggestion is that in future theory syntheses of this kind, further consideration should be given to the usefulness of theories for the specific questions underlying each review. For example, here it would have been most useful for us to focus on theories that emphasised the mechanisms by which integration of academic and health curricula produced desired outcomes. Unfortunately, for many interventions with rich description of theory, this integration was not always mentioned, or was mentioned with very little detail. As a result, our focus might have been on the quality of descriptions of integration rather than simply the interventions overall. Additionally, for fields where there is little previous theoretical cohesion, it may be necessary to synthesise every theory in equal measure, regardless of what may be somewhat arbitrary quality criteria.[3]

Our current review confirms the relative paucity of process evaluation in this field as in many others. We refer earlier to the limited examination of context in particular and recommend that this receives more focus in future process evaluations. Especially within complex public health interventions that aim to bring about behaviour change, both implementation and outcomes are inevitably influenced by context. Realist approaches are helpful here, as they aim to test hypotheses concerning how content interacts with intervention implementation and mechanisms to generate outcomes.[161,164]

Finally, in the synthesis of outcome evaluations, we noted the challenges of accounting for developmental processes while also including end point-based estimates of outcomes. In future reports, authors should be encouraged to include both types of measures where study design allows it, that is, both estimates of differences in change over time and estimates of differences between groups at measurement occasions. This will facilitate a complete picture of intervention effectiveness. In addition, evaluations of interventions integrating academic and health education should make an effort to understand subgroup effects more specifically. Although some evaluations (namely, Positive Action Hawaii and Chicago, and, to a lesser degree, I-LST) considered subgroup effects, more careful attention to moderation by school-level characteristics could yield useful knowledge about context-intervention fit, facilitating as well the application of knowledge from process and theory studies.

## Implications for policy

In terms of policy, this intervention type offers the potential to address health in busy school timetables, addressing health determinants relating to health knowledge and skills as well as to school engagement and education more generally. However, our findings suggest that integrative interventions, while attractive as ways to deliver some health, social and emotional learning in the context of school systems overwhelmingly focused on educational attainment, should not be viewed as a panacea as their implementation poses

particular challenges. Proper integration requires that teachers and their senior administrators believe in the interventions and have the time and resources to reflect on and build seamless integration between components, that interventions have enough flexibility to be applied appropriately in diverse contexts and that the baseline teaching culture of a school offers hope for this type of intervention to be included into an already positive environment. This category of interventions will not be implemented well in settings where staff are demoralised and change jobs frequently, where they are sceptical about integrating health into their lessons or where managerial and collegial support for this challenging work of integration is perceived as an issue. As a result, our findings suggest that both supportive external policies at the government level and internal policies at the school level might support such programming. Government policies might include measures of school achievement that are both related to student health and wellness as well as academic achievement, noting their often reciprocal relationship. Having a mandate from the government to focus on student health and wellness may enable greater allocations of time and resources in schools for health alongside academics. Although we found strong evidence for the effectiveness of this category of interventions in reducing substance use in KS2 and KS3, we found no evidence for the effectiveness of these interventions in reducing violence victimisation in KS2, and findings regarding other outcomes were mixed.

## Conclusions

Interventions integrating health and academic education are theorised to work via eroding boundaries between health and academic education as well as other forms of boundaries in schools. Their implementation is facilitated by institutional support and teacher motivation. Such interventions may be effective in reducing substance use but there is currently no evidence that they reduce outcomes related to violence. Most evidence originated from settings outside the UK. We suggest that where interventions integrating health and academic education are delivered in the UK, this should be in the context of evaluation studies to assess their effectiveness, and trials of interventions addressing substance use are likely to have more potential than trials of interventions addressing violence outcomes.

# Acknowledgements

The work was undertaken with the support of DECIPHer, a UK Clinical Research Collaboration Public Health Research Centre of Excellence.

We would like to thank the ALPHA young people's public input advisory group based in DECIPHer.

## Policy/practitioner stakeholders

Jonathan Baggaley (Chief Executive, PSHE Association).

Paul Ainsworth (Director, Ariel Trust).

Michael O'Toole (Chief Executive, Mentor UK).

Mary Charles (Principal Public Health Practitioner, Public Health Wales).

Claire Robson (Public Health Delivery Manager for Children, Young People and Families Team, Public Health England).

Joan Roberts (School Health Research Network Manager, School Health Research Network).

Emma Rigby (Chief Executive, Association for Young People's Health).

Ann Hagell (Research Lead, Association for Young People's Health).

Sindee Bass (Special Education Needs Co-ordinator, London Fields Primary School).

Nick Bonell (Head Teacher, St. Saviour's Church of England Junior School).

Jonathan Lewis (Newstead Wood School).

Greig Pilkington (Barnhill Community High School).

## Contributions of authors

**Tara Tancred** (Research Fellow in Evidence Synthesis, London School of Hygiene & Tropical Medicine) carried out web searches, contributed to screening and quality appraisal/data extraction of studies and led on synthesising theory and process evaluations and drafting the report.

**GJ Melendez-Torres** (Senior Lecturer in Social Science and Health, Cardiff University) contributed to project planning, screening of studies on title and abstract, led the review and synthesis of outcome evaluations and contributed towards drafting the report.

**Sara Paparini** (Research Fellow in Sociology, London School of Hygiene & Tropical Medicine) contributed to the synthesis of theory and process evaluations and drafting the report.

**Adam Fletcher** (Professor of Social Science and Health, Cardiff University) contributed to project planning and screening of studies on title and abstract, led the consultation with young people and commented on report drafts.

## Publications

Melendez-Torres GJ, Tancred T, Fletcher A, Campbell R, Thomas J, Bonell C. Integration of academic and health education for the prevention of violence in young people: systematic review, narrative synthesis and intervention components analysis. *BMJ Open* 2018;**8**:e020793.

Melendez-Torres GJ, Tancred T, Fletcher A, Thomas J, Campbell R, Bonell C. Does integrated academic and health education prevent substance use? Systematic review and meta-analyses. *Child* 2018;**44**:516–30.

Tancred T, Paparini S, Melendez-Torres GJ, Fletcher A, Thomas J, Campbell R, Bonell C. Interventions integrating health and academic interventions to prevent substance use and violence: a systematic review and synthesis of process evaluations. *Syst Rev* 2018;**7**:227.

Tancred T, Paparini S, Melendez-Torres GJ, Thomas J, Fletcher A, Campbell R, Bonell C. A systematic review and synthesis of theories of change of school-based interventions integrating health and academic education as a novel means of preventing violence and substance use among students. *Systematic Reviews* 2018;**7**:190.

## Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. Access to available anonymised data may be granted following review.

# References

1. Bonell C, Fletcher A, Melendez-Torres GJ, Thomas J, Campbell R. *How Can Interventions Integrating Health and Academic Education in Schools Help Prevent Substance Misuse and Violence, and Reduce Health Inequalities Among Young People? Systematic Review and Evidence Synthesis*. Southampton: National Institute for Health Research; 2016. URL: www.journalslibrary.nihr.ac.uk/programmes/phr/145215/#/ (accessed 9 September 2018).

2. Bier MC, Zwarun L, Fehrmann Warren V. Getting universal primary tobacco use prevention into priority area schools: a media literacy approach. *Health Promot Pract* 2011;**12**(Suppl. 2):152–8. https://doi.org/10.1177/1524839911414887

3. Jones SM, Brown JL, Lawrence Aber J. Two-year impacts of a universal school-based social-emotional and literacy intervention: an experiment in translational developmental research. *Child Dev* 2011;**82**:533–54. https://doi.org/10.1111/j.1467-8624.2010.01560.x

4. Kupersmidt JB, Scull TM, Benson JW. Improving media message interpretation processing skills to promote healthy decision making about substance use: the effects of the middle school media ready curriculum. *J Health Commun* 2012;**17**:546–63. https://doi.org/10.1080/10810730.2011.635769

5. Lynskey MT, Fergusson DM. Childhood conduct problems, attention deficit behaviors, and adolescent alcohol, tobacco, and illicit drug use. *J Abnorm Child Psychol* 1995;**23**:281–302. https://doi.org/10.1007/BF01447558

6. Nutt DJ, King LA, Phillips LD, Independent Scientific Committee on Drugs. Drug harms in the UK: a multicriteria decision analysis. *Lancet* 2010;**376**:1558–65. https://doi.org/10.1016/S0140-6736(10)61462-6

7. Lifestyle Statistics, Health and Social Care Information Centre. *Statistics on Alcohol: England, 2013*. London: Government Statistical Service; 2013.

8. Great Britain. *The Government's Alcohol Strategy*. London: The Stationery Office; 2012.

9. Droomers M, Schrijvers CT, Casswell S, Mackenbach JP. Occupational level of the father and alcohol consumption during adolescence: patterns and predictors. *J Epidemiol Community Health* 2003;**57**:704–10. https://doi.org/10.1136/jech.57.9.704

10. Hingson RW, Heeren T, Edwards EM. Age at drinking onset, alcohol dependence, and their relation to drug use and dependence, driving under the influence of drugs, and motor-vehicle crash involvement because of drugs. *J Stud Alcohol Drugs* 2008;**69**:192–201. https://doi.org/10.15288/jsad.2008.69.192

11. Viner RM, Taylor B. Adult outcomes of binge drinking in adolescence: findings from a UK national birth cohort. *J Epidemiol Community Health* 2007;**61**:902–7. https://doi.org/10.1136/jech.2005.038117

12. World Health Organization (WHO). *Evidence-based Strategies and Interventions to Reduce Alcohol-related Harm*. Geneva: WHO; 2007.

13. Masterman PW, Kelly AB. Reaching adolescents who drink harmfully: fitting intervention to developmental reality. *J Subst Abuse Treat* 2003;**24**:347–55. https://doi.org/10.1016/S0740-5472(03)00047-3

14. Department of Health and Social Care. *Healthy Lives, Healthy People: Our Strategy for Public Health in England*. London: Department of Health and Social Care; 2010.

15. Allender S, Balakrishnan R, Scarborough P, Webster P, Rayner M. The burden of smoking-related ill health in the UK. *Tob Control* 2009;**18**:262–7. https://doi.org/10.1136/tc.2008.026294

16. Nash R, Featherstone H. *Cough Up: Balancing Tobacco Income and Costs in Society*. London: Policy Exchange; 2010.

17. Dunstan S. *General Lifestyle Survey Overview: A Report on the 2010 General Lifestyle Survey*. Newport: Office for National Statistics; 2012.

18. Hibell B, Guttormsson U, Ahlström S, Balakireva O, Bjarnason T, Kokkevi A, *et al. The 2011 ESPAD Report: Substance Use Among Students in 36 European Countries*. Stockholm: The Swedish Council for Information on Alcohol and other Drugs (CAN); 2012.

19. Secretary of State for Health, Secretary of State for Scotland, Secretary of State for Wales, Secretary of State for Northern Ireland. *Smoking Kills: A White Paper on Tobacco*. London: The Stationery Office; 1998.

20. Fergusson DM, Boden JM, Horwood LJ. Cannabis use and other illicit drug use: testing the cannabis gateway hypothesis. *Addiction* 2006;**101**:556–69. https://doi.org/10.1111/j.1360-0443.2005.01322.x

21. Hall W. The mental health risks of adolescent cannabis use. *PLOS Med* 2006;**3**:e39. https://doi.org/10.1371/journal.pmed.0030039

22. Moore TH, Zammit S, Lingford-Hughes A, Barnes TR, Jones PB, Burke M, Lewis G. Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. *Lancet* 2007;**370**:319–28. https://doi.org/10.1016/S0140-6736(07)61162-3

23. Parrott AC, Lees A, Garnham NJ, Jones M, Wesnes K. Cognitive performance in recreational users of MDMA of 'ecstasy': evidence for memory deficits. *J Psychopharmacol* 1998;**12**:79–83. https://doi.org/10.1177/026988119801200110

24. Beautrais AL, Joyce PR, Mulder RT. Cannabis abuse and serious suicide attempts. *Addiction* 1999;**94**:1155–64. https://doi.org/10.1046/j.1360-0443.1999.94811555.x

25. Charlton J, Kelly S, Dunnell K, Evans B, Jenkins R. Suicide deaths in England and Wales: trends in factors associated with suicide deaths. *Popul Trends* 1993;**71**:34–42.

26. Thomas J, Kavanagh J, Tucker H, Burchett H, Tripney J, Oakley A. *Accidental Injury, Risk-taking Behaviour and the Social Circumstances in which Young People (Aged 12–24) Live: A Systematic Review*. London: The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre); 2007.

27. Calafat A, Blay N, Juan M, Adrover D, Bellis MA, Hughes K, *et al.* Traffic risk behaviors at nightlife: drinking, taking drugs, driving, and use of public transport by young people. *Traffic Inj Prev* 2009;**10**:162–9. https://doi.org/10.1080/15389580802597054

28. Great Britain. *Updated Drug Strategy 2002*. London: The Stationery Office; 2002.

29. Jayakody A, Sinha S, Curtis K, Roberts H, Viner R. *Smoking, Drinking, Drug Use, Mental Health and Sexual Behaviour in Young People in East London*. London: Department of Health and Teenage Pregnancy Unit; 2005.

30. Jessor R, Donovan JE, Costa FM. *Beyond Adolescence: Problem Behaviour and Young Adult Development*. Cambridge: Cambridge University Press; 1994.

31. Krug EG, Mercy JA, Dahlberg LL, Zwi AB. The world report on violence and health. *Lancet* 2002;**360**:1083–8. https://doi.org/10.1016/S0140-6736(02)11133-0

32. Scott S, Knapp M, Henderson J, Maughan B. Financial cost of social exclusion: follow up study of antisocial children into adulthood. *BMJ* 2001;**323**:191. https://doi.org/10.1136/bmj.323.7306.191

33. Beinart S, Anderson B, Lee S, Utting D. *Youth at Risk? A National Survey of Risk Factors, Protective Factors and Problem Behaviour among Young People in England, Scotland and Wales (JRF Findings 432)*. York: Joseph Rowntree Foundation; 2002.

34. Bender D, Lösel F. Bullying at school as a predictor of delinquency, violence and other anti-social behaviour in adulthood. *Crim Behav Ment Health* 2011;**21**:99–106. https://doi.org/10.1002/cbm.799

35. Olweus D. *Bullying at School: What We Know and What We Can Do*. Malden, MA: Blackwell Publishing; 1993.

36. Parsonage M. *Diversion: A Better Way for Criminal Justice and Mental Health*. London: Sainsbury Centre for Mental Health; 2009.

37. Fletcher A, Bonell C, Sorhaindo A. 'We don't have no drugs education': the myth of universal drugs education in English secondary schools? *Int J Drug Policy* 2010;**21**:452–8. https://doi.org/10.1016/j.drugpo.2010.09.009

38. National Association of Schoolmasters Union of Women Teachers (NASUWT). *English Baccalaureate Survey Summary*. Birmingham: NASUWT; 2011.

39. PSHE Association. *Comments on the National Curriculum Proposals Published in February 2013 from the PSHE Education Strategic Partners' Group*. London: PHSEA; 2013.

40. The Office for Standards in Education, Children's Services and Skills (OFSTED). *School Inspection Handbook: Handbook for Inspecting Schools in England Under Section 5 of the Education Act 2005*. Manchester: OFSTED; 2015.

41. Foxcroft DR, Tsertsvadze A. Cochrane review: universal school-based prevention programs for alcohol misuse in young people. *Evidence-Based Child Health* 2012;**7**:450–575. https://doi.org/10.1002/ebch.1829

42. Thomas RE, McLellan J, Perera R. School-based programmes for preventing smoking. *Cochrane Database Syst Rev* 2013;**4**:CD001293. https://doi.org/10.1002/14651858.CD001293.pub3

43. Faggiano F, Vigna-Taglianti FD, Versino E, Zambon A, Borraccino A, Lemma P. School-based prevention for illicit drugs' use. *Cochrane Database Syst Rev* 2005;**2**:CD003020. https://doi.org/10.1002/14651858.CD003020.pub2

44. Farrington DP, Ttofi MM. *School-based Programs to Reduce Bullying and Victimization*. Oslo: The Campbell Collaboration; 2009. https://doi.org/10.1037/e528362010-001

45. Hahn R, Fuqua-Whitley D, Wethington H, Lowy J, Crosby A, Fullilove M, *et al.* Effectiveness of universal school-based programs to prevent violent and aggressive behavior: a systematic review. *Am J Prev Med* 2007;**33**(Suppl. 2):114–29. https://doi.org/10.1016/j.amepre.2007.04.012

46. Vreeman RC, Carroll AE. A systematic review of school-based interventions to prevent bullying. *Arch Pediatr Adolesc Med* 2007;**161**:78–88. https://doi.org/10.1001/archpedi.161.1.78

47. Formby E, Coldwell M, Stiell B, Demack S, Stevens A, Shipton L, *et al. Personal, Social, Health and Economic (PSHE) Education: A Mapping Study of the Prevalent Models of Delivery and their Effectiveness*. London: Department for Education; 2011. https://doi.org/10.1037/e605582011-001

48. Pearson M, Chilton R, Wyatt K, Abraham C, Ford T, Woods HB, Anderson R. Implementing health promotion programmes in schools: a realist systematic review of research and experience in the United Kingdom. *Implement Sci* 2015;**10**:149. https://doi.org/10.1186/s13012–015–0338–6

49. The British Heart Foundation (BHF). *Money to Burn: Lesson Plan*. BHF; 2014. URL: http://teachers.theguardian.com/teacher-resources/15410/Money-to-burn——lesson-plan (accessed 15 January 2017).

50. Wright G, Ainsworth P. *Plastered Evaluation: Part of It's Not OK! Violence Prevention Education Programme*. Liverpool: Ariel Trust; 2008.

51. Jones SM, Brown JL, Aber JL. *Three Year Cumulative Impacts of the* 4Rs *Program on Children's Social-Emotional, Behavioral, and Academic Outcomes*. Evanston, IL: Society for Research on Educational Effectiveness; 2010.

52. Jones SM, Brown JL, Hoglund WL, Aber JL. A school-randomized clinical trial of an integrated social-emotional learning and literacy intervention: impacts after 1 school year. *J Consult Clin Psychol* 2010;**78**:829–42. https://doi.org/10.1037/a0021383

53. Aber L, Brown JL, Jones SM, Berg J, Torrente C. School-based strategies to prevent violence, trauma, and psychopathology: the challenges of going to scale. *Dev Psychopathol* 2011;**23**:411–21. https://doi.org/10.1017/S0954579411000149

54. Brown JL, Jones SM, LaRusso MD, Aber JL. Improving classroom quality: teacher influences and experimental impacts of the 4Rs program. *J Educ Psychol* 2010;**102**:153–67 https://doi.org/10.1037/a0018160

55. Flay B, Berkowitz MW, Bier MC. Elementary school-based programs theorized to support social development, prevent violence, and promote positive school climate: description and hypothesized mechanisms of change. *J Res Char Educ* 2009;**7**:21–49.

56. Jones SM, Brown JL, Aber JL. *Classroom Settings as Targets of Intervention and Research*. Washington, DC: American Psychological Association; 2008. https://doi.org/10.1093/acprof:oso/9780195327892.003.0004

57. Sung J-A. *Teachers' Perceptions and Experiences of the Implementation of a Social-Emotional Learning Program in an Inner-city Public Elementary School*. New York, NY: Columbia University; 2015.

58. Goleman D. *Emotional Intelligence*. New York, NY: Bantam; 1995.

59. Flay BR, Graumlich S, Segawa E, Burns JL, Holliday MY, Aban Aya Investigators. Effects of 2 prevention programs on high-risk behaviors among African American youth: a randomized trial. *Arch Pediatr Adolesc Med* 2004;**158**:377–84. https://doi.org/10.1001/archpedi.158.4.377

60. Patton GC, Bond L, Carlin JB, Thomas L, Butler H, Glover S, *et al.* Promoting social inclusion in schools: a group-randomized trial of effects on student health risk behavior and well-being. *Am J Public Health* 2006;**96**:1582–7 https://doi.org/10.2105/AJPH.2004.047399

61. Masten AS, Cicchetti D. Developmental cascades. *Dev Psychopathol* 2010;**22**:491–5. https://doi.org/10.1017/S0954579410000222

62. Bonell C, Dickson K, Hinds K, Melendez-Torres G, Stansfield C, Fletcher A. 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**(5). https://doi.org/10.3310/phr04050

63. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Med* 2009;**6**:e1000097. https://doi.org/10.1371/journal.pmed.1000097

64. Bonell C, Jamal F, Harden A, Wells H, Parry W, Fletcher A, *et al.* Systematic review of the effects of schools and school environment interventions on health: evidence mapping and synthesis. *Public Health Res* 2013;**1**(1). https://doi.org/10.3310/phr01010

65. Pound P, Campbell R. Exploring the feasibility of theory synthesis: a worked example in the field of health related risk-taking. *Soc Sci Med* 2015;**124**:57–65. https://doi.org/10.1016/j.socscimed.2014.11.029

66. Harden A, Rees R, Shepherd J, Brunton G, Oliver S, Oakley A. *Young People and Mental Health: A Systematic Review of Barriers and Facilitators*. London: EPPI-Centre, Social Science Research Unit; 2001.

67. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. Oxford: The Cochrane Collaboration; 2011.

68. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. *BMC Med Res Methodol* 2009;**9**:59. https://doi.org/10.1186/1471–2288–9–59

69. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol* 2008;**8**:45. https://doi.org/10.1186/1471–2288–8–45

70. Bonell CP, Fletcher A, Jamal F, Wells H, Harden A, Murphy S, Thomas J. Theories of how the school environment impacts on student health: systematic review and synthesis. *Health Place* 2013;**24**:242–9. https://doi.org/10.1016/j.healthplace.2013.09.014

71. Campbell R, Pound P, Pope C, Britten N, Pill R, Morgan M, Donovan J. Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Soc Sci Med* 2003;**56**:671–84. https://doi.org/10.1016/S0277-9536(02)00064-3

72. Noblit GW, Hare RD. *Meta-ethnography: Synthesizing Qualitative Studies*. London: SAGE Publications; 1988. https://doi.org/10.4135/9781412985000

73. Campbell M, Egan M, Lorenc T, Bond L, Popham F, Fenton C, Benzeval M. Considering methodological options for reviews of theory: illustrated by a review of theories linking income and health. *Syst Rev* 2014;**3**:114. https://doi.org/10.1186/2046–4053–3–114

74. Schutz A. *Collected Papers*. The Hague: Martinus Nijhoff; 1962.

75. Jamal F, Fletcher A, Harden A, Wells H, Thomas J, Bonell C. The school environment and student health: a systematic review and meta-ethnography of qualitative research. *BMC Public Health* 2013;**13**:798. https://doi.org/10.1186/1471–2458–13–798

76. Britten N, Campbell R, Pope C, Donovan J, Morgan M, Pill R. Using meta ethnography to synthesise qualitative research: a worked example. *J Health Serv Res Policy* 2002;**7**:209–15. https://doi.org/10.1258/135581902320432732

77. Markham WA, Aveyard P. A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice. *Soc Sci Med* 2003;**56**:1209–20. https://doi.org/10.1016/S0277-9536(02)00120-X

78. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, *et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC Methods Programme*. Lancaster: Lancaster University; 2006.

79. Cheung MW. Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychol Methods* 2014;**19**:211–29. https://doi.org/10.1037/a0032968

80. Van den Noortgate W, López-López JA, Marín-Martínez F, Sánchez-Meca J. Meta-analysis of multiple outcomes: a multilevel approach. *Behav Res Methods* 2015;**47**:1274–94. https://doi.org/10.3758/s13428–014–0527–2

81. Social Research Association (SRA). *Ethical Guidelines*. London: SRA; 2003.

82. Centre for Social Justice and Community Action (CSJCA), National Coordinating Centre for Public Engagement (NCCPE). *Community-based Participatory Research: A Guide to Ethical Principles and Practice*. Bristol: Durham University; 2012.

83. Bavarian N, Lewis KM, Dubois DL, Acock A, Vuchinich S, Silverthorn N, *et al.* Using social-emotional and character development to improve academic outcomes: a matched-pair, cluster-randomized controlled trial in low-income, urban schools. *J Sch Health* 2013;**83**:771–9. https://doi.org/10.1111/josh.12093

84. Bechtel LJ, Vicary JR, Swisher JD, Smith EA, Hopkins AM, Henry KL, *et al.* An interdisciplinary approach for the integration and diffusion of substance abuse prevention programs. *Am J Health Educ* 2006;**37**:219–25. https://doi.org/10.1080/19325037.2006.10598906

85. Beets M. *Factors Associated with the Implementation Fidelity of a School-based Social and Character Development Program: Findings from the Positive Action Program, Hawai'i*. PhD thesis. Corvallis, OR: Oregon State University; 2007.

86. Beets MW, Flay BR, Vuchinich S, Acock AC, Li KK, Allred C. School climate and teachers' beliefs and attitudes associated with implementation of the Positive Action Program: a diffusion of innovations model. *Prev Sci* 2008;**9**:264–75. https://doi.org/10.1007/s11121–008–0100–2

87. Beets MW, Flay BR, Vuchinich S, Snyder FJ, Acock A, Li KK, *et al.* Use of a social and character development program to prevent substance use, violent behaviors, and sexual activity among elementary-school students in Hawaii. *Am J Public Health* 2009;**99**:1438–45. https://doi.org/10.2105/AJPH.2008.142919

88. Bond L, Butler H. The Gatehouse Project: a Multi-level Integrated Approach to Promoting Wellbeing in Schools. In Brownson RC, Baker EA, Deshpande AD, Gillespie KN, editors. *Evidence-based Public Health*. Oxford: Oxford University Press; 2010. https://doi.org/10.1093/acprof:oso/9780199563623.003.016

89. Bond L, Glover S, Godfrey C, Butler H, Patton GC. Building capacity for system-level change in schools: lessons from the Gatehouse Project. *Health Educ Behav* 2001;**28**:368–83. https://doi.org/10.1177/109019810102800310

90. Bond L, Patton G, Glover S, Carlin JB, Butler H, Thomas L, *et al.* The Gatehouse Project: can a multilevel school intervention affect emotional wellbeing and health risk behaviours? *J Epidemiol Community Health* 2004;**58**:997–1003. https://doi.org/10.1136/jech.2003.009449

91. Bond L, Thomas L, Coffey C, Glover S, Butler H, Carlin JB, Patton G. Long-term impact of the Gatehouse Project on cannabis use of 16-year-olds in Australia. *J Sch Health* 2004;**74**:23–9. https://doi.org/10.1111/j.1746-1561.2004.tb06597.x

92. Brown EC, Catalano RF, Fleming CB, Haggerty KP, Abbott RD. Adolescent substance use outcomes in the Raising Healthy Children project: a two-part latent growth curve analysis. *J Consult Clin Psychol* 2005;**73**:699–710. https://doi.org/10.1037/0022-006X.73.4.699

93. Brown EC, Low S, Smith BH, Haggerty KP. Outcomes from a school-randomized controlled trial of steps to respect: a bullying prevention program. *School Psych Rev* 2011;**40**:423–43. https://doi.org/10.1037/e734362011-045

94. Cain G, Carnellor Y. 'Roots of empathy': a research study on its impact on teachers in Western Australia. *The Journal of Student Wellbeing* 2008;**2**:52–73. https://doi.org/10.21913/JSW.v2i1.168

95. Catalano RF, Mazza JJ, Harachi TW, Abbott RD, Haggerty KP, Fleming CB. Raising healthy children through enhancing social development in elementary school: results after 1.5 years. *J Sch Psychol* 2003;**41**:143–64. https://doi.org/10.1016/S0022-4405(03)00031-1

96. Couch L. *The Bullying Literature Project: an Evaluation of a Class-wide Bullying Intervention Program*. PhD thesis. Riverside, CA: University of California; 2015.

97. Crean HF, Johnson DB. Promoting Alternative THinking Strategies (PATHS) and elementary school aged children's aggression: results from a cluster randomized trial. *Am J Community Psychol* 2013;**52**:56–72. https://doi.org/10.1007/s10464-013-9576-4

98. DeGarmo DS, Eddy JM, Reid JB, Fetrow RA. Evaluating mediators of the impact of the Linking the Interests of Families and Teachers (LIFT) multimodal preventive intervention on substance use initiation and growth across adolescence. *Prev Sci* 2009;**10**:208–20. https://doi.org/10.1007/s11121–009–0126–0

99. Eddy JM, Barkan SE, Lanham L. *Preventive Intervention to Reduce Youth Conduct Problems and Substance Use: Linking the Interests of Families and Teachers (LIFT)*. Washington, DC: American Psychological Association; 2015.

100. Eddy JM, Reid JB, Fetrow RA. An elementary school-based prevention program targeting modifiable antecedents of youth delinquency and violence Linking the Interests of Families and Teachers (LIFT). *J Emot Behav Disord* 2000;**8**:165–76. https://doi.org/10.1177/106342660000800304

101. Eddy JM, Reid JB, Stoolmiller M, Fetrow RA. Outcomes during middle school for an elementary school-based preventive intervention for conduct problems: follow-up results from a randomized trial. *Behav Ther* 2003;**34**:535–52. https://doi.org/10.1016/S0005-7894(03)80034-5

102. Flay B, Allred C. The Positive Action Program: Improving Academics, Behavior, and Character by Teaching Comprehensive Skills for Successful Learning and Living. In Loval T, Toomey R, Clement N, editors. *International Research Handbook on Values Education and Student Wellbeing*. New York, NY: Springer; 2010. pp. 471–501. https://doi.org/10.1007/978-90-481-8675-4_28

103. Frey KS, Hirschstein MK, Edstrom LV, Snell JL. Observed reductions in school bullying, nonbullying aggression, and destructive bystander behavior: a longitudinal evaluation. *J Educ Psychol* 2009;**101**:466–81. https://doi.org/10.1037/a0013839

104. Frey KS, Hirschstein MK, Snell JL, Edstrom LV, MacKenzie EP, Broderick CJ. Reducing playground bullying and supporting beliefs: an experimental trial of the steps to respect program. *Dev Psychol* 2005;**41**:479–90. https://doi.org/10.1037/0012-1649.41.3.479

105. Gordon M. Roots of empathy: responsive parenting, caring societies. *Keio J Med* 2003;**52**:236–43. https://doi.org/10.2302/kjm.52.236

106. Greenberg M, Kusché C. Building social and emotional competence: the PATHS curriculum. In Jimerson S, Furlong M, editors. *Handbook of School Violence and School Safety*. New York, NY: Taylor & Francis Group; 2006. pp. 395–412.

107. Hanson JL. *Teachers' Beliefs About Emotions in the Classroom: Relations to Teacher Characteristics and Implementation of a Social-Emotional Learning Program*. Masters thesis. Vancouver, CA: University of British Columbia; 2012.

108. Holcomb JD, Denk JP. An interdisciplinary English/health connection: promoting health awareness and healthy behaviors. *J Health Educ* 1993;**24**:36–41. https://doi.org/10.1080/10556699.1993.10616432

109. Jenson JM, Brisson D, Bender KA, Williford AP. Effects of the Youth Matters prevention program on patterns of bullying and victimization in elementary and middle school. *Soc Work Res* 2013;**37**:361–72. https://doi.org/10.1093/swr/svt030

110. Jenson JM, Dieterich WA. Effects of a skills-based prevention program on bullying and bully victimization among elementary school children. *Prev Sci* 2007;**8**:285–96. https://doi.org/10.1007/s11121–007–0076–3

111. Jenson JM, Dieterich WA, Brisson D, Bender KA, Powell A. Preventing childhood bullying: findings and lessons from the Denver Public Schools Trial. *Res Soc Work Pract* 2010;**20**:509–17. https://doi.org/10.1177/1049731509359186

112. Kusché C, Greenberg M. The PATHS Curriculum: Promoting Emotional Literacy, Prosocial Behavior, and Caring Classrooms. In Jimerson SR, Nickerson AB, Mayer MJ, Furlong MJ, editors. *Handbook of School Violence and School Safety: International Research and Practice*. New York, NY: Routledge; 2012. pp. 435–46.

113. Lewis KM. *Evaluation of a Social-Emotional and Character Development Program: Methods and Outcomes*. Corvallis, OR: Oregon State University; 2012.

114. Lewis KM, Bavarian N, Snyder FJ, Acock A, Day J, Dubois DL, *et al.* Direct and mediated effects of a social-emotional and character development program on adolescent substance use. *Int J Emot Educ* 2012;**4**:56–78.

115. Lewis KM, Schure MB, Bavarian N, DuBois DL, Day J, Ji P, *et al.* Problem behavior and urban, low-income youth: a randomized controlled trial of positive action in Chicago. *Am J Prev Med* 2013;**44**:622–30. https://doi.org/10.1016/j.amepre.2013.01.030

116. Li KK, Washburn I, DuBois DL, Vuchinich S, Ji P, Brechling V, *et al.* Effects of the Positive Action programme on problem behaviours in elementary school students: a matched-pair randomised control trial in Chicago. *Psychol Health* 2011;**26**:187–204. https://doi.org/10.1080/08870446.2011.531574

117. Low S, Van Ryzin MJ, Brown EC, Smith BH, Haggerty KP. Engagement matters: lessons from assessing classroom implementation of steps to respect – a bullying prevention program over a one-year period. *Prev Sci* 2014;**15**:165–76. https://doi.org/10.1007/s11121–012–0359–1

118. Malloy M, Acock A, DuBois DL, Vuchinich S, Silverthorn N, Ji P, Flay BR. Teachers' perceptions of school organizational climate as predictors of dosage and quality of implementation of a social-emotional and character development program. *Prev Sci* 2015;**16**:1086–95. https://doi.org/10.1007/s11121–014–0534–7

119. Malm B, Löfgren H. Empowering students to handle conflicts through the use of drama. *J Peace Edu* 2007;**4**:1–20. https://doi.org/10.1080/17400200601171164

120. O'Toole J, Burton B. Acting against conflict and bullying: the Brisbane DRACON project 1996–2004 – emergent findings and outcomes. *Res Drama Educ* 2005;**10**:269–83. https://doi.org/10.1080/13569780500275873

121. Patton G, Bond L, Butler H, Glover S. Changing schools, changing health? Design and implementation of the Gatehouse Project. *J Adolesc Health* 2003;**33**:231–9. https://doi.org/10.1016/S1054-139X(03)00204-0

122. Patton GC, Glover S, Bond L, Butler H, Godfrey C, Di Pietro G, Bowes G. The Gatehouse Project: a systematic approach to mental health promotion in secondary schools. *Aust N Z J Psychiatry* 2000;**34**:586–93. https://doi.org/10.1080/j.1440–1614.2000.00718.x

123. Ransford CR, Greenberg MT, Domitrovich CE, Small M, Jacobson L. The role of teachers' psychological experiences and perceptions of curriculum supports on the implementation of a social and emotional learning curriculum. *School Psych Rev* 2009;**38**:510–32.

124. Reid JB, Eddy JM. *Preventive Efforts During the Elementary School Years: The Linking the Interests of Families and Teachers Project*. Washington, DC: American Psychological Association; 2002.

125. Reid JB, Eddy JM, Fetrow RA, Stoolmiller M. Description and immediate impacts of a preventive intervention for conduct problems. *Am J Community Psychol* 1999;**27**:483–517. https://doi.org/10.1023/A:1022181111368

126. Rothwell H, Segrott J. Preventing alcohol misuse in young people aged 9–11 years through promoting family communication: an exploratory evaluation of the Kids, Adults Together (KAT) programme. *BMC Public Health* 2011;**11**:810. https://doi.org/10.1186/1471–2458–11–810

127. Segrott J, Rothwell H, Hewitt G, Playle R, Huang C, Murphy S, *et al.* Preventing alcohol misuse in young people: an exploratory cluster randomised controlled trial of the Kids, Adults Together (KAT) programme. *Public Health Res* 2015;**3**(15). https://doi.org/10.3310/phr03150

128. Smith EA, Swisher JD, Vicary JR, Bechtel LJ, Minner D, Henry KL, Palmer R. Evaluation of life skills training and infused-life skills training in a rural setting: outcomes at two years. *J Alcohol Drug Educ* 2004;**48**:51.

129. Snyder F, Flay B, Vuchinich S, Acock A, Washburn I, Beets M, Li KK. Impact of a social-emotional and character development program on school-level indicators of academic achievement, absenteeism, and disciplinary outcomes: a matched-pair, cluster randomized, controlled trial. *J Res Educ Eff* 2010;**3**:26–55. https://doi.org/10.1080/19345740903353436

130. Snyder FJ, Acock AC, Vuchinich S, Beets MW, Washburn IJ, Flay BR. Preventing negative behaviors among elementary-school students through enhancing students' social-emotional and character development. *Am J Health Promot* 2013;**28**:50–8. https://doi.org/10.4278/ajhp.120419-QUAN-207.2

131. Stoolmiller M, Eddy JM, Reid JB. Detecting and describing preventive intervention effects in a universal school-based randomized trial targeting delinquent and violent behavior. *J Consult Clin Psychol* 2000;**68**:296–306. https://doi.org/10.1037/0022-006X.68.2.296

132. Vicary JR, Smith EA, Swisher JD, Hopkins AM, Elek E, Bechtel LJ, Henry KL. Results of a 3-year study of two methods of delivery of life skills training. *Health Educ Behav* 2006;**33**:325–39. https://doi.org/10.1177/1090198105285020

133. Wales A. *Peaceful Panels: Using Comics to Teach Anti-bullying to Junior High School Students*. PhD thesis. Prescott Valley, AZ: Northcentral University; 2013.

134. Wang C, Couch L, Rodriguez GR, Lee C. The Bullying Literature Project: using children's literature to promote prosocial behavior and social-emotional outcomes among elementary school students. *Contemp Sch Psychol* 2015;**19**:320–9. https://doi.org/10.1007/s40688-015-0064-8

135. Zoller U, Weiss S. 'Hashish and Marijuana': an innovative, interdisciplinary drug education curricular program for high schools. *J Drug Educ* 1981;**11**:37–46. https://doi.org/10.2190/U18B-UQUU-ATM8-R66F

136. Wang C, Goldberg T. Using children's literature to decrease moral disengagement and victimization among elementary school students. *Psychol Sch* 2017;**54**:918–31. https://doi.org/10.1002/pits.22042

137. Torrente C, Johnston B, Starkey L, Seidman E, Shivshanker A, Weisenhorn N, *et al.* Improving the quality of school interactions and student well-being: impacts of one year of a school-based program in the Democratic Republic of the Congo. *Journal on Education in Emergences* 2015;**1**:48–91.

138. Aber JL, Tubbs C, Torrente C, Halpin PF, Johnston B, Starkey L, *et al.* Promoting children's learning and development in conflict-affected countries: testing change process in the Democratic Republic of the Congo. *Dev Psychopathol* 2017;**29**:53–67. https://doi.org/10.1017/S0954579416001139

139. Espelage DL, Low S, Polanin JR, Brown EC. The impact of a middle school program to reduce aggression, victimization, and sexual violence. *J Adolesc Health* 2013;**53**:180–6. https://doi.org/10.1016/j.jadohealth.2013.02.021

140. Espelage DL, Low S, Polanin JR, Brown EC. Clinical trial of Second Step middle-school program: impact on aggression and victimization. *J Appl Dev Psychol* 2015;**37**:52–63. https://doi.org/10.1016/j.appdev.2014.11.007

141. Espelage D, Low S, Van Ryzin M, Polanin J. Clinical trial of Second Step middle school program: impact on bullying, cyberbullying, homophobic teasing, and sexual harassment perpetration. *School Psych Rev* 2015;**44**:464–79. https://doi.org/10.17105/spr-15-0052.1

142. Coleman JC. The Focal Theory of Adolescence: A Psychological Perspective. In Hurrelmann K, Engel U, editors. *The Social World of Adolescents: International Perspectives*. Oxford: Walter De Gruyter; 1989. pp. 43–56.

143. Kolb DA. *Experiential Learning: Experience as the Source of Learning and Development, Vol. 1*. Englewood Cliffs, NJ: Prentice-Hall; 1984.

144. Winslade J, Monk G. *Narrative Mediation: A New Approach to Conflict Resolution*. San Francisco, CA: Jossey-Bass; 2000.

145. Hilgard ER. A perspective on the relationship between learning theory and educational practices. In Hilgard ER, Richey HG, editors. *Theories of Learning and Instruction*. Chicago, IL: University of Chicago Press; 1964. pp. 402–15.

146. Arhar JM. Interdisciplinary Teaming and the Social Bonding of Middle Level Students. In Irvin JL, editor. *Transforming Middle Level Education: Perspectives and Possibilities*. Needham Heights, MA: Allyn & Bacon; 1992. pp. 139–61.

147. Stuhr PL. Multicultural art education and social reconstruction. *Studies in Art Education* 1994;**35**:171–8. https://doi.org/10.2307/1320218

148. Darder A, Baltodano M, Torres RD, editors. *The Critical Pedagogy Reader*. New York, NY: Routledge; 2009.

149. Axelrod J. Collaborative for Academic, Social and Emotional Learning (CASEL). In Clauss-Ehlers CS, editor. *Encyclopedia of Cross-Cultural School Psychology*. Springer, Boston, MA; 2010. https://doi.org/10.1007/978-0-387-71799-9_77

150. Tancred T, Paparini S, Melendez-Torres GJ, Thomas J, Fletcher A, Campbell R, Bonell C. A systematic review and synthesis of theories of change of school-based interventions integrating health and academic education as a novel means of preventing violence and substance use among students. *Systematic Reviews* 2018;**7**:190. https://doi.org/10.1186/s13643-018-0862-y

151. US Government. *No Child Left Behind Act of 2001*. Washington, DC: US Government Printing Office; 2001. URL: http://gpo.gov/fdsys/pkg/PLAW-107publ110/html/PLAW-107publ110.htm (accessed 15 January 2017).

152. Bonell C, Fletcher A, Jamal F, Aveyard P, Markham W. Where next with theory and research on how the school environment influences young people's substance use? *Health Place* 2016;**40**:91–7. https://doi.org/10.1016/j.healthplace.2016.05.006

153. Welsh Assembly Government. *Personal and Social Education Framework for 7 to 19-year-olds in Wales*. Cardiff: Department for Children, Education, Lifelong Learning and Skills; 2008.

154. McWhirter J, Boddington N, Barksfield J. *Understanding Personal, Social, Health and Economic Education in Secondary Schools*. London: SAGE Publications; 2017. https://doi.org/10.4135/9781473983564

155. Melendez-Torres GJ, Tancred T, Fletcher A, Campbell R, Thomas J, Bonell C. Integration of academic and health education for the prevention of violence in young people: systematic review, narrative synthesis and intervention components analysis. *BMJ Open* 2018;**8**:e020793. https://doi.org/10.1136/bmjopen-2017-020793

156. Melendez-Torres GJ, Tancred T, Fletcher A, Thomas J, Campbell R, Bonell C. Does integrated academic and health education prevent substance use? Systematic review and meta-analyses. *Child* 2018; **44**:516–30. https://doi.org/10.1111/cch.12558

157. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;**343**:d5928. https://doi.org/10.1136/bmj.d5928

158. Campbell R, Pound P, Morgan M, Daker-White G, Britten N, Pill R, *et al.* Evaluating meta-ethnography: systematic analysis and synthesis of qualitative research. *Health Technol Assess* 2011;**15**(43). https://doi.org/10.3310/hta15430

159. Cousins LH. Toward a sociocultural context for understanding violence and disruption in black urban schools and communities. *J Sociol Soc Welf* 1997;**24**:41–63.

160. Benson PL, Scales PC, Hamilton SF, Sesma A. Positive Youth Development: Theory, Research, and Applications. In Lerner RM, Damon W, editors. *Handbook of Child Psychology*. Hoboken, NJ: John Wiley & Sons Inc; 2006. pp. 894–941. https://doi.org/10.1002/9780470147658.chpsy0116

161. Pawson R, Tilley N. *Realistic Evaluation*. London: SAGE Publications; 1997.

162. Haegerich TM, Metz E. The Social and Character Development Research Program: development, goals, and opportunities. *J Char Educ* 2009;**7**:1.

163. Michie S, Fixsen D, Grimshaw JM, Eccles MP. Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implement Sci* 2009;**4**:40. https://doi.org/10.1186/1748–5908–4–40

164. Bonell C, Fletcher A, Morton M, Lorenc T, Moore L. Realist randomised controlled trials: a new approach to evaluating complex public health interventions. *Soc Sci Med* 2012;**75**:2299–306. https://doi.org/10.1016/j.socscimed.2012.08.032

165. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998;**88**:216–22. https://doi.org/10.2105/AJPH.88.2.216

166. Fixsen DL, Naoom SF, Blase KA, Friedman RM. *Implementation Research: A Synthesis of the Literature*. Florida, FL: University of South Florida; 2005.

167. Langford R, Campbell R, Magnus D, Bonell CP, Murphy SM, Waters E, *et al.* The WHO Health Promoting School framework for improving the health and well-being of students and staff. *Cochrane Database Syst Rev* 2014;**4**:CD008958. https://doi.org/10.1002/14651858.CD008958.pub2

168. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. *Implement Sci* 2013;**8**:139. https://doi.org/10.1186/1748–5908–8–139

169. Chilton R, Pearson M, Anderson R. Health promotion in schools: a scoping review of systematic reviews. *Health Educ* 2015;**115**:357–76. https://doi.org/10.1108/HE-03-2014-0033

170. Rimm-Kaufman SE, Hulleman CS. Social and Emotional Learning in Elementary School Settings: Identifying Mechanisms that Matter. In Durlak J, Weissberg R, editors. *The Handbook of Social and Emotional Learning: Research and Practice*. New York, NY: The Guilford Press; 2015. pp. 151–66.

171. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychol* 2008;**41**:327–50. https://doi.org/10.1007/s10464–008–9165–0

172. Bandura A. Social learning theory of aggression. *J Commun* 1978;**28**:12–29. https://doi.org/10.1111/j.1460-2466.1978.tb01621.x

173. Fullan M. *The New Meaning of Educational Change*. New York, NY: Teachers College Press; 2007.

174. Kolb AY, Kolb DA. Experiential Learning Theory. In Seel NM, editor. *Encyclopedia of the Sciences of Learning*. Boston, MA: Springer; 2012. pp. 1215–9. https://doi.org/10.1007/978-1-4419-1428-6_227

175. Hunter M. *Instructional Theory Into Practice*. Blacksburg, VA: Virginia Polytechnic Institute and State University; 1984.

176. Tanford S, Penrod S. Social Influence Model: a formal integration of research on majority and minority influence processes. *Psychol Bull* 1984;**95**:189–225. https://doi.org/10.1037/0033-2909.95.2.189

177. Hawkins JD, Weis JG. The social development model: an integrated approach to delinquency prevention. *J Prim Prev* 1985;**6**:73–97. https://doi.org/10.1007/BF01325432

178. Dishion TJ, Duncan TE, Eddy JM, Fagot BI, Fetrow R. The world of parents and peers: coercive exchanges and children's social adaptation. *Soc Dev* 1994;**3**:255–68. https://doi.org/10.1111/j.1467-9507.1994.tb00044.x

179. Purkey WW. *Self Concept and School Achievement*. New Jersey, NJ: Prentice-Hall; 1970.

180. Flay BR, Snyder F, Petraitis J. The Theory of Triadic Influence. In DiClemente RJ, Kegler MC, Crosby RA, editors. *Emerging Theories in Health Promotion Practice and Research*. 2nd edn. New York, NY: Jossey-Bass; 2009. pp. 451–510.

181. Simons RL, Conger RD, Whitbeck LB. A multistage social learning model of the influences of family and peers upon adolescent substance abuse. *J Drug Issues* 1988;**18**:293–315. https://doi.org/10.1177/002204268801800301

182. Rogers EM. *Diffusion of Innovations*. 4th edn. New York, NY: Simon and Schuster; 2010.

183. Sheldon K, Frederickson B, Rathunde K, Csikszentmihalyi M, Haidt J. *Positive Psychology Manifesto*. URL: www.positivepsychology.org/akumamanifesto.htm (accessed 1 May 2016).

184. DuBois DL, Flay BR, Fagen MC. Self-esteem Enhancement Theory: Promoting Health Across the Lifespan. In DiClemente RJ, Crosby RA, Kegler MC, editors. *Emerging Theories in Health Promotion Practice and Research*. 2nd edn. San Francisco, CA: Jossey-Bass; 2009. pp. 97–130.

185. Moos RH. Conceptualizations of human environments. *Am Psychol* 1973;**28**:652. https://doi.org/10.1037/h0035722

186. Weiner BJ, Lewis MA, Linnan LA. Using organization theory to understand the determinants of effective implementation of worksite health promotion programs. *Health Educ Res* 2009;**24**:292–305. https://doi.org/10.1093/her/cyn019

187. Crick NR, Dodge KA. Social information-processing mechanisms in reactive and proactive aggression. *Child Dev* 1996;**67**:993–1002. https://doi.org/10.2307/1131875

188. Huesmann LR, Guerra NG. Children's normative beliefs about aggression and aggressive behavior. *J Pers Soc Psychol* 1997;**72**:408–19. https://doi.org/10.1037/0022-3514.72.2.408

189. Alonso A. A development theory of psychodynamic supervision. *Clin Superv* 1983;**1**:23–36. https://doi.org/10.1300/J001v01n03_04

190. Bandura A. Cognitive Social Learning Theory. In Vasta R, editor. *Annals of Child Development. Vol. 6*. Greenwich, CT: JAI Press; 1989. pp. 1–60.

191. Bowlby J. Attachment theory and its therapeutic implications. *Adolesc Psychiatry* 1978;**6**:5–33.

192. Klein GS. *Psychoanalytic Theory: An Exploration of Essentials*. Madison, CT: Intl Universities Pr Inc; 1976.

193. Mayer JD, Salovey P, Caruso DR, Sitarenios G. Emotional intelligence as a standard intelligence. *Emotion* 2001;**1**:232–42. https://doi.org/10.1037/1528-3542.1.3.232

194. Catalano RF, Hawkins J. The Social Development Model: A Theory of Antisocial Behavior. In Hawkins J, editor. *Delinquency and Crime*. New York, NY: Cambridge University Press; 1996. pp. 149–97.

195. Collaborative for Academic, Social, and Emotional Learning (CASEL). *CASEL Program Guides: Effective Social and Emotional Learning Programs*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning; 2007.

196. Espelage DL, Swearer SM. A social-ecological model for bullying prevention and intervention. In Jimerson SR, Swearer SM, Espelage DL, editors. *International Handbook of Bullying in Schools: An International Perspective*. New York, NY: Routledge; 2010. pp. 257–73.

197. World Health Organization (WHO). *The Status of School Health*. Geneva: WHO; 1996.

198. Green LW, Kreuter MW. *Health Program Planning: An Educational and Ecological Approach*. 3rd edn. California, CA: Mayfield Publishing Co.; 2005.

199. Shumaker SA, Brownell A. Toward a theory of social support: closing conceptual gaps. *J Soc Issues* 1984;**40**:11–36. https://doi.org/10.1111/j.1540-4560.1984.tb01105.x

200. Szapocznik J, Coatsworth JD. An Ecodevelopmental Framework for Organizing the Influences on Drug Abuse: A Developmental Model of Risk and Protection. In Glantz M, Hartel C, editors. *Drug Abuse: Origins and Interventions*. Washington, DC: American Psychological Association; 1999. pp. 331–66. https://doi.org/10.1037/10341-014

201. Hirschi T. Self-control and Crime. In Baumeister RF, Vohs KD, editors. *Handbook of Self-regulation: Research, Theory, and Applications*. New York, NY: Guilford Press; 2004. pp. 537–52.

202. Matsueda RL, Heimer K. Race, family structure, and delinquency: a test of differential association and social control theories. *Am Sociol Rev* 1987;**52**:826–40. https://doi.org/10.2307/2095837

# Appendix 1 Protocol deviations and clarifications

## About this appendix

This appendix provides a full indication of any deviations from or clarifications to our original protocol[1] as we progressed through the review.

**TABLE 12** Deviations and clarifications from the protocol and/or proposal with rationale for change

| Deviation or clarification | From proposal or protocol? | Date implemented | Change | Rationale |
|---|---|---|---|---|
| Deviation | Proposal | 18 November 2015 | Will not search EMBASE, CINAHL, CAB Health, CISDOC, HMIC or PAIS | EMBASE strongly overlaps with MEDLINE and is unlikely to yield additional studies of school health promotion. CISDOC is focused on occupational safety and unlikely to yield relevant studies. PAIS overlaps considerably with other social policy databases searched and is unlikely to yield additional relevant studies. CINAHL and HMIC are primarily focused on health care and unlikely to yield relevant studies |
| Deviation | Proposal | 18 November 2015 | Search the following (websites rather than bibliographic databases): Campbell Library, OpenGrey (System for Information on Grey Literature in Europe), Database of Educational Research, and International Clinical Trials Registry Platform | These are websites, not conventional bibliographic databases |
| Deviation | Proposal | 20 November 2015 | The interventions on which evidence is to be synthesised have been redefined slightly. Rather than focusing on education addressing violence and substance use delivered in 'mainstream' school lessons (i.e. not PSHE or its international equivalents), we will now focus on interventions that integrate education addressing violence and substance use with academic education (literacy, numeracy, study skills or specific academic subjects) | This is because we are interested in interventions that aim to integrate health and academic learning regardless of in which particular lesson in the curriculum they are delivered. It would be scientifically indefensible, for example, if we included an intervention integrating maths and alcohol education delivered in maths lessons but excluded a very similar intervention that was delivered in PSHE |
| Deviation | Proposal | 16 December 2015 | Have not searched Sociological Abstracts | Technical problems with the database timing out and downloading references. Given that three other social science databases have been searched, CS thought this unproblematic |
| Deviation | Proposal and protocol | 16 December 2015 | Only searched ASSIA on controlled vocabulary | Technical problems getting the free text search to run |
| Deviation | Proposal and protocol | 21 December 2015 | Only search Dissertation Abstracts/ Index to Theses from 2010 | Only recent international theses are likely to be retrievable |

*continued*

**TABLE 12** Deviations and clarifications from the protocol and/or proposal with rationale for change (*continued*)

| Deviation or clarification | From proposal or protocol? | Date implemented | Change | Rationale |
|---|---|---|---|---|
| Deviation | Proposal | 11 January 2016 | Inclusion criteria modified slightly: interventions integrating health and academic biology education no longer excluded | To exclude interventions integrating health education with biology but not other academic subjects would be arbitrary. Better to include and if necessary to examine as a subgroup |
| Clarification | Proposal and protocol | 12 January 2016 | Excluding studies where intervention targets individuals on the basis of individual health outcome | We are interested in curricula delivered in mainstream schools to general populations. This is implicit in the original proposal and protocol but we needed to make it explicit to guide screening |
| Clarification | Proposal and protocol | June 2016 | Our focus on violence had to be refined slightly. Included reports had to be explicit about reducing physical violence and/or acts of aggression. Where studies referred to non-physical violence, these were included only if the study also addressed another outcome of interest, such as substance use | This change was to ensure we were only including studies with relevance to our protocol |
| Deviation | Proposal and protocol | 1 January 2017 | Shifting forward the PPI with key stakeholders and young people from October 2016 | We felt it would be more useful (given that none of the included interventions was from a UK context with the exception of one) to have preliminary results from both our theoretical and process data syntheses to use to generate the most meaningful reflections and hypotheses to speak about with participants |
| Deviation | Proposal | 1 February 2017 | Shifting forward timeline for report and publication preparations | This shift was done to accommodate TT's maternity leave |

ASSIA, Applied Social Sciences Index and Abstracts; CINAHL, Cumulative Index to Nursing and Allied Health Literature; HMIC, Health Management Information Consortium; PAIS, Public Affairs Information Service; PPI, patient and public involvement.

# **Appendix 2** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist

| Section/topic | # | Checklist item | Reported where |
|---|---|---|---|
| **TITLE** | | | |
| Title | 1 | Identify the report as a systematic review, meta-analysis, or both | v, 5 |
| **ABSTRACT** | | | |
| Structured summary | 2 | Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number | v–vi |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of what is already known | 1–3 |
| Objectives | 4 | Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS) | 3 |
| **METHODS** | | | |
| Protocol and registration | 5 | Indicate if a review protocol exists, if and where it can be accessed (e.g. Web address), and, if available, provide registration information including registration number | 5 |
| Eligibility criteria | 6 | Specify study characteristics (e.g. PICOS, length of follow-up) and report characteristics (e.g. years considered, language, publication status) used as criteria for eligibility, giving rationale | 5–7 |
| Information sources | 7 | Describe all information sources (e.g. databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched | 7–9 |
| Search | 8 | Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated | 119–122 |
| Study selection | 9 | State the process for selecting studies (i.e. screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis) | 10 |
| Data collection process | 10 | Describe method of data extraction from reports (e.g. piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators | 10–11 |
| Data items | 11 | List and define all variables for which data were sought (e.g. PICOS, funding sources) and any assumptions and simplifications made | 10–11 |
| Risk of bias in individual studies | 12 | Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis | 11–12 |
| Summary measures | 13 | State the principal summary measures (e.g. risk ratio, difference in means) | 14–15 |
| Synthesis of results | 14 | Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g. $I^2$) for each meta-analysis | 14–15 |
| Risk of bias across studies | 15 | Specify any assessment of risk of bias that may affect the cumulative evidence (e.g. publication bias, selective reporting within studies) | 15 |
| Additional analyses | 16 | Describe methods of additional analyses (e.g. sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified | 15 |

| Section/topic | # | Checklist item | Reported where |
|---|---|---|---|
| ***RESULTS*** | | | |
| Study selection | 17 | Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram | 19–20 |
| Study characteristics | 18 | For each study, present characteristics for which data were extracted (e.g. study size, PICOS, follow-up period) and provide the citations | 20–33 |
| Risk of bias within studies | 19 | Present data on risk of bias of each study and, if available, any outcome level assessment (see Item 12) | 35, 45, 56–58 |
| Results of individual studies | 20 | For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and CIs, ideally with a forest plot | 67–70, 73–76, 87 |
| Synthesis of results | 21 | Present results of each meta-analysis done, including CIs and measures of consistency | 67–70, 73–76, 87 |
| Risk of bias across studies | 22 | Present results of any assessment of risk of bias across studies (see Item 15) | 58–60 |
| Additional analysis | 23 | Give results of additional analyses, if done (e.g. sensitivity or subgroup analyses, meta-regression [see Item 16]) | Not applicable |
| ***DISCUSSION*** | | | |
| Summary of evidence | 24 | Summarise the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g. healthcare providers, users, and policy-makers) | 93–96 |
| Limitations | 25 | Discuss limitations at study and outcome level (e.g. risk of bias), and at review-level (e.g. incomplete retrieval of identified research, reporting bias) | 91–93 |
| Conclusions | 26 | Provide a general interpretation of the results in the context of other evidence, and implications for future research | 96 |
| ***FUNDING*** | | | |
| Funding | 27 | Describe sources of funding for the systematic review and other support (e.g. supply of data); role of funders for the systematic review | vi, xxiv |

# Appendix 3 Full search terms and strategies for bibliographic databases

## About this appendix

This appendix provides an example of all string searches used for PsycINFO (via Ovid). It was searched on the 18 November 2015 and returned 22,451 references. Other database searches were based on this strategy.

## Search strategy

1. ((substance? or drug? or drinking or alcohol* or solvent?) adj1 ("use" or abus* or misuse*)).ti,ab.

2. ((substance? or drug? or drinking or alcohol* or solvent?) adj1 (usage or intake or using or taking or behavio* or user?)).ti,ab.

3. (drinking adj1 (alcohol* or behavio*)).ti,ab.

4. Alcohol.ti,ab.

5. (smoke or smoking or tobacco or cigarette? or smoker? or cannabis or marijuana).ti,ab.

6. (aggression or aggressive or bully* or delinquen* or "conduct problem*" or "conduct disorder?" or "antisocial" or "anti social" or violence or violent or (volatile adj behavio*) or victimi* or hostile or hostility or perpetrat*).ti,ab.

7. (Externalising or externalizing).ti,ab.

8. emotion*.ti,ab.

9. PSHE.ti,ab.

10. ("Health literacy" or "health education" or "health promotion" or "preventive health" or "primary prevention" or "health information" or "promoting health" or "health promoting" or "health promotion" or "health maintenance").ti,ab.

11. "Public health".ti,ab.

12. ("wellbeing" or "well being").ti,ab.

13. "mental health".ti,ab.

14. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13.

15. ((curric* or lesson? or classes or classroom? or subject? or intervention? or program* or education or initiative? or learn or learning or teach or teaching or outcome* or attainment or achievement or assessment or effect* or impact* or score? or scoring* or skill? or knowledge or competen* or performance) adj3 (Academic or academically or Scholastic or scholar* or Mainstream or "main stream")).ti,ab.

16. ((curric* or lesson? or classes or classroom? or subject? or learn or learning or teach or teaching or attainment or achievement or assessment or score? or scoring* or skill? or knowledge or competen* or performance) adj3 School?).ti,ab.

17. ((intervention? or program* or initiative? or effect* or impact* or education) adj1 School?).ti,ab.

18. (class adj1 (Academic or academically or Scholastic or scholar* or School? or Mainstream or "main stream")).ti,ab.

19. ((curric* or lesson? or classes or classroom? or subject? or education or learn or learning or teach or teaching or attainment or achievement or score? or scoring* or skill? or knowledge or competen*) adj3 (study or core or generic)).ti,ab.

20. (class adj1 (study or core or generic)).ti,ab.

21. ((curric* or lesson? or classes or classroom? or subject? or attainment or achievement or assessment or score? or scoring* or competenc* or performance) adj3 ((Education not ("patient education" or "continuing education")) or educational)).ti,ab.

22. (class adj1 ((Education not ("patient education" or "continuing education")) or educational)).ti,ab.

23. (outcome* adj1 (education or educational)).ti,ab.

24. ((curric* or lesson? or classroom? or classes or subject? or intervention? or program* or initiative? or education or teach* or outcome* or attainment or achievement or assessment or effect* or impact* or score? or scoring* or skill? or knowledge or competen* or performance) adj3 (learn or learning)).ti,ab.

25. (class adj1 (learn or learning)).ti,ab.

26. ((curric* or lesson? or classes or classroom or class or subject? or education or teach* or learning or teach or teaching or learn or attainment or achievement or assessment or score? or scoring* or skill? or knowledge or competen* or performance) adj3 (art or arts or math* or science? or humanities or chemistry or physics or language* or geography or (history not ("medical history" or "health history" or "familial history" or "family history")) or numeracy or (literacy not "health literacy") or grammar or grammer or reading or writing)).ti,ab.

27. (((curric* or lesson? or classroom or classes or subject? or skill?) adj3 literature) or "literature class").ti,ab.

28. ("Education reform" or "Instructional support" or "School reform" or "Classroom organi*" or (Commit* adj3 (school or education or learning)) or (Engag* adj3 (school or education or learning)) or "Character development" or "Whole school" or "School level" or "School wide" or schoolwide).ti,ab.

29. ((Comprehensive adj3 school) and (intervention? or program* or initiative? or outcome* or effect* or impact*)).ti,ab.

30. ((Integrat* or Combin* or Infuse or infused or infusion or sustainable) adj3 (curric* or lesson? or classes or classroom or syllabus or subject? or education or learn or learning or teach or teaching)).ti,ab.

31. (((Integrat* or Combin* or Infuse or infused or infusion or sustainable) adj3 (intervention* or program* or initiative*)) and school?).ti,ab.

32. ((school or education or core or generic or teaching or learning) adj3 syllabus).ti,ab.

33. 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 24 or 25 or 26 or 27 or 28 or 29 or 30 or 32.

34. (child* or schoolchild* or youth* or "young people*" or "young person" or teen* or adolescen* or juvenile* or preadolescen* or boy? or girl?).ti,ab.

35. (curric* or lesson? or classes or classroom? or subject? or school? or syllabus or "junior high" or "senior high" or "junior education" or "elementary education" or "primary education").ti,ab.

36. 34 and 35.

37. ("secondary school?" or "primary school?" or "comprehensive school?" or "school education" or "high school?" or "grammar school?" or "private school?" or "public school?" or "mainstream school*" or "compulsory education" or "statutory education" or "middle school?" or "junior school?" or "senior school?" or "primary education" or "secondary education" or "elementary school?" or "elementary education" or "mainstream education" or "compulsory school*" or "statutory school*" or "sixth form college?" or "post-16 education" or "junior high" or "senior high" or "reception class" or "post primary").ti,ab.

38. ((school? or junior? or elementary or senior? or primary or "sixth form" or grade) adj10 student?).ti,ab.

39. pupil?.ti,ab.

40. 36 or 37 or 38.

41. (University or universities or freshmen or sophomore? or "higher education" or "tertiary education" or ((registrar* or workplace? or clinical or medical or nursing or nurse? or doctor? or continuing or adult? or patient?) adj1 (education or educating or profession* or student?)) or "professional education").ti.

42. 40 not 41.

43. 14 and 33 and 42.

44. "Elementary School Students"/ or "Intermediate School Students"/ or "Primary School Students"/ or "Middle School Students"/ or "High School Students"/ or "Junior High School Students"/ or "Kindergarten Students"/ or "High School Education"/ or "Middle School Education"/ or "Secondary Education"/ or "Junior High Schools"/ or "High Schools"/ or "Schools"/ or "Elementary Schools"/ or "Middle Schools"/.

45. "Drug Abuse Prevention"/ or "Health Education"/ or "Drug Education"/ or "Health Promotion"/ or "Public Health"/ or "Health Promotion"/ or "Preventive Medicine"/ or Health behaviour/ or Harm reduction/ or Health literacy/ or exp Health screening/ or Primary Mental health prevention/ or Prevention/ or Public health/ or Lifestyle changes/ or Lifestyle/ or Health literacy/.

46. "Tobacco Smoking"/ or "Smoking Cessation"/ or "Marijuana Usage"/ or "Drinking Behavior"/ or "Social Drinking"/ or "Binge Drinking"/ or "Underage Drinking"/ or "Alcohol Abuse"/ or "Alcohol Drinking Patterns"/ or "Alcohol Intoxication"/ or "Alcoholism"/ or "Heroin Addiction"/ or "Drug Addiction"/ or "Drug Dependency"/ or "Drug Usage"/ or "Inhalant Abuse"/ or "Drug Abuse"/ or "Glue Sniffing"/ or "Predelinquent Youth"/ or "Cyberbullying"/ or "School Violence"/ or "Teasing"/ or "Juvenile Delinquency"/ or "Physical Abuse"/ or "Verbal Abuse"/ or "Violence"/ or "Harassment"/ or "Antisocial Behavior"/ or "Bullying"/ or "Perpetrators"/ or "Threat"/ or "Victimization"/ or "Relational Aggression"/ or "Aggressive Behavior"/ or "Behavior Problems"/ or "Behavior Disorders"/ or "Conduct Disorder"/ or "Drug Education"/ or "Drug Abuse Prevention"/ or "Harm Reduction"/.

47. emotions/ or emotional development/.

48. emotional adjustment/ or emotional disturbances/ or emotional control/.

49. mental health/ or primary mental health prevention/ or well being/.

50. "Curriculum"/ or "Curriculum Based Assessment"/ or "Curriculum Development"/ or "School Learning"/ or "Classroom Environment"/ or "Academic Environment"/ or "Teacher Effectiveness"/ or "Teacher Effectiveness Evaluation"/ or "Educational Program Evaluation"/ or "Course Evaluation"/ or "learning environment"/.

51. 14 or 45 or 46 or 47 or 48 or 49.

52. 33 or 50.

53. 42 or 44.

54. 51 and 52 and 53.

# Appendix 4 Expert searches

## About this appendix

This appendix indicates which experts were contacted for this review, how they were contacted and the list of references that resulted through this communication.

## Experts contacted 19 October–20 November 2015

Larry Aber (New York University).

Rob Anderson (University of Exeter).

Lyndal Bond (Centre of Excellence in Intervention and Prevention Science).

Marc Brackett (Yale University).

Carolyn Brown (Fordham University).

Helen Butler (Australian Catholic University).

Rona Campbell (University of Bristol).

Simon Denny (University of Auckland).

Rutger Engels (Trimbos Institute).

Brian Flay (State University of Oregon).

David Foxcroft (Oxford Brookes University).

Mark Greenberg (Pennsylvania State University).

Joe Hayman (PSHE Association).

Denise Haynie (National Institutes of Health).

Marion Henderson (University of Glasgow).

Wendy Hoglund (University of Alberta).

Neil Humphrey (University of Manchester).

Charles Irwin (University of California, San Francisco).

Stephanie Jones (University of Harvard).

Kelli Komro (Emory University).

Janis Kupersmidt (Innovation of Research and Training Inc.).

Beki Langford (University of Bristol).

Laurence Moore (University of Glasgow).

Vikram Patel (London School of Hygiene & Tropical Medicine).

George Patton (University of Melbourne).

Susan Rivers (Yale University).

Tom Roderick (Morningside Centre for Teaching Social Responsibility).

Barri Rosenbluth (Safe Place).

David Ross (World Health Organization).

Peter Salovey (Yale University).

John Santelli (University of Columbia).

Tracy Scull (Innovation of Research and Training Inc.).

Bruce Simons-Morton (National Institutes of Health).

Michael Telch (University of Texas).

Russell Viner (University College London).

Michael Wigelsworth (University of Manchester).

Danny Wight (University of Glasgow).

Lara Zwarun (University of Missouri).

## E-mails sent to experts

### *First e-mail*

Dear all

I am a researcher from University College London's EPPI-Centre (http://eppi.ioe.ac.uk/cms/). We are currently undertaking a systematic review entitled **"How can mainstream subject lessons in schools help prevent substance misuse and violence, and reduce health inequalities among young people? Systematic review and evidence synthesis**".

Please find the protocol here:

http://www.crd.york.ac.uk/prospero/display_record.asp?ID=CRD42015026464

I am writing to you today as an expert in the area of health promotion for children and young people. I would like you to inform me of **any research of which you are aware that may be relevant to this review**. The table below summarises the types of study in which we are interested.

| Participation | Children and young people aged 4-18 years attending schools |
|---|---|
| Intervention | School-based health curriculum interventions delivered within 'mainstream' subject lessons (i.e. not personal/health education or international equivalents). |
| Outcome | Smoking<br><br>Alcohol use<br><br>Legal or illegal drug use<br><br>Physical violence |
| Study design | Process evaluations<br><br>Randomised controlled trial |

At the end of this email is a list of relevant studies of which we are already aware.

Ideally I would be very grateful if you could let me know of additional relevant studies by email by 1 November 2015. However, if this is not possible, please could you indicate if and by when you would be able to respond?

If there are other experts you would recommend we contact, please do let me know.

If you have any questions, please do not hesitate to get in touch.

Thank you in advance for your assistance on this matter.

Kind regards,

Chris Bonell

**Already known studies**

Bier, M.C., L. Zwarun, and V. Fehrmann Warren, *Getting universal primary tobacco use prevention into priority area schools: a media literacy approach.* Health Promotion Practice, 2011. **12**(6 Suppl 2): p. 152S-8S.

Bond, L. and H. Butler, *The Gatehouse Project: a multi-level integrated approach to promoting wellbeing in schools. Draft chapter for NICE publication*2008.

Bond, L., et al., *Building capacity for system-level change in schools: lessons from the Gatehouse project.* Health Education and Behavior, 2001. **28**(3): p. 368-383.

Bond, L., et al., *The Gatehouse Project: can a multi-level school intervention affect emotional well-being and health risk behaviours?* Journal of Epidemiology and Community Health, 2004. **58**(12): p. 997-1003.

Bond, L., et al., *Long-term impact of the Gatehouse Project on cannabis use of 16-year-olds in Australia.* Journal of School Health, 2004. **74**(1): p. 23-30.

Flay, B.R., et al., *Effects of 2 prevention programs on high-risk behaviors among African American youth: a randomized trial.* Archives of Pediatrics & Adolescent Medicine, 2004. **158**(4): p. 377-384.

Hagelskamp, C., et al., *Improving classroom quality with the RULER Approach to Social and Emotional Learning: proximal and distal outcomes.* American Journal of Community Psychology, 2013. **51**(3-4): p. 530-43.

Jones, S.M., J.L. Brown, and A. J.L., *Two-year impacts of a universal school-based social-emotional and literacy intervention: an experiment in translational developmental research.* Child Dev, 2011. **82**(2): p. 533-54.

Jones, S.M., et al., *A School-Randomized Clinical Trial of an Integrated Social–Emotional Learning and Literacy Intervention: Impacts After 1 School Year.* Journal of Consulting and Clinical Psychology, 2010. **78**(6): p. 829-842.

Kupersmidt, J.B., T.M. Scull, and J.W. Benson, *Improving media message interpretation processing skills to promote healthy decision making about substance use: the effects of the middle school media ready curriculum.* Journal of Health Communication, 2012. **17**(5): p. 546-63.

Moskowitz, J.M., et al., *An experimental evaluation of a drug education course.* Journal of Drug Education, 1984. **14**(1): p. 9-22.

Patton, G., et al., *Promoting social inclusion in schools: group-randomized trial of effects on student health risk behaviour and well-being.* Am J Public Health, 2006. **96**(9): p. 1582-1587.

Patton, G.C., et al., *The Gatehouse Project: a systematic approach to mental health promotion in secondary schools.* Australian and New Zealand Journal of Psychiatry, 2000. **34**: p. 586-593.

Patton, G., et al., *Changing schools, changing health? Design and implementation of the Gatehouse Project.* Journal of Adolescent Health, 2003. **33**: p. 231-239.

Rican, P., K. Ondrova, and J. Svatos, *The effect of a short, intensive intervention upon bullying in four classes in a Czech town.* Annals of the New York Academy of Sciences, 1996. **794**: p. 399-400.

Scaggs, L.S., *A substance abuse awareness prevention program: knowledge, attitudes and behaviours. Unpublished Ph.D. Dissertation*1985: Ohio State University.

Simons-Morton, B., et al., *Impact analysis and mediation of outcomes: the Going Places program.* Health Education & Behavior, 2005. **32**(2): p. 227-41.

Simons-Morton, B., et al., *The effects of the Going Places Program on early adolescent substance use and antisocial behavior.* Prevention Science, 2005. **6**(3): p. 187-97.

Telch, M.J., et al., *Social influences approach to smoking prevention: The effects of videotape delivery with and without same-age peer leader participation.* Addictive Behaviors, 1990. **15**: p. 21-8.

Whitaker, D.J., et al., *Expect respect: a school-based intervention to promote awareness and effective responses to bullying and sexual harassment*, in *Bullying in American Schools: A Social-ecological Perspective on Prevention and Intervention*, D.L. Espelage and S.W. Swearer, Editors. 2004, Erlbaum: Mahwah, NJ. p. 327-350.

## *Follow-up email*

Dear all

I emailed you previously about a systematic review entitled *"How can mainstream subject lessons in schools help prevent substance misuse and violence, and reduce health inequalities among young people? Systematic review and evidence synthesis"*.

One of you pointed out that the distinction between delivering health education in personal/health education lessons and 'mainstream' subjects may not be relevant in all countries. Therefore, we have amended our protocol and inclusion criteria to make clearer what sorts of studies we are interested in: namely those which **aim to integrate learning about health and other academic subjects**.

So, for example, we would be interested in interventions which educate students about alcohol as well as maths in maths lessons by calculating units of alcohol consumed or the risks associated with alcohol consumption. But we would also now be interested in health education lessons that aim to educate about substance use but that also aim to improve academic learning skills, **irrespective of which part of the school curriculum these are delivered within**.

With this in mind, I would be very grateful if you could let me know of any relevant studies you know of by email by 10 November 2015. However, if this is not possible, please could you indicate if and by when you would be able to respond?

If you have any questions, please do not hesitate to get in touch.

Thank you in advance for your assistance on this matter.

Kind regards,

Chris Bonell

# List of studies generated from expert search

## Suggestions from Beki Langford

De Vries H, Dijk F, Wetzels J, Mudde A, Kremers S, Ariza C, *et al.* The European Smoking Prevention Framework Approach (ESFA): effects after 24 and 30 months. *Health Educ Res 2*006;**21**:116–32.

De Vries H, Mudde A, Kremers S, Wetzels J, Uiters E, Ariza C, *et al.* The European Smoking Prevention Framework Approach (ESFA): short-term effects. *Health Educ Res* 2003;**18**:649–63.

De Vries H, Mudde A, Leijs I, Charlton A, Vartiainen E, Buijs G, *et al.* The European Smoking Prevention Framework Approach (EFSA): an example of integral prevention. *Health Educ Res* 2003;**18**:611–26.

Holm K, Kremers S, De Vries H. Why do Danish adolescents take up smoking? *Eur J Public Health* 2003;**13**:67–74.

Vartiainen E, Pennanen M, Haukkala A, Dijk F, Lehtovuori R, De Vries H. The effects of a three-year smoking prevention programme in secondary schools in Helsinki. *Eur J Public Health* 2007;**17**:249–56.

## Suggestions from Brian Flay

Caring School Communities (URL: www.collaborativeclassroom.org/research-articles-and-papers).

PATHS (the module for grade 5 of which is believed to include study skills).

Positive Action (URL: www.positiveaction.net/) [see also bflay.net (URL: www.bflay.net)].

Second Step (related publications), which explicitly focuses on school connection and engagement with teachers.

Conduct Problems Prevention Group. Initial impact of the fast track prevention trial for conduct problems: I. The high-risk sample. *J Consult Clin Psych* 1999;**67**:631–47.

Conduct Problems Prevention Research Group. The effects of a multiyear universal social-emotional learning program: the role of student and school characteristics. *J Consult Clin Psych* 2010;**78**:156–168.

Crean HF, Johnson DB. Promoting Alternative THinking Strategies (PATHS) and elementary school aged children's aggression: results from a cluster randomized trial. *Am J Comm Psych* 2013;**52**:56–72.[97]

Curtis C, Norgate R. An evaluation of the Promoting Alternative Thinking Strategies curriculum at key stage 1. *Educ Psych Pract* 2007;**23**:33–44.

Greenberg MT, Kusche CA. Preventive intervention for school-aged deaf children: the PATHS Curriculum. *J Deaf Stud Deaf Educ* 1998;**3**:49–63.

Greenberg MT, Kusche CA, Cook ET, Quamma JP. Promoting emotional competence in school-aged children: the effects of the PATHS curriculum. *Dev Psychopath* 1995;**7**:117–36.

Kam C, Greenberg MT, Kusché CA. Sustained effects of the PATHS curriculum on the social and psychological adjustment of children in special education. *J Emot Behav Disord* 2004;**12**:66–78.

Kam C, Greenberg MT, Walls CT. Examining the role of implementation quality in school-based prevention using PATHS curriculum. *Prevention Science* 2003;**4**:55–63.

Little M, Berry V, Morpeth L, Blower S, Axford N, Taylor R, Bywater T, Lehtonen M. Tobin K. The impact of three evidence-based programmes delivered in public systems in Birmingham, UK. *Int J Confl Violence* 2012;**6**:260–72.

Malti T, Ribeaud D, Eisner MP. The effectiveness of two universal preventive interventions in reducing children's externalising behavior: a cluster randomized controlled trial. *J Clin Child Adolesc Psychol* 2011;**40**:677–92.

Malti T, Ribeaud D, Eisner M. Effectiveness of a universal school-based social competence program: the role of child characteristics and economic factors. *Int J Confl Violence* 2012;**6**:249–59.

Riggs NR, Greenberg MT, Kusché CA, Pentz MA. The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: effects of the PATHS curriculum. *Prevention Science* 2006;**7**:91–102.

Schonfeld DJ, Adams RE, Fredstrom BK, Weissberg RP, Gilman R, Voyce C, *et al.* Cluster-randomized trial demonstrating impact on academic achievement of elementary social-emotional learning. *School Psychology Quarterly* 2014.

Seifert R, Gouley K, Miller AL, Zabriski A. Implementation of the PATHS curriculum in an urban elementary school. *Early Educa Dev* 2004;**15**:471–86.

### *Suggestions from Tom Roderick*

Brown JL, Jones SM, LaRusso MD, Aber JL. Improving classroom quality: teacher influences and experimental impacts of the 4Rs program. *J Educ Psychol* 2010;**102**:153–67.[54]

### *Suggestions from Mark Greenberg*

Mark Greenberg shared the grade 5 curriculum of the PATHS programme and suggested that this year in particular integrates with literacy development. Possible publications including study of this grade include the following.

Crean HF, Johnson DB. Promoting Alternative Thinking Strategies (PATHS) and elementary school aged children's aggression: results from a cluster randomized trial. *Am J Commun Psychol* 2013;**52**:56–72.

Kam CM, Greenberg MT, Walls CT. Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prev Sci* 2003;**4**:55–63.

Kam C, Greenberg MT, Kusché CA. Sustained Effects of the PATHS® Curriculum on the Social and Psychological Adjustment of Children in Special Education. *J Emot Behav Disord* 2004;**12**:66–78.

Seifer R, Gouley K, Miller AL, Zakriski A. Implementation of the PATHS curriculum in an urban elementary school. *Early Educ Dev* 2004;**15**:471–85.

# Appendix 5 Data extraction and quality assessment for theory studies

## About this appendix

Below are the data extraction and quality assessment tools that were used by reviewers for reports on theory.

## Data extraction tool

Consider if the study has reported the following. Provide justifications for your response wherever possible.

1. Does the paper describe a theory of change (a description of the intended core components of an intervention and pathways via which these are intended to produce outcomes)?
2. Does the paper describe the rationale for the integration of academic and health curricula?

   - Yes
   - No

3. Which existing academic theories are used to inform the theory of change indicated?

   - None mentioned
   - Name(s) and author(s) of other theory/ies

4. Does this theory differ from others that have been reviewed thus far?

   - Yes
   - No

5. Any other comments (write here any points not covered above, e.g. your view of the novelty and importance of this paper).

### Quality assessment tool

Consider if the study has reported the following. Provide justifications for your response wherever possible.

1. Clarity

   - Are constructs clearly defined?
   - Is there a clear description of the intended pathways from intervention inputs to behaviour change?

2. Plausibility/feasibility

   - Is the theory logical (e.g. do the pathways from intervention inputs to behaviour change seem plausible)?
   - Is there a description of existing empirical evidence supporting the theory?

3. Testability

   - Do the authors refer in any way to empirical testing of the theory?

4. Ownership

- Was the theory developed with or by practitioner stakeholders (e.g. teachers)?
- Was the theory developed with or by community stakeholders (e.g. students)?

5. Generalisability

- Is the theory presented as one that is applicable to different contexts?
- Does the theory describe how the mechanisms might play out differently in different contexts?
- Do authors present empirical evidence of the generalisability of the theory?

# Appendix 6 Data extraction and quality assessment for process evaluations

## About this appendix

Below are the data extraction and quality assessment tools that were used by reviewers for reports on process.

## Data extraction tool

Consider if the study has reported the following. Provide justifications for your response wherever possible.

### *Intervention*

1. Intervention location

   - Details [country; area(s) of country]
   - Not stated

2. Description of intervention [the nature of the intervention(s)/components and how it was delivered to intervention group(s)]

   - Details (overall aim of the intervention: what knowledge and skills it aims to build, key components of the intervention, who delivered the intervention; deviations from original protocol, etc.)
   - Not stated

3. Description of the integration of academic and health curricula

   - Details (description of integration; rationale for integration)
   - Not stated

4. Frequency and duration of the intervention

   - Details
   - Not stated

5. Target population

   - Details (ages/grades; any mention of a universal vs. targeted intervention component; specific characteristics, i.e. from a low SES neighbourhood; how they were recruited)
   - Not stated

6. Provider (teachers, counsellors, nurses, external facilitators, etc.)

   - Details (qualifications, level of experience, how they were recruited)
   - Not stated

7. Provider organisation description

- Details (area of focus, years of experience doing similar work, aims of organisation, etc.)
- Not stated

8. How intervention was developed

- Details [background; founder(s); are there guiding theories and, if so, how are these explained in terms of theory of change data extraction, etc.]
- Not stated.

## *Evaluation*

9. Research questions or hypotheses

- Details
- Not stated

10. Evaluation examined (tick all which apply): feasibility/fidelity/quality/intensity, coverage/reach/ accessibility, acceptability/satisfaction, mechanism, and/or context
11. Overall design within which evaluation is part: RCT, matched comparison, uncontrolled before/after, cross-sectional, other (please state)
12. Timing of evaluation

- Details
- Not stated

13. Sampling methods

- Details
- Not stated

14. Sample size, response rates at baseline

- Details
- Not stated

15. Sample size, response rates at follow-up

- Not stated
- Details
- Not appropriate

16. Sociodemographic characteristics of participants at baseline/follow-up

- Details
- Not stated

17. Methods of data collection

- Details (period of data collection, types of data, comments on quality of data, etc.)
- Not stated

18.  Methods of data analysis

- Details
- Not stated.

## Quality assessment tool

Consider if the study has reported the following. Provide justifications for your response wherever possible.

1.  Were steps taken to minimise bias and error/increase rigour in sampling? [Yes/no, plus free text explaining decision (for example, was the sampling strategy appropriate to the questions being asked? Were all stakeholders included?).]
2.  Were steps taken to minimise bias and error/increase rigour in data collection? [Yes/no, plus free text explaining decision (for example, were data collection tools validated or piloted? Was data collection comprehensive, flexible and/or sensitive to provide a rich description of processes?).]
3.  Were steps taken to minimise bias and error/increase rigour in data analysis? [Yes/no, plus free text explaining decision (for example, were analysis methods systematic? Was diversity in perspectives explored?).]
4.  Were the findings of the study grounded in/supported by data? [Yes/no, plus free text explaining decision (for example, were enough data presented to show how the authors arrived at their findings? Do the data presented fit the interpretation provided?).]
5.  Was there good breadth and depth achieved in the findings? [Yes/no, plus free text explaining decision (for example, were a range of process issues covered in the evaluation? Were the perspectives of participants fully explored in terms of breadth – contrast of two or more perspectives – and depth –insight into a single perspective?).]
6.  Were the perspectives of young people privileged? [Yes/no, plus free text explaining decision (for example, were young people included in the evaluation? Was there a balance between open-ended and fixed-response options?).]
7.  How reliable or trustworthy overall are the findings (i.e. the extent to which the methods employed were rigorous/could minimise bias and error in the findings)? (Low/medium/high, plus free text explaining the decision.)
8.  How useful, overall, were the findings for shedding light on factors relating to the research questions? (Low/medium/high, plus free text explaining the decision.)

# Appendix 7 Data extraction and quality assessment for outcome evaluations

## About this appendix

Below are the data extraction and quality assessment tools that were used by reviewers for reports on outcomes.

## Data extraction tool

Consider if the study has reported the following. Provide justifications for your response wherever possible.

### *Intervention details*

1. Study location (country, area of country)

   - Details
   - Not stated

2. Intervention start date

   - Details
   - Not stated

3. Intervention duration

   - Details
   - Not stated

4. Target population (indicate whether this is a student, classroom, school, teacher or other population)

   - Details
   - Not stated

5. Description of intervention activities and how they were delivered to intervention group(s)

   - Details
   - Not stated

6. Description of components delivered to control group(s)

   - Details
   - Not stated

7. Description of integration

   - Details
   - Not stated

8. How intervention was developed

- Details
- Not stated

9. Provider organisation description

- Details
- Not stated.

## *Evaluation details*

10. Research question(s) or hypotheses

- Details
- Not stated

11. Timing of outcome evaluation

- Details
- Not stated

12. Unit of allocation

- School
- Classroom
- Student
- Other (please indicate)
- Not stated

13. Generation of allocation sequence (e.g. What was the random component of the sequence generation process – this may include minimisation?)

- Details
- Not stated

14. Concealment of allocation

- Details
- Not stated

15. Blinding

- Details
- Not stated

16. Sampling strategy

- Details
- Not stated

17. Sample size, overall response rates at baseline

   - Details (intervention group)
   - Details (control group)
   - Not stated

18. Sample size, overall response rates at follow-up

   - Details (intervention group)
   - Details (control group)
   - Not stated
   - Not appropriate

19. Sociodemographic characteristics at baseline/follow-up

   - Details
   - Not stated

20. Were baseline equivalence/differences between arms reported?

   - Details
   - Not stated

21. How were differences between intervention and comparison groups controlled?

   - Details
   - Not stated

22. For *each* outcome measure, report the following

   (a) Description of outcome measure provided

      - Details
      - Not stated

   (b) Outcomes stated (in protocol, from outset)

      - Primary outcomes
      - Secondary outcomes
      - Not stated

   (c) Post hoc outcomes reported

      - Description
      - Not stated

   (d) Evidence of reliability

      - Description
      - Not stated

(e) Data collection methods

- Description
- Not stated

(f) Baseline response rate

- Details (intervention group)
- Details (control group)
- Not stated

(g) Follow-up response rate

- Details (intervention group)
- Details (control group)
- Not stated

(h) Effect sizes

- Overall
- By sex
- By age
- By SES
- By ethnic subgroup
- No effect size

(i) Study analysis was intention to treat?

- Yes
- No
- Not stated

(j) Study analysis appropriately accounted for clustering?

- Yes
- No
- Not stated

(k) Study analysis adjusted for confounders?

- Yes
- No
- Not stated.

## Quality assessment tool

Consider if the study has reported the following. Provide justifications for your response wherever possible.

1.  Adequate generation (random) of the allocation sequence

    - Yes
    - No
    - Not clear
    - Not stated

2.  Concealed allocation

    - Yes
    - No
    - Not clear
    - Not stated

3.  Blinding

    - Yes
    - No
    - Not clear
    - Not stated

4.  Complete outcome data (i.e. low attrition): were complete data for each outcome reported, and if not, were reasons given for incomplete reporting?

    - Yes
    - No
    - Not clear
    - Not stated

5.  Reporting complete, not selective by measure

    - Yes
    - No
    - Not clear
    - Not stated

6.  Controlled for confounding

    - Yes
    - No
    - Not clear
    - Not stated

7.  Accounted for clustering

    - Yes
    - No
    - Not clear
    - Not stated

8. Aimed to reduce other forms of bias that may have entered the study

- Yes
- No
- Not clear
- Not stated

9. Suitable intensity and type of control group?

- Yes
- No
- Not clear
- Not stated.

# **Appendix 8** Author contact e-mail for additional information

## **About this appendix**

Below is the template used to contact authors of outcome evaluations for additional information.

## **E-mail template**

Dear _____,

Hello from the University of Warwick. I am currently collaborating with Prof Chris Bonell and Dr Tara Tancred of the London School of Hygiene and Tropical Medicine on a systematic review and meta-analysis of the effectiveness of interventions that integrate academic and health education to prevent substance use and violence, funded by the UK National Institute for Health Research. (Our protocol can be found here: www.crd. york.ac.uk/PROSPEROFILES/26464_PROTOCOL_20160011.pdf.) We wish to include results from the evaluation of _____ in our systematic review.

Because our key outcomes include _____, we wish to include estimates for _____.
To assist us in including estimates for these outcomes, would you be able to provide _____?

We would be most grateful for your assistance as we aim to complete this project in a timely fashion. Would it be possible to answer our queries in the next two weeks (i.e. by _____)? Please do not hesitate to get in touch with any questions you might have.

With best wishes,

GJ Melendez-Torres DPhil RN MFPH FHEA

Associate Professor in Health Technology Assessment, Epidemiology and Evidence Synthesis

The University of Warwick

# Appendix 9 Quality appraisal of reports of theories of change

**TABLE 13** Quality assessment of intervention theories of change

| Intervention name | Clarity | | | | Plausibility and feasibility | | | | Testability | | Ownership | | | | Generalisability | | | | | | Overall score (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Constructs defined | | Clear pathways from inputs to outcomes | | Theorised pathways are logically plausible | | Empirical evidence in support of theory | | Evidence of empirical testing of theory | | Theory developed with practitioners | | Theory developed with community members | | Theory presented as general | | Theory describes its application to different contexts | | Authors present evidence of a theory's generalisability | | |
| | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | |
| ***4Rs*** | | | | | | | | | | | | | | | | | | | | | |
| Aber 2011[53] | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | N | N | Y | N | Y | N | N | N | 55 |
| Brown 2010[54] | Y | N | Y | N | Y | Y | Y | Y | Y | Y | Y | N | N | N | Y | N | Y | N | N | N | 55 |
| [a]Flay 2009[55] | Y | Y | Y | Y | Y | Y | N | N | Y | Y | Y | N | N | N | Y | Y | N | N | N | N | 55 |
| Jones 2008[56] | Y | Y | N | Y | Y | Y | Y | Y | Y | Y | Y | N | N | N | Y | Y | N | Y | N | Y | 70 |
| Jones 2010[52] | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | N | N | N | Y | Y | Y | N | Y | N | 70 |
| Jones 2011[3] | N | Y | N | Y | Y | Y | Y | Y | Y | Y | Y | N | N | N | Y | Y | Y | N | N | N | 60 |
| Sung 2015[57] | N | Y | N | N | N | Y | Y | N | Y | N | Y | N | N | N | Y | N | Y | N | N | N | 35 |
| ***Bullying Literature Project*** | | | | | | | | | | | | | | | | | | | | | |
| Couch 2015[96] | N | N | N | N | N | N | Y | Y | Y | Y | N | N | N | N | Y | Y | N | N | N | N | 30 |
| Wang 2015[134] | Y | Y | Y | Y | Y | Y | Y | N | Y | Y | N | N | N | N | Y | N | N | N | N | N | 50 |
| ***DRACON*** | | | | | | | | | | | | | | | | | | | | | |
| Malm 2007[119] | N | Y | N | Y | Y | Y | N | N | N | Y | Y | Y | Y | N | Y | N | Y | N | Y | N | 55 |
| ***English classes (no name)*** | | | | | | | | | | | | | | | | | | | | | |
| Holcomb 1993[108] | N | Y | N | Y | N | Y | N | N | N | Y | Y | N | N | N | N | Y | Y | N | N | N | 35 |
| ***Hashish and Marijuana*** | | | | | | | | | | | | | | | | | | | | | |
| Zoller 1981[135] | N | Y | N | Y | N | Y | N | N | Y | N | N | N | N | N | N | Y | N | N | N | N | 25 |
| ***I-LST*** | | | | | | | | | | | | | | | | | | | | | |
| Bechtel 2006[84] | N | N | N | N | Y | N | Y | N | Y | N | N | N | N | N | Y | Y | N | N | N | N | 25 |

| Intervention name | Clarity | | | | Plausibility and feasibility | | | | Testability | | Ownership | | | | Generalisability | | | | | | Overall score (%) |
| | Constructs defined | | Clear pathways from inputs to outcomes | | Theorised pathways are logically plausible | | Empirical evidence in support of theory | | Evidence of empirical testing of theory | | Theory developed with practitioners | | Theory developed with community members | | Theory presented as general | | Theory describes its application to different contexts | | Authors present evidence of a theory's generalisability | | |
| | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | |
| **KAT** | | | | | | | | | | | | | | | | | | | | | |
| Segrott 2015[127] | N | N | N | Y | N | Y | N | Y | Y | Y | N | N | N | N | Y | Y | N | N | N | N | 35 |
| **LIFT** | | | | | | | | | | | | | | | | | | | | | |
| DeGarmo 2009[98] | Y | N | Y | Y | Y | Y | Y | Y | Y | Y | N | N | N | N | Y | Y | N | N | Y | N | 60 |
| Eddy 2000[100] | N | N | N | Y | Y | Y | Y | N | Y | N | N | N | N | N | N | Y | N | Y | N | N | 35 |
| Eddy 2015[99] | N | Y | Y | N | Y | Y | Y | Y | Y | Y | N | N | N | N | Y | Y | Y | N | N | N | 55 |
| Reid 1999[125] | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | N | N | N | Y | Y | N | N | N | N | 65 |
| Reid 2002[124] | N | N | N | Y | N | Y | Y | N | Y | Y | N | N | N | N | Y | Y | Y | N | N | N | 45 |
| **Peaceful Panels** | | | | | | | | | | | | | | | | | | | | | |
| Wales 2013[133] | N | N | N | Y | N | N | N | N | N | Y | N | N | N | N | Y | N | N | N | N | N | 15 |
| **Positive Action** | | | | | | | | | | | | | | | | | | | | | |
| Beets 2008[86] | N | Y | N | Y | Y | Y | N | N | N | N | N | N | N | N | Y | N | N | N | N | N | 30 |
| Beets 2009[87] | N | N | N | N | N | N | Y | N | Y | N | N | N | N | N | Y | Y | N | N | N | N | 20 |
| [a]Flay 2009[55] | N | Y | N | Y | Y | Y | Y | Y | Y | Y | N | N | N | N | Y | Y | N | N | N | Y | 60 |
| Flay 2010[102] | N | Y | Y | Y | Y | Y | Y | N | Y | Y | N | N | N | N | Y | Y | Y | N | N | N | 60 |
| Lewis 2012[113] | N | Y | N | Y | N | Y | N | N | N | N | N | N | N | N | N | Y | N | N | N | N | 25 |
| Malloy 2015[118] | N | N | N | N | N | Y | N | N | Y | Y | N | N | N | N | Y | Y | N | N | N | N | 25 |
| **PATHS** | | | | | | | | | | | | | | | | | | | | | |
| Crean 2013[97] | N | Y | N | Y | Y | Y | Y | N | Y | Y | N | N | N | N | Y | Y | N | N | N | N | 45 |
| [a]Flay 2009[55] | Y | Y | N | N | N | N | Y | Y | Y | Y | N | N | N | N | Y | Y | N | N | N | N | 35 |
| Greenberg 2006[106] | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | N | N | N | N | Y | Y | N | N | Y | N | 60 |
| Kusché 2012[112] | N | Y | Y | N | Y | Y | Y | N | Y | N | N | N | N | N | Y | Y | N | N | Y | N | 45 |

**TABLE 13** Quality assessment of intervention theories of change (*continued*)

| Intervention name | Clarity | | | | Plausibility and feasibility | | | | Testability | | Ownership | | | | Generalisability | | | | | | Overall score (%) |
| | Constructs defined | | Clear pathways from inputs to outcomes | | Theorised pathways are logically plausible | | Empirical evidence in support of theory | | Evidence of empirical testing of theory | | Theory developed with practitioners | | Theory developed with community members | | Theory presented as general | | Theory describes its application to different contexts | | Authors present evidence of a theory's generalisability | | |
| | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | CB | TT | |
| *RHC* | | | | | | | | | | | | | | | | | | | | | |
| Brown 2005[92] | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | N | N | N | N | Y | Y | Y | N | N | N | 60 |
| Catalano 2003[95] | Y | N | Y | N | Y | N | Y | N | Y | Y | N | N | N | N | Y | Y | N | N | N | N | 40 |
| *Roots of Empathy* | | | | | | | | | | | | | | | | | | | | | |
| Cain 2008[94] | N | Y | N | Y | Y | Y | Y | N | N | N | Y | N | N | N | Y | Y | N | N | N | N | 40 |
| Gordon 2003[105] | N | Y | N | N | N | Y | Y | N | Y | Y | Y | N | N | N | Y | Y | N | N | N | Y | 45 |
| Hanson 2012[107] | N | N | Y | N | Y | Y | Y | N | Y | Y | Y | N | N | N | Y | Y | N | N | Y | N | 50 |
| *Steps to Respect* | | | | | | | | | | | | | | | | | | | | | |
| Brown 2011[93] | Y | Y | Y | N | Y | N | Y | Y | Y | Y | N | N | N | N | Y | Y | N | N | Y | N | 55 |
| Frey 2009[103] | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | N | N | N | N | Y | Y | N | N | N | N | 55 |
| *The Gatehouse Project* | | | | | | | | | | | | | | | | | | | | | |
| Bond 2010[88] | N | N | N | Y | Y | Y | N | N | Y | Y | N | N | N | N | Y | Y | N | N | N | N | 35 |
| Patton 2000[122] | N | Y | N | Y | Y | Y | Y | N | Y | N | N | Y | N | Y | Y | Y | Y | N | Y | N | 60 |
| Patton 2003[121] | N | N | N | N | N | N | Y | N | Y | N | N | N | N | N | Y | Y | N | N | N | N | 20 |
| *Youth Matters* | | | | | | | | | | | | | | | | | | | | | |
| Jenson 2007[110] | N | Y | N | Y | Y | Y | N | N | Y | Y | N | N | N | N | Y | Y | N | N | N | N | 40 |

N, no; Y, yes.
a Note that Flay 2009[55] appears three times as it describes 4Rs, Positive Action and PATHS.

# Appendix 10 Theory studies: example coding template

## About this appendix

An example of two reviewers' node structure reports done in NVivo 8 (QSR International, Warrington, UK) for the PATHS intervention is provided below. Of note is that this intervention was slightly atypical in that it had a fairly robust description of five different constituent theoretical models underlying the intervention, although a limited description of if and how these models come together to produce the intervention's theory of change in practice.

### Sara Paparini node structure report

#### Inputs
Teacher training.

#### Intervention goals

#### Intervention goals linked to theory models
A–B–C–D [Affective–Behavioural–Cognitive–Dynamic]\learning that integrates affect cognition and behaviour

Eco-Behavioural System Model\caring classroom and school

Environment

Eco-Behavioural System Model\combining instruction with environmental support

Eco-Behavioural System Model\parent participation

Neurobiology\teaching strategies that promote brain structuralisation

Neurobiology\teaching strategies that promote brain structuralisation\teaching horizontal communication through labelling etc.

Neurobiology\teaching strategies that promote brain structuralisation\vertical structuralisation (regulating emotion and controlling impulse)

Intervention goals linked to theory models\Psychodynamic Education

Intervention goals linked to theory models\Psychodynamic Education\children 'own' or internalise values and concepts

Intervention goals linked to theory models\Psychodynamic Education\teach pro-social values and their importance

Intervention goals linked to theory models\psychology of emotional awareness

Intervention goals linked to theory models\psychology of emotional awareness\belonging to a broader human community is emphasised (v us v others)

Intervention goals linked to theory models\psychology of emotional awareness\children feel valued and learn to value themselves and others

## Key theoretical concepts and their mechanisms of change

### *Neurobiology and brain structuralisation and organisation*
Children can be unaware of the feelings they are having

Children can perform well cognitively but struggle with social functioning

### *Psychodynamic education*
Enhance development (not treatment)

Enhance development (not treatment)\internalisation of values (independent of external environment)

Enhance development (not treatment)\learning as joyful discovery

Enhance development (not treatment)\positive teacher–child relationships enhance learning

Enhance development (not treatment)\promotion of self-control and self-motivation develops autonomy and empathy

Enhance development (not treatment)\students learn from others' reactions towards them

Enhance development (not treatment)\Psychodynamic Education\teachers are powerful role models

### *Psychology of emotional awareness*
Children plan emotional strategies

Emotional plans and real-life situations

Importance of modelling and reinforcement of peers and adults

The importance of emotional intelligence and definition

### *A–B–C–D model*
Affective development precedes learning

Affective development precedes learning\0–3 preschool years

Affective development precedes learning\0–3 preschool years\affective perspective-taking skills

Affective development precedes learning\0–3 preschool years\pre-school child begins to verbalise (v act on) feelings

Affective development precedes learning\5–7 years

Affective development precedes learning\5–7 years\responsibility, independence and social roles development

Affective development precedes learning\elementary school years

Affective development precedes learning\elementary school years\integration of affect behaviour and cognition (language)

### The eco-behavioural systems model
Systems-level change (integration of the child within the environment)

### Outcomes
A sense of belonging to a broad human community deters violent behaviours

Optimal growth and mental health prevents antisocial behaviour, violence and substance use

## Tara Tancred node structure report

### Inputs
Training of staff (teachers)

### Intervention goals

### Psychodynamic Education
Psychodynamic Education\child development

Psychodynamic Education\enhanced learning through teacher–child relationships

Psychodynamic Education\internalisation

Psychodynamic Education\positive peer relationships

Psychodynamic Education\role of the teachers

### Improved psychological issues
Improved psychological issues\emotional intelligence

Improved psychological issues\psychological and emotional development

Improved psychological issues\self-awareness

Improved psychological issues\sense of belonging

### Key theoretical concepts and their mechanisms of change

### A–B–C–D model
Childhood development

Different domains affected

Emotional intelligence (linking effect and cause)

### Eco-behavioural systems model
Development of a positive classroom environment

Multiple levels targeted

Systems-level change

*Neurobiology and brain structuralisation and organisation*
Frontal cortex communication

Horizontal communication

Neurological development with emotional development

Vertical control

## Outcomes
Caring for others and the world

Decreased problem behaviours

Decreased problem behaviours\reduced substance use

Decreased problem behaviours\self-management

Decreased problem behaviours\violence prevention

Greater social interactions and social problem solving

Improved classroom and school health

Optimised mental health

Promoting positive protective factors

# Appendix 11 Meta-ethnography coding framework for theoretical studies

**TABLE 14** First-, second- and third-order codes used within a meta-ethnographic approach to synthesise mechanisms of change

| First-order codes | Second-order codes | Third-order codes | Central overarching code |
|---|---|---|---|
| Academic platform used to simultaneously build academic and health skills | Degree of integration of academic and health curricula (full or partial) | Eroding boundaries between health and academic education | Eroding boundaries at multiple and mutually reinforcing levels – between academic and health education, between teachers and students, between classrooms and the wider school environment, and between schools and families – will lead to the development of a community of engaged students oriented towards pro-social behaviour and away from substance use, violence and other risk behaviours |
| Health curricula bridged into academic curricula, with academic curricula sometimes intended to reinforce health messages and vice versa | | | |
| Decreasing time during the school day for health education owing to an emphasis on academic achievement (measured through standardised test scores) | Pragmatic rationale for integration | | |
| Mutually reinforcing effect of improved academic and health outcomes | Scientific rationale for integration | | |
| Students being less resistant to health messaging in integrated curricula | Additional rationale for integration | | |
| Integrated curriculum providing opportunities for repetition | | | |
| Opportunities for experiential learning provided | | | |
| Teachers' internalisation of curricula | Normalisation of pro-social behaviours through teacher internalisation of curricula | Eroding boundaries between teachers and students | |
| Teachers' role-modelling of curricula | | | |
| Forming bonds between pro-social peers and adults and acceptance of behaviours demonstrated in these relationships | | | |
| Establishment of good relationships between students and teachers | Establishing student connectedness to the classroom through positive teacher–student relationships | | |

TABLE 14 First-, second- and third-order codes used within a meta-ethnographic approach to synthesise mechanisms of change (*continued*)

| First-order codes | Second-order codes | Third-order codes | Central overarching code |
|---|---|---|---|
| Rewards for pro-social behaviours being given to students | Positive reinforcement of pro-social behaviours | Eroding boundaries between classroom and the wider school | |
| Students internalising pro-social beliefs and feeling positive about themselves following the demonstration of pro-social behaviours, leading to more pro-social behaviours | | | |
| Students having the opportunity to practise valued skills at multiple levels | | | |
| A sense of connectedness and bonding with the school is linked to overall emotional well-being and security experienced by students in school | Establishing student connectedness to the school | | |
| Connectedness to school fosters better academic learning as the school becomes a more positive environment that students are invested in | | | |
| Role-modelling of pro-social behaviours at home | Normalisation of pro-social behaviours | Eroding boundaries between schools and families | |
| Parents aware of curricula and expect pro-social behaviours (e.g. conflict resolution skills) to be practised at home | Provision of opportunities for practical skill development | | |

# Appendix 12　Coding framework for process studies

**TABLE 15** Coding framework for process studies

| Theme | Subtheme | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Student attitudes towards programme | Student perception of the integrated curriculum or its messages as being valuable is a facilitator | Student responsiveness / Students who perceived the programme as worthwhile achieved more positive outcomes | Topics relevant to students was a facilitator of the curriculum | Curricula not meeting student expectations for an academic class was a barrier | Student support of intervention was a facilitator | Early introduction (i.e. in primary school) of curricula was a facilitator | Targeted age of participant (primary school) was a facilitator | |
| | Student pre-existing attitudes aligning with programme ideals is a facilitator | Student (pre-existing) pro-social behaviours and feelings / More positive behaviours and feelings led to increased responsiveness, which had an indirect effect, reducing substance use | Student attitudes against bullying were facilitators of implementation | Positive student climate was a facilitator of implementation | | | | |
| Teacher attitudes towards programme | Teacher belief in their responsibility to teach and own the integrated curriculum was a facilitator | Teacher beliefs about responsibility to teach integrated curriculum and its perceived effectiveness was a facilitator | Shifts in teacher perceptions of their role was a facilitator | Teacher connectedness to and internalisation of the curriculum was a facilitator | Teacher belief in the importance of the programme was a facilitator | Teacher bond beliefs were facilitators | Teacher instruction/modelling beliefs were barriers | Teacher expressiveness beliefs were facilitators · Teacher project beliefs were barriers · Teacher display/control beliefs were facilitators · Teacher ownership of integrated lessons was a facilitator |
| | Positive teacher attitude towards and belief in the potential of the integrated programme was a facilitator | Teacher attitude to the programme was positively linked to implementation fidelity | Teacher attitude to SACD programs was positively associated with implementation fidelity | Teachers with a negative perception of the curriculum had low implementation fidelity | Teacher agreement about the effectiveness of integrated curricula was a facilitator | Positive staff climate was a facilitator | Lack of teacher faith in programme's abilities was a barrier | Teacher perception of the programme as a disruption to learning was a barrier · However, quality and fidelity of implementation decreased with grade level, which saw increasingly worse implementation |
| Teacher working environment | Teacher freedom to be innovative and have flexibility within the curriculum was a facilitator | Teacher perception of school innovativeness was positively linked to implementation fidelity | Flexibility of curricula enabling classroom-specific modifications was a facilitator | Teachers having autonomy and flexibility within the curriculum was a facilitator | However, lack of curriculum flexibility was a barrier | | | Teacher perception of school participatory nature was negatively linked to perceptions of programme quality |
| | Teachers working collaboratively and learning from one another was a facilitator | Teacher–teacher affiliation was a positive predictor of use of supplementary materials | Working collaboratively with colleagues was a facilitator | Finding common points of effort between committees working on the Gatehouse Project was a facilitator | Support from teachers with previous experience with the curriculum to new teachers was a facilitator | | | Teacher familiarity with integrated academic/health curricula was a facilitator |
| | Teachers feel well prepared to deliver the curriculum | Having practical and instructional support from a staff developer was a facilitator | Regular training and mentoring was a facilitator · Positive perception of training and coaching was a facilitator | Adequate teacher training was a facilitator | Ongoing integration of feedback into the programme was a facilitator | Teachers feeling unprepared or disempowered by the programme was a barrier | An organised and easy to follow curriculum was a facilitator · DRACON resources (namely the handbook) was a facilitator · Programme manual was a facilitator | The manual and guidance were facilitators |

| Theme | Subtheme | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Teacher workload and burnout is a barrier; therefore, good curricular and administrative support is essential | A heavy workload for teachers was a barrier to implementation | Lack of teacher time was a barrier | Teachers with high burnout and low administrative support were less likely to implement PATHS with quality and to generalise its concepts | | Staff turnover was a barrier | Teacher time required to develop new integrated lessons was a barrier to implementation | | Teacher burnout was a barrier to use of supplemental materials but not to the quality of the overall curriculum |
| Higher-level support of programme | Positive, supportive school climate that aligns with programme goals was a facilitator | Teacher perception of school climate (connectedness and administrative support) were positively linked to teacher beliefs about responsibility to teach SACD and teacher attitude to programme | Positive school climate was positively linked to programme fidelity through positive teacher beliefs about responsibility to teach SACD and teacher attitude to programme | Alignment of programme with school district ethos was a facilitator | School-wide approach to conflict was either a facilitator or barrier to the programme | Student connectedness to school was a facilitator | Higher perceived school-level bullying was a barrier | Higher percentage of students receiving school lunch was a barrier | Older teachers were more likely to report administrative support and good mentoring and coaching than younger teachers, which made teacher age a facilitator |
| | | A political and administrative environment that is amenable to integrated curriculum is necessary | Support from administrators (specifically the principal) was positively associated with implementation | Ongoing practical support from leadership to mainstream integrated health and academic learning was a facilitator | | | Higher-level support in funding and school policy was a facilitator | Supportive administration was a facilitator | |
| Parental support | Parental support of the programme, either through direct participation or indirectly through role-modelling of programme messaging at home, was a facilitator | Poor parental involvement, especially in family components of the programme, was a barrier | Parental engagement, encouraged through children's natural pestering tendencies, was a facilitator | Parents willingness to provide support to children was a facilitator | | Parent willingness to provide support to school was a facilitator | Conflicting messages at home vs. in the curriculum was a barrier | Conflicting messaging at home was a barrier | Relative school affluence was a barrier to implementation (poorer school populations had higher adult engagement, wealthier schools had less adult engagement) |

**TABLE 15** Coding framework for process studies (*continued*)

| Theme | Subtheme | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Teaching methods and programme inputs associated with integrated curricula | Having platforms to make health topics enjoyable, interactive and approachable was a facilitator | Student enjoyment of curriculum was a facilitator | Enjoyment of parent component was a facilitator | Student enjoyment of interactive/facilitative classroom environment was a facilitator | Non-judgemental approach was a facilitator | Harm reduction approach was a facilitator | Interactive teaching and learning was a facilitator | Interactive curricula was a facilitator | Opportunities to collaborate with peers was a facilitator | Student enjoyment of the use of art and humour as platforms for conflict resolution learning was a facilitator | Student ability to work independently was motivating and was a facilitator |
| | Integrating health and academic curricula was a facilitator, if teachers were supported in its delivery | | Integration of programme and academic curricula was a facilitator | | Repetition of health messaging across lessons was a facilitator | | Some life skills being difficult to integrate into academic classes was a barrier | Some teachers not being able to cover all academic material owing to integration was a barrier | | Drama as a neutral platform to approach conflict was a facilitator | |
| | Programme-specific factors linked to theory of change must be facilitated with high fidelity | School-specific data on risk and protective factors were enabling | | Building on other initiatives and ideas was a facilitator of good school strategies | Logistical problems when bringing students together was a barrier | Developing a team with shared vision was a facilitator | Establishing links with the wider community was a facilitator | External facilitators was essential in implementing the programme | Targeted support networks outside the school was a facilitator | Strong connection between different intervention components was a facilitator | Intervention facilitation of parent–school contact was a facilitator |

SACD, super audio compact disc.
Shaded cells indicate barriers to implementation. Unshaded cells indicate facilitators of implementation.

# Appendix 13 Data transformations

**TABLE 16** Data transformations by intervention

| Evaluation | Transformation |
|---|---|
| **Substance use** | |
| The Gatehouse Project[60,88,90,91] | Logistic transformation on ORs |
| I-LST[128,132] | At the last measurement wave, SDs for MDs from females used for males as well |
| LIFT[98,100,101,125,131] | Only significance reported; use within-study effect-size variance for alcohol for other estimates |
| RHC[92,95] | Logistic transformation on ORs derived from proportions for prevalence of use |
| | SEs for frequency of use effect sizes derived using model-implied means; for frequency of use of cigarettes, effect size set to 0 and variance estimated using mean of other within-study effect sizes for frequency of substance use |
| **Violence victimisation** | |
| Bullying Literature Project[96,134] | Use pre-test ICCs to account for clustering by classroom |
| Bullying Literature Project–Moral Disengagement[136] | Use pre-test ICCs to account for clustering by classroom |
| The Gatehouse Project[60,88,90,91] | Logistic transformation on ORs |
| Steps to Respect[103,104] | Assume *F*-test value of 0.5 for observed victimisation |
| Youth Matters[109–111] | Logistic transformation on ORs for dichotomous estimates of bullying victimisation |

ICC, intracluster correlation coefficient.

# Appendix 14 User involvement

## About this appendix

We presented policy and practitioner stakeholders with the following information in advance of our first discussions with them. A summary of questions posed and feedback from young people follows.

## Intervention summary

These interventions are those that meet the following criteria. They:

- are delivered in schools as part of the regular school day (we were not interested in after-school programmes, camps, etc.)
- are delivered to school-aged children and adolescents (we were not interested in nursery/ kindergarten years)
- are universal (we were not interested in targeted interventions, often for high-risk students)
- aim to reduce substance use (drugs and alcohol) and/or violence (physical acts of violence or aggression, bullying, etc.)
- integrate academic and health education (see more below) to achieve improved health/ behavioural outcomes.

## Integration of academic and health education

There is often the recognition that improvement of academic outcomes will also improve behavioural outcomes and vice versa. We generally found that there were two types of academic integration: 'full' or 'partial'. In fully integrated interventions, the academic curriculum serves as the platform from which the health curriculum is delivered, achieving both academic and health-related learning objectives simultaneously. An example of this is a programme called 4Rs, in which children's literature is used as a springboard in English classes to discuss themes related to violence and bullying. As children develop literacy skills, they also develop social and emotional skills that help to reduce violence.

Partially integrated interventions are those that include separate learning activities oriented towards health and academic learning objectives in the same overall package. For example, another intervention called PATHS has a literacy-building component as part of its year 5 curriculum, but it does not primarily focus on building academic skills.

## Intervention types

These interventions had the following characteristics (*Table 17*) depending on how they were implemented.

## What we have learned about theory of change (how these interventions are expected to work)

Although these interventions have different theories of change, there were a few common components that seemed to stand out.

**TABLE 17** Typology of interventions

| Intervention characteristic | Options |
|---|---|
| Target | Primary students, secondary students or both |
| Led by | Teachers, external facilitators or both |
| Type | Full or partial |
| Class to be integrated into | English, drama/theatre, art, biology, chemistry, maths, multiple |
| Outcome of interest | Substance use, violence or both |
| Components | Classroom only; classroom and whole school; classroom and family; or classroom and whole school and family |

Typically, these interventions are led directly by teachers. Thus, these teachers (and any other contributing school staff) are trained in a specific curriculum that they are expected to deliver/support in their classroom/ school. Sometimes these interventions have components at different levels. For example, there might be a curriculum around bullying introduced in a drama class, plus a school-wide intervention around recognising bullying and supporting affected students, plus a home component that encourages parents to model good problem-solving at home. In almost all interventions, teachers play a really critical role, and often their role-modelling of the behaviours stressed by the intervention, as well as their development of healthy, positive relationships with students, was felt to be a powerful mechanism by which these interventions functioned. Beyond that, the following were felt to be important ways through which these interventions work:

- students having opportunities to practise skills they learn (for example, active listening)
- students being given opportunities to reinforce what they have learned (for example, being rewarded for positive behaviours demonstrated during breaks or throughout other times in the school day)
- positive behaviours encouraged by these interventions generally being normalised in classrooms/schools and therefore internalised by students
- students developing positive bonds with 'pro-social others' (i.e. those who are demonstrating desired behaviours), leading to a sense of connectedness to classrooms and schools.

Ultimately, students are expected to develop better social–emotional skills and emotional intelligence (being able to recognise emotions in oneself and in others), leading to better self-management (being able to respond appropriately to those emotions). Fewer negative behaviours, such as violence or substance use, are then thought to occur.

## What we have learned about implementation

Beyond understanding how these interventions are expected to work, we have also learned about several important factors that affect the implementation of these interventions. Unsurprisingly, many hinge on the role of the teacher and the school in supporting them. *Table 18* briefly summarises some of these key facilitators of these interventions.

## What we are interested in learning from you

- Could this type of intervention be delivered in the UK?

  - If so, which types of intervention characteristics would be the most appropriate?

- Do you think that schools would be receptive to these types of interventions in the UK?
- What factors (practical, administrative or otherwise) would facilitate or inhibit the implementation of these types of interventions in the UK?

TABLE 18 Facilitators of integrated academic and health education interventions in schools

| Teacher-level | Student-level | School-level | Parent-level | Programme-level |
|---|---|---|---|---|
| Teachers are motivated to deliver the curriculum and believe in its messages | Students perceive the programme to be of relevance | School administrators provide practical and instructional support and leadership | Parents are involved | External support is available as needed |
| Teachers feel that they have a responsibility to teach the intervention's curriculum | Student is provided opportunities to collaborate with peers | School has an ethos linked to the intervention | Parents provide reinforcing (rather than conflicting) messages at home | Useful intervention curricula and other materials are taught to implementers (teachers) and are readily available |
| Teachers have autonomy and flexibility with the intervention's curriculum | Student enjoys the intervention's curriculum | | | Regular training and coaching is provided to implementers (teachers) |
| Teachers must feel that they have time for the intervention and are prepared for the intervention | | | | |

## Summary prior to second round of consultations

In advance of the second round of consultations, the following summary was shared with participants.

### How can interventions integrating health and academic education in schools help prevent substance misuse and violence? Systematic review and evidence synthesis

### Background and rationale

Our review focused on the prevention of substance use (alcohol consumption, smoking and drug use) and violence because these are important and interconnected health issues.

In the UK, many schools are reducing lessons that address these health issues because schools increasingly focus narrowly on academic attainment; school inspectors have limited reporting on schools' promotion of student health and personal development; and, in England, PSHE is currently not a statutory subject.

Existing reviews suggest that school curriculum-based health interventions can reduce alcohol consumption, smoking, drug use and violence, but these are increasingly difficult to deliver within constrained school timetables.

In this context, schools may deliver health education in other subjects, integrating it with academic learning. This approach may be effective because it could allow for larger 'doses' of health education, it may be less prone to student resistance and 'prevention fatigue' and it may allow health education taught across different lessons to reinforce learning.

However, the evaluation evidence on these sorts of interventions has not been reviewed.

### Aim and review questions

Aim: to search systematically for, appraise the quality of and synthesise evidence to address the following research questions.

Research question 1: what types of curriculum interventions integrating health and academic education in schools addressing substance use and violence have been evaluated?

Research question 2: what theories inform these interventions and what do these suggest about the way they are intended to work?

Research question 3: what characteristics of interventions, deliverers, participants and school contexts facilitate or limit successful implementation and receipt of such interventions, and what are the implications of these for delivery in the UK?

Research question 4: how effective are such interventions in reducing alcohol consumption, smoking, drug use and violence?

## Methods

We carried out a review of theories, process evaluations and outcomes evaluations of school-based curriculum interventions integrating health and academic education among students aged 4–18 years addressing substance use or violence. This review used systematic methods to define which studies were appropriate to include, search comprehensively for studies and then assess their quality and pool their results.

## Results

### Included studies

A total of 39 reports described theories, 16 reports (15 studies) evaluated process and 35 reports (13 studies) evaluated outcomes.

### Research question 1: types of interventions

Among our included studies, health curricula are either partially or fully integrated in an academic class. Fully integrated curricula have both academic and health goals that are addressed seamlessly within lessons. For example, one programme uses literature to teach themes about bullying, aiming to both reduce violence in children and improve literacy. Partially integrated programmes have separate components addressing academic learning and health education taught in the same lesson.

### Research question 2: theories of change

The interventions in this review aimed to integrate and thus erode boundaries between academic and health education. The role of teachers was emphasised. Teachers aimed to model positive behaviours and there were attempts to develop positive teacher–student and pro-social peer relationships. Classroom work was supported by other components delivered, for example, at the level of the overall school environment or to families. These aimed to ensure that learning and reinforcement of positive behaviours occurred beyond the classroom.

### Research question 3: characteristics facilitating or limiting successful implementation

Key facilitators were supportive senior management and fitting with the existing school ethos, a positive teaching environment including the ability to be flexible in the adaptation and delivery of lessons, positive pre-existing student and teacher attitudes towards intervention content, and parental support of interventions. Important barriers were overburdened teachers with little time to both learn and implement integrated curricula.

### Research question 4: effectiveness in reducing substance use and violence

The strongest evidence for the effectiveness of these interventions was found in the reduction of substance use in KS2 and KS3. There was no evidence for their effectiveness in reducing violence.

## Conclusions

Multilevel interventions that aim to erode boundaries (or, rather, strengthen pro-social relationships) between academic and health curricula, teachers and students, classrooms and schools, and schools and families are likely to have the greatest impact on risk taking and academic outcomes. These programmes may be effective in reducing substance use but do not appear to reduce violence.
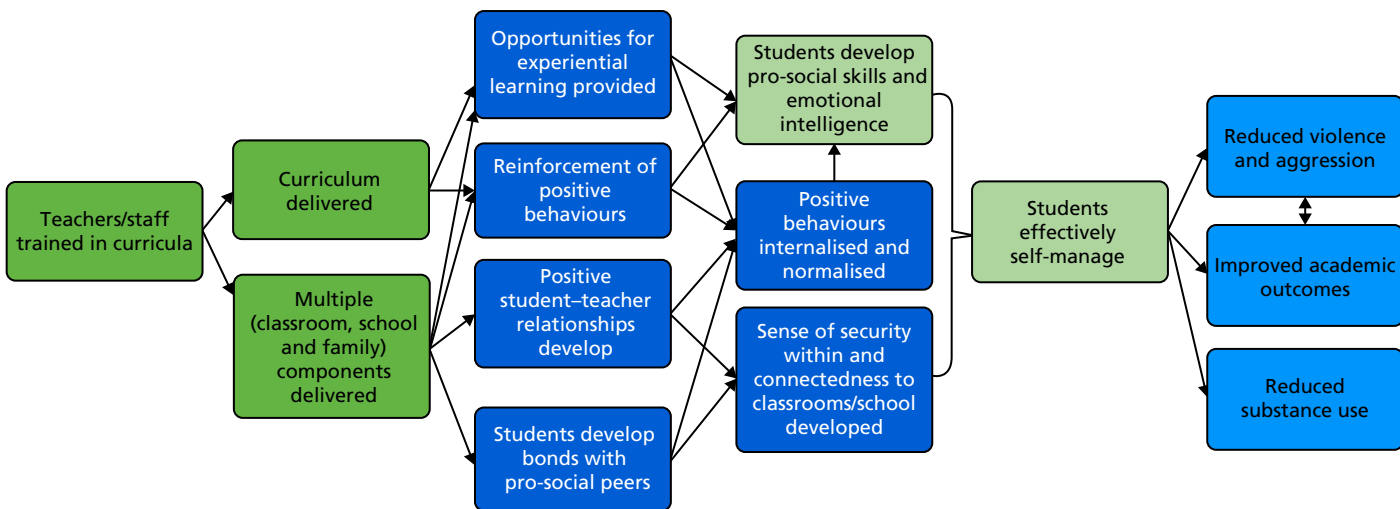
**FIGURE 18** Proposed logic model for interventions integrating academic and health curricula for reduced substance use and/or violence. Dark green, intervention inputs; dark blue, mechanism of change expected to facilitate outcomes; light green, proximal outcomes; light blue, distal outcomes.

# Appendix 15 List of included studies and intervention descriptions

TABLE 19 Summary of reports included in the review by research question answered

| Research question 2 (n = 39) | Research question 3 (n = 16) | Research question 4 (n = 35) |
|---|---|---|
| Aber 2011[53] | Bechtel 2006[84] | Aber 2011[53] |
| Bechtel 2006[84] | Beets 2007[85] | Aber 2017[138] |
| Beets 2008[86] | Beets 2008[86] | Bavarian 2013[83] |
| Beets 2009[87] | Bond 2001[89] | Beets 2009[87] |
| Bond 2010[88] | Cain 2008[94] | Bond 2004[91] |
| Brown 2005[92] | Hanson 2012[107] | Bond 2004[90] |
| Brown 2010[54] | Holcomb 1993[108] | Bond 2010[88] |
| Brown 2011[93] | Low 2014[117] | Brown 2005[92] |
| Cain 2008[94] | Malloy 2015[118] | Brown 2011[93] |
| Catalano 2003[95] | O'Toole 2005[120] | Catalano 2003[95] |
| Couch 2015[96] | Ransford 2009[123] | Couch 2015[96] |
| Crean 2013[97] | Rothwell 2011[126] | Crean 2013[97] |
| DeGarmo 2009[98] | Segrott 2015[127] | DeGarmo 2009[98] |
| Eddy 2000[100] | Sung 2015[57] | Eddy 2000[100] |
| Eddy 2015[99] | Wales 2013[133] | Eddy 2003[101] |
| Flay 2009[55] | Zoller 1981[135] | Espelage 2013[139] |
| Flay 2010[102] | | Espelage 2015[140] |
| Frey 2009[103] | | Espelage 2015[141] |
| Gordon 2003[105] | | Frey 2005[104] |
| Greenberg 2006[106] | | Frey 2009[103] |
| Hanson 2012[107] | | Jenson 2007[110] |
| Holcomb 1993[108] | | Jenson 2010[111] |
| Jenson 2007[110] | | Jenson 2013[109] |
| Jones 2008[56] | | Jones 2010[52] |
| Jones 2010[52] | | Jones 2010[51] |
| Jones 2011[3] | | Jones 2011[3] |
| Kusché 2012[112] | | Lewis 2012[113] |
| Lewis 2012[113] | | Lewis 2012[114] |
| Malloy 2015[118] | | Lewis 2013[115] |
| Malm 2007[119] | | Li 2011[116] |
| Patton 2000[122] | | Patton 2006[60] |
| Patton 2003[121] | | Reid 1999[125] |

**TABLE 19** Summary of reports included in the review by research question answered (*continued*)

| Research question 2 (*n* = 39) | Research question 3 (*n* = 16) | Research question 4 (*n* = 35) |
|---|---|---|
| Reid 1999[125] | | Segrott 2015[127] |
| Reid 2002[124] | | Smith 2004[128] |
| Segrott 2015[127] | | Snyder 2010[129] |
| Sung 2015[57] | | Snyder 2013[130] |
| Wales 2013[133] | | Torrente 2015[137] |
| Wang 2015[134] | | Stoolmiller 2000[131] |
| Zoller 1981[135] | | Vicary 2006[132] |
| | | Wang 2015[134] |
| | | Wang 2017[136] |

**TABLE 20** Included intervention descriptions

| Intervention | Studies | Country | Targeted grades/ages | Providers and organisation | Intervention development | Contents: delivery and dose | Integration of health and academic education |
|---|---|---|---|---|---|---|---|
| 4Rs | Aber 2011[53] Brown 2010[54] Flay 2009[55] Jones 2008[56] Jones 2010[52] Jones 2010[51] Jones 2011[3] Sung 2015[57] | USA | Kindergarten to grade 5 (aged 6–11 years) | Led by the Morningside Centre for Teaching Social Responsibility, a community-based non-profit organisation | Evolved from a predecessor, the Resolving Conflict Creatively Program, which was inspired by shifts in the USA towards academic achievement and standardised test performance, often at the expense of 'whole child' curriculum. The evolution of 4Rs was in response to an increasingly busy school day, and thus aimed to create a platform to teach conflict resolution within existing classes | A literacy-based social–emotional learning curriculum for elementary school students. There are two components: (1) a seven-unit, 21–35 lesson literacy-based curriculum in conflict resolution and social–emotional learning for children in primary school (to grade 5) and (2) intensive professional development for teachers | Each unit uses children's literature as a springboard to help students understand anger and develop skills in listening, co-operation, assertiveness and negotiation. Additionally, 4Rs aims to develop literacy skills in children to capitalise on the mutual positive effect that social–emotional and academic learning have on each other |
| Bullying Literature Project | Couch 2015[96] Wang 2015[134] | USA | Grade 4 (aged 10 years) | Not clear | Not clear, although the development of the Bullying Literature Project seems to have a basis in recognising and building on the limitations of existing bibliotherapy-based antibullying interventions | The Bullying Literature Project incorporates explicit antibullying instruction, opportunities for role-modelling of behavioural strategies introduced and bibliotherapy (the use of children's literature with reflection on themes related to the antibullying lesson, namely developing emotional intelligence, engaging in positive bystander behaviour, changing attitudes towards bullying and using effective strategies such as walking away, ignoring, talking it out, seeking help or using humour and deflection) to create dialogue and increased understanding of bullying. There are five sessions, one per week, lasting 30–45 minutes | The basis of the intervention stems from the use of children's literature to frame a discussion around bullying, after which students participate in writing activities to develop literacy skills and to reinforce the lessons' messages |
| Bullying Literature Project–Moral Disengagement | Wang 2017[136] | USA | Grade 3 (aged 9 years) | Not clear | Developed as a subsequent iteration of the Bullying Literature Project, with an increased focus on moral disengagement. The theoretical focus includes consideration of Bandura's concept of moral disengagement[172] | Similar to the Bullying Literature Project, but tasks and topics included a stronger focus on moral disengagement included in each lesson | The basis of the intervention stems from the use of children's literature to frame a discussion around bullying, after which students participate in writing activities to develop literacy skills and to reinforce the lesson's messages |

**TABLE 20** Included intervention descriptions (*continued*)

| Intervention | Studies | Country | Targeted grades/ages | Providers and organisation | Intervention development | Contents: delivery and dose | Integration of health and academic education |
|---|---|---|---|---|---|---|---|
| DRACON | Malm 2007[119]<br><br>O'Toole 2005[120] | Australia | Primary and secondary school (aged 7–16 years) | The intervention had overall co-ordination from the University of Gothenburg, Sweden | The original intervention was a project comparing cultural components of conflict between schools in Malaysia and Sweden. It then involved two more projects in Australia. Co-ordinated by Professor Mats Friberg from the University of Gothenburg, Sweden | The goals of DRACON are to use drama to develop cognitive understanding of conflict and bullying and to empower students to manage their own conflict both personally and in the broader school community. It is unclear, but it appears there are nine cycles, each run for 8 weeks with two 100-minute classes per week in the school year. Cycle two onwards has the aim of combining drama and peer teaching from older classes to younger classes. Classes also aim to develop group communication, familiarity, empathy, etc. | Conflict literacy is taught through 'enhanced forum theatre' and other drama techniques. However, the peer learning classes taught by the students are typically not delivered in drama classes but in other classes (e.g. English) owing to the perceived opportunities in the curriculum to discuss conflict and bullying |
| English classes (no name) | Holcomb 1993[108] | USA | Grades 8 and 9 (aged 14 and 15 years) | Baylor College of Medicine was responsible for developing and implementing the intervention | Unclear, but suggested to be inspired by very limited amount of dedicated time for health instruction | Teachers are trained and, working in pairs, develop integrated health/English material with a specific emphasis on the prevention of drug and alcohol use. It is unclear, but it seems that the intervention is expected to be delivered in (potentially) every English class, with health lessons integrated wherever appropriate | Health topics are infused into English classes. English was chosen as it was felt to be the subject into which non-traditional concepts could be discussed and is taken by all students |
| Hashish and Marijuana | Zoller 1981[135] | Israel | Upper high school (aged 17 and 18 years) | Unclear. Study was authored by two individuals from the Department of Education and the Division of Chemical Studies within the School of Education of the Kibbutz Movement | This intervention was inspired by the failure of other preventative interventions and a recognition of the need to engage students both academically and socially to effect change in their behaviour | The goal of the curriculum is to develop scientific knowledge of hashish and marijuana and to strengthen students' problem-solving and decision-making skills. Lessons about hashish and marijuana are delivered through lectures in chemistry classes and complemented by group work, class discussions, independent study, projects, games, field trips and any other curricula additions that develop social and decision-making skills | The intervention is entirely integrated into chemistry classes, where lessons around hashish and marijuana take place, teaching the chemical aspects of the drugs. Behaviour change is also addressed through more participatory teaching methods |

| Intervention | Studies | Country | Targeted grades/ages | Providers and organisation | Intervention development | Contents: delivery and dose | Integration of health and academic education |
|---|---|---|---|---|---|---|---|
| I-LST | Bechtel 2006[84] Smith 2004[128] Vicary 2006[132] | USA | Middle/junior high school (aged 13–15 years) | Princeton Health Press was the commercial enterprise responsible for supporting the I-LST curriculum | LST interventions were first established as tobacco prevention interventions. Different studies of LST interventions, when applied to different areas, were also found to be successful in reducing alcohol and other substance use. However, findings have been contested and the need for testing of different models of LST delivery inspired the I-LST intervention and its study | I-LST aims to promote social skills, personal self-management skills and drug resistance skills, all for reduced substance use. Core components include self-image and self-improvement; decision-making; smoking, marijuana and alcohol myths and realities; smoking and advertising awareness; coping with anxiety; communication skills; social skills; and assertiveness. Lessons are integrated into the existing curriculum | Participating teachers delivering I-LST develop a matrix of LST principles that are then mapped onto curricular content areas. The hypothesis is that the I-LST curriculum achieves dual objectives of both LST and the usual academic subject. 'In this way, more teachers can provide the LST lessons, allowing for repeated exposure to the constructs in a variety of topics and eliminating the necessity for time set aside for prevention in an already burdened curriculum'[132] |
| KAT | Rothwell 2015[126] Segrott 2015[127] | UK | Grades 5 and 6 (aged 11 and 12 years) | Not stated | The KAT intervention was derived from an Australian predecessor (P-KAT) that focused more broadly on substance misuse (vs. alcohol only). The intervention was developed by the 'program organiser' and a working group, of which two members were head teachers from the two participating schools in the included studies within this review | The KAT intervention aims to reduce drinking and antisocial behaviours in young people. It has three components: (1) a classroom component in which students learn material and prepare presentations for a parent evening, (2) a parent evening in which students demonstrate what they have learned to their parents so that parents and children can discuss key issues together and (3) a follow-up DVD for children and parents to watch together | The classroom component integrates lessons around the effects and consequences of alcohol use into various classes. For example: in literacy classes, in addition to learning about alcohol, students also learn how to collect data; in art classes, learning is tied into development of skills in art and design, for example, by designing alcohol awareness posters; and in drama classes, students use role play and other techniques to raise awareness about the effects of alcohol in families and the broader community |

**TABLE 20** Included intervention descriptions (*continued*)

| Intervention | Studies | Country | Targeted grades/ages | Providers and organisation | Intervention development | Contents: delivery and dose | Integration of health and academic education |
|---|---|---|---|---|---|---|---|
| Learning to Read in a Healing Classroom | Torrente 2015[137]<br><br>Aber 2017[138] | Democratic Republic of Congo | Years 3 to 5 | International Rescue Committee with New York University | This intervention was developed from a prior intervention, Healing Classrooms, created by the International Rescue Committee to support classroom learning in conflict-affected settings. This iteration explicitly integrated literacy learning into the intervention | The intervention has two components delivered specifically to active treatment schools: teacher learning circles, which are opportunities for teachers to learn from master educators about classroom practices that enhance growth and development, and teacher resource materials, which included literacy components and social–emotional learning components. These resource materials are intended to integrate social–emotional learning alongside learning to read. Materials were provided in French | The classroom component includes materials to support teachers in creating safe and supportive learning environments as well as using literature to teach students emotional regulation, self-expression and how to interact with peers |
| LIFT | DeGarmo 2009[98]<br><br>Eddy 2000[100]<br><br>Eddy 2003[101]<br><br>Eddy 2015[99]<br><br>Reid 1999[125]<br><br>Reid 2002[124]<br><br>Stoolmiller 2000[131] | USA | Grades 1 and 5 (aged 7 and 11 years) | Not stated | Owing to rising levels of delinquency in the 1990s, school-based interventions to reduce delinquency were created. LIFT was one such intervention. The authors refer to the existence of scientific literature and clinical knowledge around antisocial behaviours that informed the development of the intervention | The LIFT intervention has classroom and playground components as well as a parent training intervention. It aims to reduce antecedents of violence and delinquency in children. The classroom component is 1 hour, twice per week, for 10 weeks. Each class had the following structure: lecture and role play around social and problem-solving skills and small- and large-group skills practice. The playground component involves free play in the playground and skills review and presentation of daily rewards. In the skills review, negative behaviours are noted and subtracted from positive ones, which are rewarded. In the parent component, LIFT instructors meet with parents from 10–15 families once per week for 6 weeks. Parents are given home activities for the week to reflect their children's curriculum and have their own lecture and role plays | The fifth-grade classroom component also develops study skills that align with the fifth-grade curricula |

| Intervention | Studies | Country | Targeted grades/ages | Providers and organisation | Intervention development | Contents: delivery and dose | Integration of health and academic education |
|---|---|---|---|---|---|---|---|
| Peaceful Panels | Wales 2013[133] | USA | Grades 8 and 9 (aged 14 and 15 years) | One teacher (the study author) | The author drew extensively from the social reconstructionist art education paradigm, which aims to develop critical thinking skills. The author also chose to teach violence prevention as taught by the Second Step intervention | Throughout what appears to be all art classes in a school year, students participate in antibullying lessons (from the Second Step intervention for eighth-grade students on empathy and communication in handling a grievance) and comic-making lessons. They then prepare artwork to demonstrate their understanding of how to resolve conflict | Conflict resolution lessons are integrated fully within art classes and creation of art is used as a medium to reflect on learning |
| Positive Action | Bavarian 2013[83] Beets 2007[85] Beets 2008[86] Beets 2009[87] Flay 2009[55] Flay 2010[102] Lewis 2012[113] Lewis 2012[114] Lewis 2013[115] Li 2011[116] Malloy 2015[118] Snyder 2010[129] Snyder 2013[130] | USA | Kindergarten to grade 12 (aged 6–18 years) | Written by Carol Allred and first delivered in 1977. The programme was launched under the auspices of the Social and Character Development Cooperative Agreement, funded by the Institute for Education Science of the US Department of Education | Positive Action is grounded in the theories of self-concept, particularly self-esteem enhancement theory, which suggests that people inherently want to feel good and will use cognitive, affective and behavioural strategies to help achieve and sustain this feeling. Thus, the programme's classroom curriculum introduces motivation and opportunities for students to feel good about themselves, plus ecological supports to sustain it. Authors also mentioned the shifts in curriculum following the No Child Left Behind Act of 2001,[151] which placed huge emphasis on standardised test performance, squeezing out non-academic curricula. There has also been a movement towards integrated interventions that involve families and communities. Positive Action was developed to address all of these aspects | Positive Action is a social–emotional and character development intervention aimed at encouraging positive behaviours through positive thoughts and actions. Students are given 140 lessons, each 15–20 minutes, throughout the school year by their teachers. Lessons cover six units: self-concept, positive actions for mind and body, positive social–emotional actions, managing oneself, being honest with oneself and continually improving oneself. Teachers receive training and principals at participating schools receive a 'school climate kit' in which they are provided with directions to support the core elements of the Positive Action curriculum throughout the school | Positive Action recognises the relationship between students' academic performance, their learning and life skills and their behaviours. The methods of education used in Positive Action promote active learning and learning skills development alongside behaviour change. Students are encouraged to solidify academic learning by applying it to real-life scenarios. For example, unit two on positive actions for the mind and body emphasises thinking skills, problem-solving and study skills. Teachers are also encouraged to tie lessons into academic content |

**TABLE 20** Included intervention descriptions (*continued*)

| Intervention | Studies | Country | Targeted grades/ages | Providers and organisation | Intervention development | Contents: delivery and dose | Integration of health and academic education |
|---|---|---|---|---|---|---|---|
| PATHS | Crean 2013[97] Flay 2009[55] Greenberg 2006[106] Kusché 2012[112] Ransford 2009[123] | USA | Kindergarten to grade 5 (aged 6–11 years) | Not clear | PATHS was motivated by the psychological concepts of emotional intelligence and self-regulation | PATHS aims to develop social and emotional competencies in children for reduced aggression and behaviour problems. Throughout the primary school years, students receive 131 lessons grouped under three units: readiness and self-control, feelings and relationships, and inter-personal problem-solving. These units cover five domains: (1) self-control, (2) emotional understanding, (3) positive self-esteem, (4) healthy relationships and (5) interpersonal problem-solving skills. PATHS lessons are taught approximately two or three times per week, with lessons lasting 20–30 minutes, throughout the school year | PATHS was designed to be taught by regular classroom teachers (initially with support from consultant trainers) as an integrated component of the regular year-long curriculum. English skills (both reading and writing) are bridged to PATHS in most lessons by including supplementary suggestions for teachers to utilise such things as quality children's literature to reinforce lesson concepts. One of the chapters in the PATHS instructional manual provides many ideas for how teachers can directly tie PATHS concepts to English, social studies/history and other subject areas |

| Intervention | Studies | Country | Targeted grades/ages | Providers and organisation | Intervention development | Contents: delivery and dose | Integration of health and academic education |
|---|---|---|---|---|---|---|---|
| RHC | Brown 2005[92] Catalano 2003[95] | USA | Grades 1 to 8 (aged 7–14 years) | Not stated | RHC is guided by the social development model, which emphasises the importance of early intervention prior to the onset of antisocial behaviours and individual, family and community (including school and peer influence) characteristics | RHC is a four-component intervention that promotes positive youth development by targeting risk and protective factors at different stages of development. The components are (1) school intervention strategies (staff and teacher development and monthly coaching of teachers), (2) student intervention strategies (after-school tutoring and study clubs in grades 4 to 6 and group-based workshops in high school, which are for academic achievement, pro–social bonding, teaching refusal skills and developing pro-social beliefs around healthy behaviours), (3) peer intervention strategies (through classroom instruction and annual summer camps in primary school and social-skills booster retreats in secondary schools for social–emotional learning and developing problem-solving skills) and (4) family intervention strategies (parenting workshops to increase parenting skills, decrease familial conflict, clarify family standards and expectations around substance use and to practice peer resistance skills) | Staff are trained to promote reading as part of school intervention strategies. Also, in grades 4 to 6 there are after-school study groups and group-based workshops in secondary school, all of which aim to increase academic achievement |

**TABLE 20** Included intervention descriptions (*continued*)

| Intervention | Studies | Country | Targeted grades/ages | Providers and organisation | Intervention development | Contents: delivery and dose | Integration of health and academic education |
|---|---|---|---|---|---|---|---|
| Roots of Empathy | Cain 2008[94]<br><br>Gordon 2003[105]<br><br>Hanson 2012[107] | Australia (Cain), Canada (Gordon) and the UK (Hanson) | Grades 1 to 9 (aged 7–15 years) | The Roots of Empathy facilitator primarily delivers the intervention. However, teachers can also work to reinforce lesson material. The Department of Education and Training oversaw the intervention's implementation in Australia. In Canada, this intervention has local provincial support | The Western Australian Government has legislated that core shared values as social outcomes must be part of the curriculum. In 2004, the Department of Education and Training sought out a trial of the Canadian-based Roots of Empathy intervention, which was developed in response to poor academic outcomes in inner city children in Toronto | Roots of Empathy is an intervention that brings a visiting baby and mother (and/or father) into a classroom and, through students' exploration of the baby's interactions with his/her parent(s), a platform for learning empathy is provided. Visits are facilitated by a certified Roots of Empathy instructor who works with the class before, during and after each baby visit. Students build empathy by observing the baby's development, milestones, interacting with the baby and learning about its needs and temperament and, through this, learn messages of social inclusion, respect, how to build consensus and how to contribute to a safe and caring classroom while developing emotional intelligence. Throughout the school year, a neighbourhood baby will visit the class with its parent(s) nine times | Connections with literacy, writing, art, music, mathematics and science are made. Lessons may involve literature to help develop a theme to encourage children to explore their own emotions and experiences. Children may also make artwork to reflect these feelings, which is often compiled and used in follow-up activities |
| Second Step | Espelage 2013[139]<br><br>Espelage 2015[102,140]<br><br>Espelage 2015[141] | USA | Grade 6 (aged 11–12 years) | Intervention developed by Committee for Children, delivered by teachers after training | The intervention was developed from a perspective of social learning theory, with a focus on modelling, role playing and skills acquisition as well as performance coaching. Group work with peers provides reinforced opportunities for skills acquisition. The intervention was developed to address shared risk and protective factors between bullying, sexual harassment and sexual violence, and substance use | The intervention was delivered by teachers after implementation training. The dose includes 15 weeks of classroom lessons taught weekly or every 2 weeks throughout the school year for 3 years. The intervention consists of manualised content including didactic lessons, group activities and multimedia content including modelling of skills | Students receive academic homework integration assignments and teachers are encouraged to connect lessons to current events |

| Intervention | Studies | Country | Targeted grades/ages | Providers and organisation | Intervention development | Contents: delivery and dose | Integration of health and academic education |
|---|---|---|---|---|---|---|---|
| Steps to Respect | Brown 2011[93] Frey 2005[104] Frey 2009[103] | USA | Grades 3 to 6 (aged 9–12 years) | Committee for Children (no additional detail provided) | Not clear | Steps to Respect is an antibullying intervention that has both classroom and school-wide components. In the classroom, students learn social–emotional skills to counter bullying (including how to recognise, refuse and report bullying) and to promote the development of healthy friendships. At the school level, policy changes to increase staff awareness of, and responsiveness towards, bullying are created | Bi-weekly lessons in the Steps to Respect curriculum are supported by eight to 10 literature-based lessons, which also develop literacy skills alongside furthering understanding of the Steps to Respect curricular themes |
| The Gatehouse Project | Bond 2001[89] Bond 2004[91] Bond 2004[90] Bond 2010[88] Patton 2000[122] Patton 2003[121] Patton 2006[60] | Australia | Grade 8 (aged 14 years) | Experienced educators developed the intervention. They became facilitators of the intervention and then acted as liaisons between the schools and the intervention | The intervention was developed to address shortcomings in school health promotion, building on an intervention that emphasises whole-school change. The Gatehouse Project is in line with the health-promoting schools framework. The conceptual framework drew on adolescent risks and incorporated work on attachment and social support theories | The Gatehouse Project is a school health promotion intervention that includes both institutional- and individual-level components. The intervention aims to build a sense of security and trust in students, enhance skills and opportunities for good communication and build a sense of positive regard through participation in school life. The teaching curriculum is delivered over each term in normal classes. It focuses on cognitive and interpersonal skills that underlie emotional well-being relevant to teenagers. At the institutional level, a school-based adolescent health team (comprising administrators, teachers and community members) is established to identify risks in the school and create strategies to address these risks | Teaching resources are delivered in usual English, health or personal development classes. The intervention also adopts a critical literacy approach. Concepts are taught using literature, poetry, song, film and visual materials. The integration of the Gatehouse Project and normal curricula is intended to embed the work such that it could be continued beyond the project. There is a deliberate link made between the intervention's goals and academic goals |

**TABLE 20** Included intervention descriptions (*continued*)

| Intervention | Studies | Country | Targeted grades/ages | Providers and organisation | Intervention development | Contents: delivery and dose | Integration of health and academic education |
|---|---|---|---|---|---|---|---|
| Youth Matters | Jenson 2007[110]<br><br>Jenson 2010[111]<br><br>Jenson 2013[109] | USA | Grades 4 and 5 (aged 10 and 11 years) | 'Educational specialists' taught the curriculum | The social development model, which integrates social controls, social learning and differential association theories provided some basis for the intervention | Youth Matters promotes the development of healthy relationships between students, staff and schools. It also promotes skills around social resistance and social competency through a curriculum. There are four curriculum models, each with 10 lessons, delivered across 2 years. These were taught through instruction followed by student discussion and projects to promote positive social norms in schools. All sessions emphasise the perspective of both the bully and the victim | Each module uses a 30- to 40-page story that is intended to help schools meet academic standards in both health education and English |

P-KAT, Parents and Kids Acting Together.

# Appendix 16 Included theoretical studies

**TABLE 21** Summary of interventions and reports included in the theory synthesis

| Intervention | Description | Countries in which intervention has been evaluated | Targeted grade of schooling and approximate ages of participants | Integration of health and academic education | Included reports that provide details of theory of change |
|---|---|---|---|---|---|
| 4Rs | A literacy-based social–emotional learning curriculum for elementary school students. There are two components: (1) a seven-unit, 21–35 lesson literacy-based curriculum in conflict resolution and social–emotional learning and (2) intensive professional development for teachers | USA | Kindergarten to grade 5 (aged 6–11 years) | Each unit uses children's literature to help students understand anger and develop skills in listening, co-operation, assertiveness and negotiation. Additionally, 4Rs aims to develop literacy skills in children to capitalise on the mutual positive effect that social–emotional and academic learning have on each other | Aber 2011[53]  Brown 2010[54]  Flay 2009[55]  Jones 2008[56]  Jones 2010[52]  Jones 2011[3]  Sung 2015[57] |
| Bullying Literature Project | This intervention aims to reduce bullying by introducing themes related to bullying through children's literature. It also provides an opportunity for children to role-model practical skills to address or avoid bullying | USA | Grade 4 (aged 10 years) | The Bullying Literature Project integrates themes related to bullying into the children's literature used in a standard English curriculum | Couch 2015[96]  Wang 2015[134] |
| DRACON | This intervention uses drama to develop cognitive understanding of conflict and bullying and to empower students to manage their own conflict, both personally and in the broader school community | Australia | Primary and secondary school students (aged 7–16 years) | Conflict literacy is taught through 'enhanced forum theatre' and other drama techniques | Malm 2007[119] |
| English classes (no name) | Teachers were trained and, working in pairs in the summer, they developed integrated health/English material, with a specific emphasis on the prevention of drug and alcohol use | USA | Grades 8 and 9 (aged 14 and 15 years) | Health topics were infused into English classes. English was chosen as it was felt to be the subject into which non-traditional concepts could be discussed and is taken by all students | Holcomb 1993[108] |
| Hashish and Marijuana | The goal of the curriculum is to develop scientific knowledge of hashish and marijuana and to strengthen students' problem-solving and decision-making skills through both didactic and participatory learning approaches | Israel | Upper secondary school (aged 17 and 18 years) | The intervention was entirely integrated into chemistry classes, where hashish and marijuana lessons teach the chemical aspects of the drugs. Behaviour change was addressed through more participatory teaching methods | Zoller 1981[135] |

| Intervention | Description | Countries in which intervention has been evaluated | Targeted grade of schooling and approximate ages of participants | Integration of health and academic education | Included reports that provide details of theory of change |
|---|---|---|---|---|---|
| I-LST | This intervention integrates the messaging of standard LST intervention (self-image and self-improvement, decision-making, smoking, marijuana, alcohol, etc.) into core academic classes such as English, science and maths for healthier behaviours, including reduced substance use | USA | Middle/junior high school (aged 13–15 years) | The entire intervention hinges on the 'infusion' of health messaging into the core academic curriculum | Bechtel 2006[84] |
| KAT | This intervention comprised a classroom component in which children learn about alcohol, a parent evening in which students prepare presentations for parents and a take-home DVD about alcohol that is to be watched with parents for normalisation of reduced alcohol use | UK | Grades 5 and 6 (aged 11 and 12 years) | The classroom curriculum is delivered across many subjects and students develop academic skills alongside understanding of alcohol at every opportunity. For example, in arts classes, students will learn design skills by making anti-alcohol posters | Segrott 2015[127] |
| LIFT | The intervention aims to reduce future delinquency by positively rewarding and reinforcing good behaviours. The intervention includes classroom, playground and family components | USA | Grades 1 and 5 (aged 7 and 11 years) | The fifth-grade classroom component also develops study skills that align with the fifth-grade curriculum | DeGarmo 2009[98] Eddy 2000[100] Eddy 2015[99] Reid 1999[125] Reid 2002[124] |
| Peaceful Panels | Throughout art classes, students participated in antibullying lessons on empathy and communication in handling a grievance and comic-making lessons. They then prepared artwork to demonstrate their understanding of how to resolve conflict | USA | Grades 8 and 9 (aged 14 and 15 years) | Conflict resolution lessons were integrated fully in art classes and creation of art was used as a medium to reflect on learning | Wales 2013[133] |

**TABLE 21** Summary of interventions and reports included in the theory synthesis (*continued*)

| Intervention | Description | Countries in which intervention has been evaluated | Targeted grade of schooling and approximate ages of participants | Integration of health and academic education | Included reports that provide details of theory of change |
|---|---|---|---|---|---|
| Positive Action | Positive Action is a social–emotional and character development intervention aimed at encouraging positive behaviours through positive thoughts and actions. Lessons cover six units: self-concept, positive actions for mind and body, positive social–emotional actions, managing oneself, being honest with oneself and continually improving oneself | USA | Kindergarten to grade 12 (aged 6–18 years) | The methods of education used promote active learning and learning skills development alongside behaviour change. Teachers are also encouraged to tie lessons into academic content | Beets 2008[86] Beets 2009[87] Flay 2009[55] Flay 2010[102] Lewis 2012[113] Malloy 2014[118] |
| PATHS | PATHS aims to develop social and emotional competencies in children for reduced aggression and behaviour problems. Throughout the elementary school years, students receive lessons grouped under three units: readiness and self-control, feelings and relationships, and interpersonal problem-solving | USA | Kindergarten to grade 5 (aged 6–11 years) | Both reading and writing are bridged to PATHS in most lessons by including supplementary suggestions for teachers to utilise such things as quality children's literature to reinforce lesson concepts. Teachers can directly tie PATHS concepts to English, social studies/history and other subject areas | Crean 2013[97] Flay 2009[55] Greenberg 2006[106] Kusché 2012[112] |
| RHC | A four-component (school, student, peer and family) intervention that promotes positive youth development by targeting risk and protective factors at different stages of development. The focus is on pro-social bonding, social–emotional learning and the development of problem-solving skills | USA | Grades 1 to 8 (aged 7–14 years) | Staff were trained to promote reading as part of school intervention strategies. Also, in grades 4 to 6 there were after-school study groups and group-based workshops in high school, all of which aimed to increase academic achievement | Brown 2005[92] Catalano 2003[95] |
| Roots of Empathy | An intervention that brings a visiting baby and their parent(s) into a classroom as a springboard for learning empathy. Students learn messages of social inclusion, respect, how to build consensus, how to contribute to a safe and caring classroom, and develop emotional literacy | Australia, Canada, UK | Grades 1 to 9 (aged 7–15 years) | Connections with literacy, writing, art, music, mathematics and science are made. Lessons may involve literature to encourage children to explore their own emotions and experiences. Children may also make artwork to reflect these feelings, which is often compiled and used in follow-up activities | Cain 2008[94] Gordon 2003[105] Hanson 2008[107] |

| Intervention | Description | Countries in which intervention has been evaluated | Targeted grade of schooling and approximate ages of participants | Integration of health and academic education | Included reports that provide details of theory of change |
|---|---|---|---|---|---|
| Steps to Respect | A bullying prevention intervention that has both a classroom curriculum aimed at teaching emotional intelligence, bullying prevention and bystander skills to children and a schoolwide component that trains teachers and administrators to change policies around disciplining bullying and providing mentoring to prevent future occurrences | USA | Grades 3 to 6 (aged 9–12 years) | In the classroom curriculum, children's literature is used to frame lessons and develop social–emotional and antibullying skills | Brown 2011[93] Frey 2009[103] |
| The Gatehouse Project | Through teaching a curriculum and establishing a school-wide adolescent health team, the Gatehouse Project aims to build a sense of security and trust in students, enhance skills and opportunities for good communication and build a sense of positive regard through participation in school life | Australia | Grade 8 (aged 14 years) | The intervention adopts a critical literacy approach. Concepts are taught using literature, poetry, song, film and visual materials. There is a deliberate link made between the intervention's health goals and academic goals | Bond 2010[88] Patton 2000[122] Patton 2003[121] |
| Youth Matters | Youth Matters promotes the development of healthy relationships between students, staff and schools. It also promotes skills around social resistance and social competency through a curriculum. The curricula also emphasises both the bully and the victim perspective | USA | Grades 4 and 5 (aged 10 and 11 years) | Each module uses a 30- to 40-page story that is intended to help schools meet academic standards in both health education and English | Jenson 2007[110] |

# Appendix 17 Characteristics of theory reports

**TABLE 22** Summary of intervention theories of change

| Intervention | Studies | Summary of theory of change | Existing theories drawn on |
|---|---|---|---|
| 4Rs | Aber 2011[53]<br><br>Brown 2010[54]<br><br>Flay 2009[55]<br><br>Jones 2008[56]<br><br>Jones 2010[52]<br><br>Jones 2011[3]<br><br>Sung 2015[57] | As an intervention integrated into classroom curricula, the 4Rs theory of change draws from the theory of developmental cascades. There are three principles guiding the 4Rs theory of change: (1) the intervention promotes change at multiple levels, (2) if functioning in one domain (e.g. interpersonal relationships) improves, functioning in other domains (e.g. academic performance) will also improve and (3) the intervention promotes the establishment of dynamic systems that change over time at multiple levels in multiple domains. The intervention's outcomes are expected to be cumulative and have a stronger effect (initially) on the most proximal outcomes and then later expand to the intervention's distal outcomes across levels after more exposure. There are three classroom domains: emotional support, classroom organisation and instructional support. These are all felt to be strengthened through introducing teachers to social–emotional learning skills and concepts, which are to be used daily in interactions with students, other teachers and school administrators. Overall, strengthening the classroom domains will promote student academic learning and social–emotional development<br><br>To target underlying processes that lead to aggression and violence, the 4Rs classroom curriculum focuses on mental, interpersonal and social–cognitive processes. These are the children's attributional biases towards interpreting social cues, normalised beliefs around aggression, and interpersonal negotiation strategies with peers. As such, with 4Rs implemented in a school, there are three assumptions that are key to the theory of change: (1) students are expected to develop social–emotional competencies by reacting to norms in their school environments, with an emphasis on positive behaviours, (2) through environmental consequences of their actions, particularly reinforcement of positive behaviours (e.g. praise, small rewards), students are thought to develop social–emotional competencies and (3) social–emotional competencies are built by observing others modelling positive social–emotional behaviours | Developmental cascades theory (Masten and Ciccheti[61]); Fullan's theory of educational change[173] |
| Bullying Literature Project | Couch 2015[96]<br><br>Wang 2015[134] | The Bullying Literature Project integrates themes related to bullying into the children's literature used in a standard English curriculum. The Bullying Literature Project also provides an opportunity for children to role-model practical skills to address or avoid bullying. Ultimately, the intervention aims to ensure that children respond constructively to bullying, with reduced demonstration of aggressive and bullying behaviours | Bandura's social learning theory[172] |

**TABLE 22** Summary of intervention theories of change (*continued*)

| Intervention | Studies | Summary of theory of change | Existing theories drawn on |
|---|---|---|---|
| | | The authors suggest that the role-modelling component of the Bullying Literature Project intervention is what distinguishes it from other bibliotherapy-based school antibullying interventions. In particular, learning to walk away, ignore, talk it out or seek help are strategies emphasised to address bullying. Additional mechanisms of change that emerged include identifying feelings in oneself and others, using humour to deal with bullying and intervening in bullying rather than being a bystander | |
| DRACON | Malm 2007[119] | DRACON is an integrated school-based intervention that uses educational drama to teach conflict management to students, emphasising the importance of experiential learning. Specifically, educational drama can be a way of processing conflict, but students can also remain distant through the fictionalisation of the experience, enabling them to see alternative perspectives and outcomes. The authors specifically outline the following assumptions: (1) conflict literacy can be improved through drama, (2) school provides a platform for learning conflict handling and literacy skills, (3) early adolescence, when students experience high frequency of conflicts, is an important period in which to learn conflict resolution skills, (4) conflict resolution is culturally determined and (5) students need to be empowered to build up self-help and other capacities to solve conflict | Coleman's model of adolescent development,[142] Kolb's modes and characteristics of experiential learning[174] and Winslade and Monk's approaches to peer mediation[144] |
| English classes (no name) | Holcomb 1993[108] | This intervention uses learning theory to suggest that the repetition of health concepts in varied settings (in this case, English classes) will increase the retention of the concepts taught. Furthermore, if subject matter is integrated, it is thought to help students understand concepts across disciplines | Learning theory[145] and the notion of the impact of integrated subject matter[146] |
| Hashish and Marijuana | Zoller 1981[135] | This intervention is said to sit within the 'fourth generation' of preventative intervention, in which the approach of integration into chemistry classes with additional opportunities for participatory learning is interdisciplinary, integrates cognitive–informational and affective–behavioural domains, fosters students' value judgement and decision-making capacity and emphasises affective–behavioural–personal–societal components of learning to change behaviour | None mentioned |
| I-LST | Bechtel 2006[84] | I-LST is an intervention that hinges on the integration of health education into the regular curriculum. Lessons are designed to be relevant to students, interactive and with opportunities for repetition and reinforcement across subject areas. Provision of health education through this platform is therefore felt to optimise learning, leading students to internalise health messages and exhibit positive health behaviours<br><br>The primary mechanism of change appears to be integration of messaging around health topics (substance use in particular) into academic learning in an engaging way | Hunter's integration model, 'Instructional Theory into Practice',[175] which indicates that lessons must simultaneously address academic and life skills objectives |

**TABLE 22** Summary of intervention theories of change (*continued*)

| Intervention | Studies | Summary of theory of change | Existing theories drawn on |
|---|---|---|---|
| KAT | Segrott 2015[127] | The KAT intervention is intended to reduce alcohol use in young people. It has an anti-alcohol curriculum integrated into other subject areas to develop academic skills as well as practical skills that young people will need to avoid and decline the use of alcohol. Additionally, there is a parent evening in which students present to their parents what they have learned about alcohol. A DVD is made available to students and their parents to watch and discuss at home. Beyond the knowledge and practical skills that the classroom curriculum aims to develop, the crux of the intervention's intended mechanism of action lies in the development of pro-social bonds between students and adults, especially their parents. These positive bonds are felt to decrease the likelihood of negative behaviours and provide positive reinforcement of positive behaviours. Furthermore, parents who participate in the intervention are expected to reinforce and role-model its messages at home, further decreasing the likelihood of their children using alcohol if a norm against using alcohol has been established in their household | The underlying theory around which the intervention is based concerns two models: the social influence model[176] and the social development model.[177] In the former, 'children can be "inoculated" against social pressure to adopt undesirable behaviours such as drug and alcohol use' by training resistance skills, and in the latter, 'interactions with others, [result] in the formation of attachments which, if strong, can have a lasting effect on behaviour through supporting the acquisition of skills and influencing norms and values'[127] |
| LIFT | DeGarmo 2009[98]<br><br>Eddy 2000[100]<br><br>Eddy 2015[99]<br><br>Reid 1999[125]<br><br>Reid 2002[124] | LIFT has three components: classroom and playground components and a parent training intervention. The overall aim of LIFT is to reduce violence and delinquency in children. The theory of change specified does not relate to the curricular components of the intervention. LIFT refers to coercion theory in which negative reinforcement is hypothesised to drive the development of child problem behaviours, whereas positive reinforcement (or a reward paradigm) is used to reward positive behaviours. The authors propose that the most likely behaviour is what has been reinforced to lead to the desired outcome. Children who are in a cycle of negative reinforcement are likely to behave poorly, be disliked, go on to exhibit further negative behaviours later in life and engage with other antisocial youth. Thus, through each component of LIFT, there are aspects of problem-solving skill development to train children to use good behaviours and positively reinforce these behaviours | Coercion theory[178] |
| Peaceful Panels | Wales 2013[133] | The intervention uses the social reconstructionist art education paradigm, which suggests that art curricula can be used to foster critical thinking (owing to alignment with critical pedagogy) and a sense of social responsibility in students. Through this paradigm, 'students can be equipped to explore, inspect, deconstruct and reconstruct their world'[133] | Stuhr's 1994 paradigm of social reconstructionist art education[147] and critical pedagogy[148] |
| Positive Action | Beets 2008[86]<br><br>Beets 2009[87]<br><br>Flay 2009[55]<br><br>Flay 2010[102]<br><br>Lewis 2012[113] | Positive Action is a school-based intervention that teaches skills for successful learning and living. Students develop skills to better manage negative emotions and to exhibit a positive sense of self. In doing so, it is expected that their academic outcomes as well as their character will improve in the long term and that they will engage in fewer health-compromising behaviours such as violence and substance use | There are a considerable number of existing theories drawn on in Positive Action, including theory of self-concept,[179] theory of triadic influence,[180] multistage social learning theory,[181] Roger's diffusion of innovations framework,[182] Sheldon's notion of positive psychology,[183] self-esteem |

**TABLE 22** Summary of intervention theories of change (*continued*)

| Intervention | Studies | Summary of theory of change | Existing theories drawn on |
|---|---|---|---|
| | Malloy 2015[118] | The intervention is supported by theories of triadic influence (that behaviour is influenced by factors at different levels, proximal and distal) and self-concept and self-esteem (positive actions will reinforce a positive sense of self, which will further result in positive actions) | enhancement theory,[184] Moos's conceptualisation of social environments[185] and Weiner's theory around contextual factors such as organisational functioning and intervention implementation[186] |
| PATHS | Crean 2013[97]<br><br>Flay 2009[55]<br><br>Greenberg 2006[106]<br><br>Kusché 2012[112] | PATHS targets aggression, violence and substance use, which are assumed to result from maladaptation, neurobiological challenges and poor emotional intelligence manifested as lack of self-management. It operates on the assumption that emotional development precedes most forms of cognition (e.g. verbalising). Thus, when children do start to self-regulate their emotions through verbal self-regulation, that is a critical period to intervene to ensure that this self-regulation is positive. The intervention's theoretical framework therefore has four assumptions: (1) both the development of communication skills and self-control are important in understanding and discussing emotions, (2) socialisation practices affect children's emotional literacy, (3) emotional literacy is a central component of social problem-solving and (4) the school environment can be central in change as a fundamental ecology. An overall emphasis is on understanding emotions and engaging in social problem-solving<br><br>There are five theories that are drawn from to make up the conceptual model for this intervention, although the primary model is the A–B–C–D model, which emphasises integration of feelings and moods with emotion language, behaviour and cognitive understanding for improved social and emotional competence. The affective domain centres on a child's emotional intelligence to understand and control emotions. The behavioural domain refers to controlling behaviours and having appropriate behavioural skills. The cognitive domain seeks to build a child's analytical and local reasoning and to improve independent thinking. The dynamic domain aims to develop a child's self-esteem and personality in a healthy way. The second conceptual model, the eco-behavioural model, emphasises the role of context and promotes healthy classroom environments through infusion of intervention material wherever possible. The third model centres around neurobiology, suggesting that PATHS lessons build on neurological processing that helps the brain to develop in such a way that fosters social–emotional thinking and behaviours. The fourth model follows psychodynamic theory. The fifth model emphasises emotional intelligence and self-regulation | The authors draw from a number of existing theories, including the social information-processing theory of aggression,[187] Huesmann and Guerra's theory around latent social knowledge structures and the normalisation of aggression leading to more aggressive behaviour,[188] psychodynamic developmental theory,[189] cognitive social-learning theory,[190] attachment theory,[191] psychoanalytic theory[192] and concepts of emotional intelligence[193] |
| RHC | Brown 2005[92]<br><br>Catalano 2003[95] | RHC is guided by the social development model, which integrates social control and learning and differential association theories. This framework insists on early prevention (before antisocial behaviour sets in), the importance of individual, family and large social contexts such as school, community and peer influences (which become increasingly important in adolescence and lead to | Catalano and Hawkins' social development model[194] |

**TABLE 22** Summary of intervention theories of change (*continued*)

| Intervention | Studies | Summary of theory of change | Existing theories drawn on |
|---|---|---|---|
| | | inclusion of risk management rather than only rewards for pro-social behaviour) and the importance of recognising changing needs of a target population in terms of risk and protective factors. As such, four components were targeted by the intervention: providing opportunities for interaction with pro-social others (family, teachers and non-substance-using peers), academic, social and cognitive skills, rewarding pro-social behaviours and substance use avoidance. The combination of increased pro-social involvement with positive reinforcement is theorised to lead to better pro-social bonds to pro-social others. Once strong bonds are established, individuals will tend to behave in a manner consistent with the norms and values of the individuals and groups with whom they associate. In turn, stronger pro-social bonds support positive belief formation against antisocial behaviours such as adolescent substance use | |
| Roots of Empathy | Cain 2008[94] <br><br> Gordon 2003[105] <br><br> Hanson 2012[107] | The competencies that are believed to form the basis of effective social–emotional learning, and thus which Roots of Empathy aims to develop in students, include self-awareness (of one's own emotions, accurate self-perception and self-efficacy, spirituality), social awareness (empathy and respect for others), responsible decision-making (problem-solving, evaluation and reflection with a sense of personal responsibility), self-management (controlling impulses and stress, self-motivation and discipline, goal-setting and organisational skills) and relationship management (communication, building relationships and working well with others, conflict management and help-seeking and providing). A supportive classroom environment that rewards positive behaviour and develops and recognises social–emotional learning is thought to be essential to the development of these competencies. Thus, through Roots of Empathy, with increased social–emotional learning, increased engagement in learning overall is theorised, resulting in students having a higher commitment to school and improved academic outcomes | The authors draw from the Collaboration for Academic, Social, and Emotional Learning (CASEL)'s framework in the design of the curriculum[195] |
| Steps to Respect | Brown 2011[93] <br><br> Frey 2009[103] | The Steps to Respect intervention has both classroom and school-wide components and is intended to reduce bullying and victimisation. The classroom component involves a literature-based curriculum that teaches students a number of bullying prevention skills, social–emotional skills and friendship skills. Additionally, children are taught how to engage as bystanders of bullying and how to report bullying. At the school level, there are policies established to change disciplinary action towards bullying, including increased monitoring of bullying and providing individual coaching for victims and perpetrators of bullying. The school climate overall is expected to become more positive and, together, these components are intended to change social norms around bullying | The intervention is aligned with the social-ecological model of bullying,[196] which suggests that bullying behaviour is shaped by contextual factors |

**TABLE 22** Summary of intervention theories of change (*continued*)

| Intervention | Studies | Summary of theory of change | Existing theories drawn on |
|---|---|---|---|
| The Gatehouse Project | Bond 2010[88]<br><br>Patton 2000[122]<br><br>Patton 2003[121] | The Gatehouse Project operates on the premise that both institutional and individual-focused intervention components come together to promote emotional and behavioural well-being in students in secondary schools. Healthy attachments, especially those with teachers and peers, are emphasised in the intervention. There are three highlighted actions: (1) building security and trust, (2) improving communication and social connectedness and (3) instilling positive regard through participation in school life. The intervention's conceptual framework suggests that the 'whole school' strategy involves skills and opportunities being introduced through the curriculum. These skills can lead to changes in the school (building security, communication and positive regard), which strengthens connectedness and a sense of belonging (within both schools and communities), leading to improved learning and emotional well-being<br><br>Attachment theory reinforces the understanding that social and learning environments may foster pro-social skills and provide opportunities for students to establish secure connections, good communication and positive feelings towards peers and teachers, which, in turn, creates a sense of belonging and connectedness, which ultimately leads to improved learning outcomes and emotional well-being | The authors refer to the World Health Organization's Health Promoting Schools Framework[197] as well as Green and Kreuter's 'ecological approach' to addressing environmental and social contexts alongside individual-level factors,[198] Bowlby's attachment theory,[191] social support theories[199] and the risk and protective factor framework[200] |
| Youth Matters | Jenson 2007[110] | The Youth Matters curriculum is guided by the social development model, which derives its thought from social control theory, social learning theory and differential association theory. This posits that four characteristics should be developed and will ultimately inhibit antisocial behaviours in children: (1) bonding (with family, the school and pro-social peers), (2) belief in shared values and norms (with the people the youth bond with), (3) external constraints (external barriers to antisocial behaviour) and (4) social, cognitive and emotional skills (which provide problem-solving skills for children to be used in social situations, including resisting negative impulses that lead to antisocial behaviour) | The authors draw primarily from the social development model,[194] although social control theory,[201] social learning theory[172] and differential association theory[202] are also briefly mentioned |

# Appendix 18 Included studies of process

**TABLE 23** Summary of studies of process

| Intervention | Description | Location | Targeted grade of schooling and approximate ages of participants | Process data collected on | Study/studies |
|---|---|---|---|---|---|
| 4Rs | A literacy-based social–emotional learning curriculum for elementary school students. There are two components: (1) a seven-unit, 21–35 lesson literacy-based curriculum in conflict resolution and social–emotional learning for children in primary school (to grade 5) and (2) intensive professional development for teachers | USA | Kindergarten to grade 5 (aged 6–11 years) | Fidelity and acceptability | Sung 2015[57] |
| DRACON | This intervention uses drama to develop cognitive understanding of conflict and bullying and to empower students to manage their own conflict, both personally and in the broader school community | Australia | Primary and secondary school students (aged 7–16 years) | Implementation, mechanisms of change, acceptability and context | O'Toole 2005[120] |
| English classes (no name) | Teachers were trained and, working in pairs in the summer, developed integrated health/English material, with a specific emphasis on the prevention of drug and alcohol use | USA | Grades 8 and 9 (aged 14 and 15 years) | Fidelity, acceptability, quality and mechanisms of change | Holcomb 1993[108] |
| Hashish and Marijuana | The goal of the curriculum is to develop scientific knowledge of hashish and marijuana and to strengthen students' problem-solving and decision-making skills through both didactic and participatory learning approaches | Israel | Upper secondary school (aged 17 and 18 years) | Implementation | Zoller 1981[135] |
| I-LST | A substance abuse prevention and competency curriculum that focuses on social and psychological protective factors affecting substance use. It is integrated into the existing subject curriculum by the classroom teachers | USA | Middle/junior high school (aged 13–15 years) | Fidelity, quality, dose and acceptability | Bechtel 2006[84] |
| KAT | The intervention aims to reduce drinking and antisocial behaviours in young people through a classroom curriculum, a parent evening and follow-up family activities | UK | Grades 5 and 6 (aged 11 and 12 years) | Acceptability and satisfaction | Rothwell 2011[126] |
| | | | | Fidelity, reach and mechanisms of change | Segrott 2015[127] |
| Peaceful Panels | Throughout art classes, students participated in antibullying lessons (from the Second Step intervention for eighth-grade students on empathy and communication in handling a grievance) and comic-making lessons. They then prepared artwork to demonstrate their understanding of how to resolve conflict | USA | Grades 8 and 9 (aged 14 and 15 years) | Acceptability and satisfaction | Wales 2013[133] |

| Intervention | Description | Location | Targeted grade of schooling and approximate ages of participants | Process data collected on | Study/studies |
|---|---|---|---|---|---|
| Positive Action | Positive Action is a social–emotional and character development intervention aimed at encouraging positive behaviours through positive thoughts and actions. Lessons cover six units: self-concept, positive actions for mind and body, positive social–emotional actions, managing oneself, being honest with oneself and continually improving oneself | USA | Kindergarten to grade 12 (aged 6–18 years) | Coverage and acceptability | Beets 2007[85] |
| | | | | Acceptability and satisfaction | Beets 2008[86] |
| | | | | Implementation, fidelity, dosage and quality | Malloy 2015[118] |
| PATHS | An intervention to reduce conflict by improving students' social–emotional and thinking skills through a curriculum, the establishment of a positive classroom environment and generalised positive social norms throughout the school environment | USA | Kindergarten to grade 5 (aged 6–11 years) | Quality, coverage (dose) and context | Ransford 2009[123] |
| Roots of Empathy | An intervention that brings a visiting baby and their parent into a classroom as a springboard for learning empathy. Students learn messages of social inclusion, respect, how to build consensus, how to contribute to a safe and caring classroom, and develop emotional literacy | Australia | Grades 1 to 9 (aged 7–15 years) | Implementation, mechanisms of change and acceptability | Cain 2008[94] |
| | | Canada, UK | | Implementation and context | Hanson 2012[107] |
| Steps to Respect | This is an antibullying intervention with both school-wide and classroom components. The school-wide components create new disciplinary policies for bullying and improve monitoring of and intervention in bullying. Classroom curricula positive social norms and improve social–emotional skills for better engagement with bullying | USA | Grades 3 to 6 (aged 9–12 years) | Fidelity, context and acceptability | Low 2014[117] |
| The Gatehouse Project | Through teaching a curriculum and establishing a school-wide adolescent health team, the Gatehouse Project aims to build a sense of security and trust in students, enhance skills and opportunities for good communication and build a sense of positive regard through participation in school life | Australia | Grade 8 (aged 14 years) | Coverage, quality and mechanisms of change | Bond 2001[89] |

# Appendix 19 Characteristics and quality appraisal of process evaluations

TABLE 24 Summary of process evaluation characteristics and quality appraisal

| Details | Characteristics | |
|---|---|---|
| **4Rs: Sung 2015[57]** | | |
| Stated aim of study | | To describe how 4Rs has been implemented in an inner-city public elementary school and to better understand teachers' experiences with its implementation |
| Methods | Data collection | Three semi-structured interviews [with (1) teachers (to assess implementation and experiences with 4Rs), (2) the principal and (3) the intervention developer], classroom observation, study of field notes, 'reflective conversations' and document review (about the school and its 4Rs intervention) were used to provide contextual data and school characteristics |
| | Data analysis | Iterative and combined with data collection and constant comparative analysis. Data were read carefully and open coded. Data were then re-coded to identify emergent themes |
| Participants | Country | USA |
| | Actual sample | Teachers |
| | Sample number | Four teachers, one principal, one intervention developer |
| | Age | One teacher of the first grade, one of the second grade, one of the third grade and another of the fifth grade. Their number of years at the school ranged from 1 to 25 years. Ages not given |
| | Sex | All female |
| | Ethnicity | Not stated |
| | SES | Not stated |
| Intervention | Description | A literacy-based social–emotional learning curriculum for elementary school students |
| | Target population | Kindergarten to grade 5 (aged 6–11 years) |
| | Theory | There are guiding theoretical assumptions around students' social–emotional development. First, students are expected to develop social–emotional competencies by reacting to norms in the school environment that are changed to emphasise more positive social behaviours. Second, they develop social–emotional competencies through environmental consequences of their actions, particularly reinforcement for positive behaviours (e.g. praise or rewards for helping others). Third, they develop these competencies through observing others who model the desired skills and characteristics as well as through practising them in social contexts |
| | Provider(s) | Teachers |
| | Training | 25 hours of training followed by ongoing coaching of teachers with a minimum of 12 contacts in one school year |
| | Setting | Classrooms |
| | Content | There are two components – (1) a seven-unit, 21–35 lesson literacy-based curriculum in conflict resolution and social–emotional learning for children in primary school (to grade 5) and (2) intensive professional development for teachers |
| | Length/intensity | 21–35 lessons |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | No | Purposeful sampling but of only high-performing classrooms |
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | Different instruments piloted and used; findings triangulated |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | Yes | Author verified data through 'reflexive conversations' and member checking |
| Were the findings of the study grounded in/supported by data? | Yes | Clear results followed methods |
| Was there good breadth and depth achieved in the findings? | Yes | Very comprehensive data collected from a number of classrooms |
| Were the perspectives of young people privileged? | No | |

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability | Medium | Selection of only high performers limits transferability of findings |
| Usefulness | High | Detailed information about implementation provided |

| Details | Characteristics | |
|---|---|---|
| *DRACON: O'Toole 2005[120]* | | |
| Stated aim of study | To describe emergent outcomes following nine rounds of action research and 'tell the story' | |
| Methods | Data collection | Interviews with teachers, students, principals and education officers; class questionnaires; focus group discussions with teachers and students; observation (by both teachers and external observers); and examination of video footage from classes |
| | | Interpretive analysis based on 'corroborative statement' and ongoing interpretation within the scope of action research |
| | Data analysis | |
| Participants | Country | Australia |
| | Actual sample | 15 secondary schools have taken part and 16 linked primary schools. Between one and five classrooms were involved in each school for both secondary and primary schools |
| | Sample number | Unclear. From participating schools, authors state that 'nearly all' teachers and a 'proportion' of students, from five to the whole class, participated |
| | Age | Not stated |
| | Sex | Not stated |
| | Ethnicity | Not stated |
| | SES | Not stated |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Details | Characteristics | |
|---|---|---|
| Intervention | Description | This intervention uses drama to develop cognitive understanding of conflict and bullying and to empower students to manage their own conflict, both personally and in the broader school community |
| | Target population | Primary and secondary school students (aged 7–16 years) |
| | Theory | A basic premise of the project has been that drama is used with the aim not of direct conflict resolution but to provide 'conflict literacy' in the hope that the learners will be able to utilise this later in dealing with their own conflicts. The Brisbane team believe that the two conditions necessary for drama (empathy and intellectual distance embodied in fiction) do not exist in a real conflict situation |
| | Provider(s) | Teachers and students (as peer educators) |
| | Training | Not stated |
| | Setting | Classrooms |
| | Content | Conflict literacy is taught through 'enhanced forum theatre' and other drama techniques |
| | Length/intensity | Unclear, but at least 'several weeks' per school year |

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | No | No detail provided |
| Were steps taken to minimise bias and error/increase rigour in data collection? | No | No detail provided |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | No | No detail provided |
| Were the findings of the study grounded in/supported by data? | No | No quotations present to support qualitative data and no links to questionnaire data |
| Was there good breadth and depth achieved in the findings? | No | Good breadth of findings, but limited depth |
| Were the perspectives of young people privileged? | Yes | |

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | Low | Limited data on methods and links to results |
| Usefulness of findings | Low | Limited detail on implementation |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Details | Characteristics | |
|---|---|---|
| **English Classes: Holcomb 1993[108]** | | |
| Stated aim of study | Not stated, although study describes experiences implementing an integrated health curriculum into English classes | |
| Methods | Data collection | Monthly reports from teachers on progress/failures, teacher interviews each year and a concluding teacher survey that built on the interview responses |
| | Data analysis | Not stated |
| Participants | Country | USA |
| | Actual sample | Teachers |
| | Sample number | Eight teachers |
| | Age | Not stated |
| | Sex | Not stated |
| | Ethnicity | Not stated |
| | SES | Teaching experience ranged from 3–15 years, with an average of 11.6 years. No SES factors stated |
| Intervention | Description | Teachers were trained and, working in pairs in the summer, developed integrated health/English material, with a specific emphasis on the prevention of drug and alcohol use |
| | Target population | Grades 8 and 9 (aged 14 and 15 years) |
| | Theory | Learning theory suggests that students learn best through repetition of material, presented in classes that they view as important. Furthermore, integrated subject matter helps students understand concepts across disciplines |
| | Provider(s) | Teachers |
| | Training | Teachers were briefed about their responsibilities in implementation and reporting in a series of meetings. Resources to supplement their knowledge of health topics were provided routinely |
| | Setting | Classrooms |
| | Content | Health topics and activities were infused into required English classes to increase students' awareness and practice of healthy behaviours |
| | Length/intensity | Not stated |

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | No | No detail provided |
| Were steps taken to minimise bias and error/increase rigour in data collection? | No | No detail provided |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | No | No detail provided |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| Were the findings of the study grounded in/supported by data? | Yes | Survey results followed clearly; qualitative results presented without supporting quotations |
| Was there good breadth and depth achieved in the findings? | Yes | Mixed methods enabled exploration of both breadth and depth |
| Were the perspectives of young people privileged? | No | |

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | Low | Limited detail on the rigour of methods used |
| Usefulness of findings | Medium | One of the few studies in which integration was core to the study's design and some good detail around implementation is provided |

| Details | Characteristics | |
|---|---|---|
| *Hashish and Marijuana: Zoller 1981[135]* | | |
| Stated aim of study | To understand student drug use and to gain preliminary insights of the programme's implementation, including feedback from students and teachers to identify areas to improve | |
| Methods | Data collection | Ongoing observation of class and students' behaviours, interviews of teachers, and a pre- and post-pilot student questionnaire around drug use |
| | Data analysis | A single, weighted attitudinal score was created to assess student changes in response to the curriculum |
| Participants | Country | Israel |
| | Actual sample | Students and teachers |
| | Sample number | Not stated |
| | Age | Not stated |
| | Sex | Not stated |
| | Ethnicity | Not stated |
| | SES | Not stated |
| Intervention | Description | The goal of the curriculum is to develop scientific knowledge of hashish and marijuana and to strengthen students' problem-solving and decision-making skills through both didactic and participatory learning approaches |
| | Target population | Upper secondary school (aged 17 and 18 years) |
| | Theory | A 'fourth generation' of preventative programme, in which the approach is interdisciplinary, integrates cognitive–informational and affective–behavioural domains, fosters students' value judgement and decision-making capacity and emphasises affective, behavioural, personal and societal components of learning to change behaviour |
| | Provider(s) | Teachers |
| | Training | Not stated |
| | Setting | Classrooms |

continued

TABLE 24 Summary of process evaluation characteristics and quality appraisal (*continued*)

| Details | Characteristics | |
|---|---|---|
| | Content | Lessons around hashish and marijuana take place in a chemistry class, teaching the chemical aspects of the drugs and also addressing behaviours around their misuse |
| | Length/intensity | Not stated |

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | No | No detail provided |
| Were steps taken to minimise bias and error/increase rigour in data collection? | No | No detail provided |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | No | No detail provided |
| Were the findings of the study grounded in/ supported by data? | No | Scant data were provided and it was unclear how these were produced |
| Was there good breadth and depth achieved in the findings? | No | Minimal findings reported |
| Were the perspectives of young people privileged? | Yes | |

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | Low | Poor reporting of methods and minimal results |
| Usefulness of findings | Low | Lack of detail in findings restricted the use of this study |

| Details | Characteristics | |
|---|---|---|
| *I-LST: Bechtel 2006[84]* | | |
| Stated aim of study | To determine the dosage and fidelity with which the intervention was implemented, student and teacher receptiveness to materials and teacher satisfaction with training and support | |
| Methods | Data collection | Lesson plan assessment by teachers using the classroom implementation evaluation, on-site teacher observation and a year-end focus group |
| | Data analysis | Not stated |
| Participants | Country | USA |
| | Actual sample | Teachers |
| | Sample number | Not stated |
| | Age | Not stated |
| | Sex | Not stated |
| | Ethnicity | Not stated |
| | SES | Not stated |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Details | Characteristics | |
|---|---|---|
| Intervention | Description | A substance abuse prevention and competency curriculum integrated into the existing subject curriculum by the classroom teachers. Curriculum focuses on social and psychological protective factors affecting substance use |
| | Target population | Middle/junior high school (aged 13–15 years) |
| | Theory | Constructivist approach to learning where natural links among subjects are identified. Learning is expected to occur when information is embedded in meaningful contexts, with repetition across disciplines and opportunities for application are provided |
| | Provider(s) | Teachers supported by external specialists |
| | Training | Teachers were trained in the regular life skills curriculum with additional information on integration and embedding lessons in different disciplines. Teachers supported one another in multidisciplinary teams |
| | Setting | Classrooms |
| | Content | Topics: self-image and self-improvement; decision-making; smoking, marijuana, and alcohol myths and realities; smoking and biofeedback; advertising awareness; coping with anxiety; communication skills; social skills; and assertiveness |
| | Length/intensity | No set number of lessons as material is expected to be integrated wherever appropriate |

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | No | No detail provided |
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | Multiple methods and instruments used; findings triangulated |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | No | No detail provided |
| Were the findings of the study grounded in/supported by data? | No | No primary data provided, only authors' accounts of the data |
| Was there good breadth and depth achieved in the findings? | Yes | Different aspects of implementation explored from students, teachers and administrators |
| Were the perspectives of young people privileged? | No | |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | Low | Poor reporting of methods and minimal results |
| Usefulness of findings | Medium | Paper provides interesting insights and is the only one to compare with non-integrated curriculum implementation, but detail on methods is lacking |

| Details | Characteristics | |
|---|---|---|
| *KAT: Rothwell 2011[126]* | | |
| Stated aim of study | An examination of the program's development and implementation | |
| Methods | Data collection | Document analysis, semi-structured in-depth interviews (with intervention personnel, head teachers and teachers delivering the intervention), observation of classroom preparation and KAT family events, focus-group discussions with children, interviews and focus groups with parents who attended the KAT family events and questionnaires from parents whose children were involved in the classroom preparation |
| | Data analysis | Thematic content analysis carried out in NVivo 8. Coding was informed by a framework that was developed based on topics in interviews and after document analysis |
| Participants | Country | Wales |
| | Actual sample | Working group members, head teachers, programme organiser, assistant programme organiser, KAT DVD producer, organiser of the Australian P-KAT intervention, children and parents |
| | Sample number | Six working group members, two head teachers, the 'program organiser' and their assistant, the KAT DVD producer, the organiser of the Australian P-KAT intervention, 54 children, 12 parents attending KAT events and 110 other parents |
| | Age | Students were aged 9–11 years; not stated for others |
| | Sex | Not stated |
| | Ethnicity | Most students were white British; not stated for others |
| | SES | Schools were generally of lower SES and with lower attendance rates than average; not stated for others |
| Intervention | Description | The intervention aims to reduce drinking and antisocial behaviours in young people through a classroom curriculum, a parent evening and follow-up family activities |
| | Target population | Grades 5 and 6 (aged 11 and 12 years) |
| | Theory | The social influence model, which suggests that students will be less likely to partake in risky behaviours if they have strong bonds with pro-social other |
| | Provider(s) | Teachers, supported by the KAT organiser |
| | Training | Not stated |
| | Setting | Classrooms, schools and children's homes |
| | Content | A classroom component in which students prepare presentations for the parent evening by developing skills across the normal curriculum around alcohol use effects and consequences, a 'fun' evening with parents and a follow-up DVD for children and parents to watch together |
| | Length/intensity | Not stated |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | No | No detail provided |
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | Multiple methods used at different data points to ensure comprehensive perspectives |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | Yes | Comparative coding used to refine analytical framework |
| Were the findings of the study grounded in/supported by data? | Yes | Clear results followed methods |
| Was there good breadth and depth achieved in the findings? | No | Good depth around acceptability; limited detail on other aspects of implementation |
| Were the perspectives of young people privileged? | Yes | |

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | Medium | Insufficient detail to determine possible bias introduced in sampling, but data collection and analysis seem appropriate |
| Usefulness of findings | Low | Nothing about the integration of academic and health curricula in findings |

| Details | Characteristics | |
|---|---|---|
| *KAT: Segrott 2015[127]* | | |
| Stated aim of study | To assess quality and fidelity of programme delivery; to refine the programme's logic model and theory of change and to determine the programme's reach | |
| Methods | Data collection | Classroom observations, semi-structured interviews (with parents, intervention deliverers and head teachers) and focus group discussions |
| | Data analysis | Analysis of process data unclear, but used to develop the theoretical framework and explore implementation fidelity and acceptability of KAT |
| Participants | Country | Wales |
| | Actual sample | Schools, children and parents |
| | Sample number | 257/418 children and 27 parents/carers |
| | Age | Not stated |
| | Sex | Not stated |
| | Ethnicity | Not stated |
| | SES | Not stated |
| Intervention | See Rothwell 2011[126] | |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | No | No detail provided |
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | Comprehensive qualitative data were collected |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | Yes | Data were triangulated; constant comparison of data was done; authors increased validity of instruments |
| Were the findings of the study grounded in/supported by data? | Yes | Actual quotations and results from process evaluation were limited |
| Was there good breadth and depth achieved in the findings? | Yes | Data were collected on many aspects of implementation |
| Were the perspectives of young people privileged? | No | |

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | Medium | A lack of data on methods makes reliability impossible to ascertain |
| Usefulness of findings | Medium | This study has interesting findings but it would be better to see them grounded in primary data |

| Details | Characteristics | |
|---|---|---|
| *Peaceful Panels: Wales 2013*[133] | | |
| Stated aim of study | To explore and describe how comic-making can be used to teach antibullying | |
| Methods | Data collection | The comics and portfolios created by students themselves were collected and analysed for evidence of understanding of violence prevention. Otherwise, interviews, observation notes, videotapes and photographs were used. A final post-experience survey was also carried out |
| | Data analysis | Data were analysed for themes, commonalities, trends and divergent cases. Notes were taken on emerging themes and a matrix was made to link evidence to main themes. Visual data were also synthesised into a comic created by the author |
| Participants | Country | USA |
| | Actual sample | Students |
| | Sample number | 56 students |
| | Age | Not stated |
| | Sex | Not stated |
| | Ethnicity | Not stated |
| | SES | Not stated |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Details | Characteristics | |
|---|---|---|
| Intervention | Description | Throughout art classes, students received antibullying and comic-making training. They then prepared artwork to demonstrate their understanding of how to resolve conflict |
| | Target population | Grades 8 and 9 (aged 14 and 15 years) |
| | Theory | Social reconstructionist art education paradigm, which states that art curricula can be used to foster critical and a sense of social responsibility in students |
| | Provider(s) | Teacher |
| | Training | Not stated |
| | Setting | Classrooms |
| | Content | Students participated in antibullying lessons (from the Second Step intervention for eighth-grade students on empathy and communication in handling a grievance) and comic-making lessons |
| | Length/intensity | Not clear (some art classes throughout the school year) |

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | No | Convenience sample drawn from the author's classroom |
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | A range of methods used to collect data and an independent peer audited the author's methods |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | Yes | Author employed reflexivity, debriefs with peers and member checking to increase robustness |
| Were the findings of the study grounded in/supported by data? | Yes | Clear results followed methods |
| Was there good breadth and depth achieved in the findings? | Yes | Considerable detail on a number of implementation factors reported |
| Were the perspectives of young people privileged? | Yes | |

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | Medium | Convenience samples in the opinion of the study team meant that the analyses were of a medium level of robustness |
| Usefulness of findings | Medium | Detailed information about implementation processes but limited information about influencing factors |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Details | Characteristics | |
|---|---|---|
| **Positive Action: Beets 2007[85]** | | |
| Stated aim of study | To determine factors influencing the implementation fidelity of Positive Action | |
| Methods | Data collection | Year-end student survey about the Positive Action intervention, perceptions around substance use and, in years 3 and 4, the perceived benefits of the Positive Action intervention |
| | Data analysis | Descriptive statistics generated and presented; *t*-tests comparing perceptions and practices of behaviours were carried out. A measure of student responsiveness to Positive Action was generated and logistic regression models were used to produce ORs for perception and behaviour variables by level of responsiveness |
| Participants | Country | USA |
| | Actual sample | Students |
| | Sample number | 1399–1574 students, with an average of 1523 throughout 4 years, including at baseline; 458 fifth-graders were specifically asked about substance use |
| | Age | Not stated |
| | Sex | 50% male, 50% female |
| | Ethnicity | 25% Hawaiian (full or part), 15% Filipino, 16% white, 6% Japanese or other |
| | SES | Not stated |
| Intervention | Description | Positive Action is a social–emotional and character development intervention aimed at encouraging positive behaviours through positive thoughts and actions |
| | Target population | Kindergarten to grade 12 (aged 6–18 years) |
| | Theory | Purkey's[179] theory of self-concept (suggests that positive self-worth leads to positive action, which reinforces self-worth) and Flay *et al.*'s[180] theory of triadic influence (suggests that positive school climate and good relationships with teachers and parents will develop positive self-concept in students) |
| | Provider(s) | Teachers |
| | Training | 3–4 hours of teacher training in the first year and 1–2 hours in subsequent years with booster sessions and mini-conferences (with other participating schools) each year |
| | Setting | Classrooms |
| | Content | Lessons cover six units: self-concept, positive actions for mind and body, positive social–emotional actions, managing oneself, being honest with oneself and continually improving oneself |
| | Length/intensity | 140 lessons per grade per year |

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | Yes | Sampling of schools was random and there was an attempt to reach a census of all participating students |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | Validated tools that collected data around a variety of measures of implantation were used |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | Yes | Data were analysed using statistically appropriate methods |
| Were the findings of the study grounded in/supported by data? | Yes | Clear results followed methods |
| Was there good breadth and depth achieved in the findings? | Yes | Various features of implementation were explored in detail. These were generated from a large sample of diverse students. However, no qualitative data |
| Were the perspectives of young people privileged? | Yes | |
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | High | Methods were appropriate, efforts were made to increase rigour and the findings and interpretations lead clearly from the methods used |
| Usefulness of findings | High | This paper gives good information about important aspects of implementation |

| Details | Characteristics | |
|---|---|---|
| ***Positive Action: Beets 2008[86]*** | | |
| Stated aim of study | To explore how relationships among teacher beliefs and attitudes towards a Positive Action, school administrative support, perceptions of school connectedness and school climate influenced programme implementation | |
| Methods | Data collection | End-of-year process evaluation survey (on implementation, attitude towards PA, perception of administrative support and beliefs about responsibility to teach social and character development) |
| | Data analysis | Structural equation modelling was used to create a model of the hypothesised relationships between tested constructs. Maximum likelihood estimation used to derive model estimates |
| Participants | Country | USA |
| | Actual sample | Teachers |
| | Sample number | Year 2: 171 teachers |
| | | Year 3: 191 teachers |
| | Age | Not stated |
| | Sex | Year 2: 83% female, 17% male |
| | | Year 3: 90% female, 10% male |
| | Ethnicity | Year 2: 41% Japanese, 21% other Asian descent, 18% white |
| | | Year 3: 36% Japanese, 25% white, 12% full- or part-Hawaiian |
| | SES | Not stated |
| Intervention | See Beets 2007[85] | |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | No | Census of teachers attempted without success and no explanation provided |
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | High reliability of tools used |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | Yes | Analysis were appropriate and data were entirely quantitative |
| Were the findings of the study grounded in/supported by data? | Yes | Clear results followed methods |
| Was there good breadth and depth achieved in the findings? | Yes | Good range of process measures covered in considerable depth |
| Were the perspectives of young people privileged? | No | |

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | High | Study was well-conducted and statistically robust |
| Usefulness of findings | High | Useful discussion of key implementation factors including the perspectives of implementers |

| Details | Characteristics | |
|---|---|---|
| *Positive Action: Malloy 2014[118]* | | |
| Stated aim of study | To determine if and how teachers' perceptions of school organisational climate predict the dose and quality of Positive Action implementation | |
| Methods | Data collection | Teacher-completed unit implementation reports (to assess PA dose and quality) and a teacher work climate survey (to assess teacher–teacher affiliation, school innovation, participatory decision-making and attitude towards social and character development) |
| | Data analysis | Stata® 12.1 (StataCorp LP, College Station, TX, USA) was used for all analyses, including the generation of descriptive statistics, bivariate correlation and missing data reports. Scale analysis of three school climate scales and model estimation of relationships between constructs using the least squares regression were done |
| Participants | Country | USA |
| | Actual sample | Teachers |
| | Sample number | 46/63 possible teachers |
| | Age | Not stated |
| | Sex | 79% female, 21% male |
| | Ethnicity | 43% black, 45% white, 13% Hispanic |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Details | Characteristics | |
|---|---|---|
| | SES | 14.7 mean years of teaching experience (1–34) and 47% have a graduate degree; no other SES details provided |
| Intervention | See Beets 2007[85] | |

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | Yes | Relevant sampling criteria used with a very high response rate |
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | Multiple data sources used and triangulated |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | Yes | Analytical approach was appropriate and robust |
| Were the findings of the study grounded in/supported by data? | Yes | Clear results followed methods |
| Was there good breadth and depth achieved in the findings? | Yes | Multiple data sources provided information about many aspects of implementation, in detail, with description of relationships between these |
| Were the perspectives of young people privileged? | No | |

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | Medium | Although methods were robust, there was no qualitative data to answer the more useful 'why' questions, particularly behind the relationships between implementation factors |
| Usefulness of findings | High | Good data provided around multiple aspects of implementation |

| Details | Characteristics | |
|---|---|---|
| *PATHS: Ransford 2009[123]* | | |
| Stated aim of study | To determine the influence of teachers' psychological experiences and perceptions of curriculum supports on PATHS implementation | |
| Methods | Data collection | Online survey assessing teacher burnout, efficacy, available support and implementation quality |
| | Data analysis | Descriptive statistics were generated and a set of regression models were developed between teacher variables and implementation outcomes. A second series of regression models were developed to test moderation hypotheses, including main effects and an interaction term |
| Participants | Country | USA |
| | Actual sample | Teachers |
| | Sample number | 133/156 possible teachers (multigrade, 11; grade 1, 20; grade 2, 17; grade 3, 19; grade 4, 9; grade 5, 16) |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Details | Characteristics | |
|---|---|---|
| | Age | Average age 40 years |
| | Sex | 92% female, 8% male |
| | Ethnicity | 86.2% white, 9.2% black, 1.8% Asian, 2.8% multiracial/other |
| | SES | 25% < 1 year in PATHS, 43% 1–3 years in PATHS, 31.6% ≥ 3 years in PATHS |
| | | Average length of time at school 14.7 years |
| | | No SES details provided |
| Intervention | Description | An intervention to reduce conflict by improving students' social–emotional and thinking skills through a curriculum, the establishment of a positive classroom environment and generalised positive social norms throughout the school environment |
| | Target population | Kindergarten to grade 5 (aged 6–11 years) |
| | Theory | Not stated |
| | Provider(s) | Teachers |
| | Training | Ongoing mentoring and coaching (weekly in first year of implementation, bi-weekly thereafter) following initial programme training |
| | Setting | Classrooms |
| | Content | Lessons are designed to improve students' social–emotional and thinking skills and to facilitate a positive classroom environment |
| | Length/intensity | 2–3 lessons per week |

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | Yes | Sufficient detail provided; high (85%) response rate |
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | Alpha-reliability coefficients acceptable and provided; other measures of validity lacking |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | Yes | Data analyses were appropriate |
| Were the findings of the study grounded in/supported by data? | Yes | Clear results follow methods |
| Was there good breadth and depth achieved in the findings? | No | Study limited to teachers' psychological factors |
| Were the perspectives of young people privileged? | No | |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | High | This is a well-conducted study |
| Usefulness of findings | Medium | Although methodologically sound, comprehensive results are lacking |

| Details | Characteristics | |
|---|---|---|
| *Roots of Empathy: Cain 2008[94]* | | |
| Stated aim of study | To investigate the impact of social–emotional learning training on teaching practice | |
| Methods | Data collection | An introductory questionnaire (on teacher knowledge of social and emotional learning, teaching experiences and their perceptions of Roots of Empathy) and follow-up interviews |
| | Data analysis | Authors suggest this was a phenomenological study. A thematic analysis of questionnaires and interview transcripts was done |
| Participants | Country | Australia |
| | Actual sample | Teachers |
| | Sample number | Eight teachers |
| | Age | Not stated |
| | Sex | Not stated |
| | Ethnicity | Not stated |
| | SES | Not stated |
| Intervention | Description | An intervention that brings a visiting baby and their parent into a classroom as a springboard for learning empathy |
| | Target population | Grades 1 to 9 (aged 7–15 years) |
| | Theory | Not clear, although authors refer to a framework of teacher competencies that facilitate social and emotional learning through teacher role-modelling, including self-awareness, social awareness, responsible decision-making, self-management and relationship management |
| | Provider(s) | Teachers, external Roots of Empathy facilitators |
| | Training | A teacher curriculum and manual are available; Roots of Empathy teachers are mentored throughout their first year by facilitators |
| | Setting | Classrooms |
| | Content | Students learn messages of social inclusion, respect, how to build consensus, how to contribute to a safe and caring classroom and develop emotional literacy |
| | Length/intensity | Nine baby visits per year, plus additional classroom work |

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | No | No detail provided |

*continued*

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| Were steps taken to minimise bias and error/increase rigour in data collection? | No | No detail provided |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | No | No detail provided |
| Were the findings of the study grounded in/supported by data? | Yes | Clear results followed methods |
| Was there good breadth and depth achieved in the findings? | No | Breadth around implementation from a teacher perspective, but little depth |
| Were the perspectives of young people privileged? | No | |
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | Low | A lack of methodological detail make trustworthiness questionable |
| Usefulness of findings | Medium | Useful data on some aspects of implementation provided, but lacking methodological rigour |

| Details | Characteristics | |
|---|---|---|
| *Roots of Empathy: Hanson 2012[107]* | | |
| Stated aim of study | | To explore how teacher's beliefs and emotions impact Roots of Empathy implementation |
| Methods | Data collection | Pre- and post-test teachers questionnaires (to assess background, beliefs around emotions, beliefs around social–emotional learning and responsibility to teach it and Roots of Empathy implementation where applicable) |
| | Data analysis | Data were analysed using IBM SPSS Statistics version 19 (IBM Corporation, Armonk, NY, USA). Descriptive statistics were generated and correlation coefficients assessing relationships among variables were computed |
| Participants | Country | Canada; the Isle of Man |
| | Actual sample | Teachers |
| | Sample number | 38 teachers (19 teachers hosting Roots of Empathy) from 16 large urban schools in Western Canada and 20 teachers instructing primary grades from 18 schools spread across the Isle of Man |
| | Age | Not stated |
| | Sex | 78–100% teachers were female |
| | Ethnicity | 70–90% were of Western European descent |
| | SES | 78–100% had an undergraduate teaching degree; 33%–50% had > 11 years' teaching experience; 52% had received some social–emotional learning training; no SES details provided |
| Intervention | See Cain 2008[94] | |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | Yes | Participants were from an ongoing RCT |
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | Reliability of instruments was good |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | Yes | Data were merged to increase study power |
| Were the findings of the study grounded in/supported by data? | Yes | Constructs were well-defined and studied. Slight bias to Canadian results |
| Was there good breadth and depth achieved in the findings? | No | Lack of qualitative data |
| Were the perspectives of young people privileged? | No | |

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | High | This is a methodologically rigorous study |
| Usefulness of findings | Medium | Focus on teacher characteristics and implementation is valuable, but qualitative findings are limited |

| Details | Characteristics | |
|---|---|---|
| ***Steps to Respect: Low 2014[117]*** | | |
| Stated aim of study | To examine predictors and outcomes associated with Steps to Respect implementation | |
| Methods | Data collection | Teachers completed online implementation checklists and assessments of student behaviour as well as a pre- and post-test school environment survey (of school antibullying policies, school climate, staff bullying interventions and school bullying-related problems) and students completed pre- and post-test surveys (around bullying) |
| | Data analysis | Two-level hierarchical linear models to predict classroom implementation were created on pre-test student survey measures (averaged by classroom) and staff survey data and school characteristics. Three-level hierarchical models of implementation were also created, with teachers' intervention implementation measures predicting student post-test outcomes |
| Participants | Country | USA |
| | Actual sample | Students and teachers |
| | Sample number | Students: 1424/1515 |
| | | Teachers: 128/128 |
| | Age | Students: 7–11 years |
| | | Teachers: not stated |

*continued*

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Details | Characteristics | |
|---|---|---|
| | Sex | Students: 51% male, 49% female |
| | | Teachers: not stated |
| | Ethnicity | Students: 52% white, 7% black, 6% Asian-American, 35% other or mixed race, 42% Hispanic |
| | | Teachers: not stated |
| | SES | Not stated, but schools were 25% rural, 10% from small towns, 50% from suburban areas and 15% in mid-sized cities; 40% of students received free or reduced-price lunch |
| Intervention | Description | This is an antibullying intervention with both school-wide and classroom components. The school-wide components create new disciplinary policies for bullying and improve monitoring of and intervention in bullying. Classroom curricula promote positive social norms and improve social–emotional skills for better engagement with bullying |
| | Target population | Grades 3 to 6 (aged 9–12 years) |
| | Theory | A social–ecological model of bullying that suggests behaviour is shaped by many factors at the school, peer and individual levels |
| | Provider(s) | Teachers |
| | Training | 1-day on-site teacher and staff training in the curriculum and to develop effective disciplinary policies, to improve monitoring of students and to effectively intervene with students involved in bullying situations |
| | Setting | Classrooms and whole-schools |
| | Content | Recognition of bullying, empathy, friendship skills, communication skills and appropriate bystander responses to bullying |
| | Length/intensity | 11 lessons throughout the year, plus whole-school approaches throughout |

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | Yes | Participants are from an ongoing RCT; high response rate |
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | Questionnaire had high face validity and reliability |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | Yes | Data analysis were appropriate |
| Were the findings of the study grounded in/supported by data? | Yes | Although qualitative exploration was lacking |
| Was there good breadth and depth achieved in the findings? | No | Concepts explored were limited |
| Were the perspectives of young people privileged? | No | |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | High | This is a methodologically rigorous study |
| Usefulness of findings | Medium | Focus on teacher characteristics and implementation is valuable, but qualitative findings are limited |

| Details | Characteristics | |
|---|---|---|
| *The Gatehouse Project: Bond 2001[89]* | | |
| Stated aim of study | | To describe and account for how system-level changes have been made in schools through the process of capacity building within the Gatehouse Project |
| Methods | Data collection | Gatehouse Project adolescent health questionnaire with students (which assessed social connectedness, victimisation and measures of anxiety and depression); annual key informant interviews with administrators (on Gatehouse Project implementation); intervention field notes; semi-structured interviews with other teachers; and a school background audit |
| | Data analysis | Field notes, key informant interviews and interviews from the school background audits were analysed for themes |
| Participants | Country | Australia |
| | Actual sample | Not clear |
| | Sample number | Not clear, but included year-8 students in 26 schools and student welfare co-ordinators, principals and teachers |
| | Age | Average age of students is 13 years; not stated for others |
| | Sex | Not stated |
| | Ethnicity | Not stated |
| | SES | Not stated |
| Intervention | Description | Through teaching a curriculum and establishing a school-wide adolescent health team, the Gatehouse Project aims to build a sense of security and trust in students, enhance skills and opportunities for good communication and build a sense of positive regard through participation in school life |
| | Target population | Grade 8 (aged 14 years) |
| | Theory | The Health-Promoting Schools Framework, which suggests the schools' ethos and environment influence student behaviours |
| | Provider(s) | Not clear, but suggested to be principal or assistant principal, curriculum leader, student welfare co-ordinator, teachers and community representatives |
| | Training | The Centre for Adolescent Health liaison team provided training, resources and mentorship to schools |
| | Setting | Whole-school |
| | Content | Classroom curriculum unclear |
| | Length/intensity | Not stated |

**TABLE 24** Summary of process evaluation characteristics and quality appraisal (*continued*)

| Questions used to judge rigour and relevance | Reviewer judgement | Description |
|---|---|---|
| *Quality appraisal* | | |
| Were steps taken to minimise bias and error/increase rigour in sampling? | No | No detail provided |
| Were steps taken to minimise bias and error/increase rigour in data collection? | Yes | Multiple methods used to collect data at multiple points in the year |
| Were steps taken to minimise bias and error/increase rigour in data analysis? | No | No detail provided |
| Were the findings of the study grounded in/supported by data? | Yes | Clear results followed methods |
| Was there good breadth and depth achieved in the findings? | Yes | Multiple aspects of implementation were explored from multiple stakeholder perspectives |
| Were the perspectives of young people privileged? | No | |

| Variable | Level | Description |
|---|---|---|
| *Overall reliability and usefulness of findings* | | |
| Reliability of findings | Medium | More detail on methodological rigour would be required to make a fair assessment of robustness |
| Usefulness of findings | High | Very useful data provided around implementation characteristics |

P-KAT, Parents and Kids Acting Together.

# Appendix 20 Characteristics and quality appraisal of outcome evaluations

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention

| Details | Characteristics | |
|---|---|---|
| **4Rs: Aber 2011,[53] Jones 2010,[51] Jones 2011,[53] Jones 2010[52]** | | |
| Methods | Unit of randomisation | School |
| Participants | Country | USA |
| | Sample number | Intervention group: nine schools, 515 students |
| | | Control group: nine schools, 427 students |
| | School years included | Year 4 |
| | Sex | 51.2% female, 48.8% male |
| | Ethnicity | 41.1% African American, 45.6% Hispanic, 4.7% white, 8.6% other |
| | SES | 31% low parental education, 15.1% parental unemployment, 53.4% single-parent household, 61.8% living in poverty |
| Outcomes | At the end of first year: | |

At the end of first year:

- aggression, $\beta = 0.02$ (SE 0.05)
- academic skill (36-point scale), $b = -0.02$ (SE 0.08)
- maths, $b = -7.40$ (SE 5.69)
- reading, $b = -5.35$ (SE 4.96)

At the end of second year:

- aggression, $d = -0.21$ ($p < 0.05$)
- maths, $b = -4.35$ (SE 10.25)
- reading, $b = -0.99$ (SE 10.73)

| Details | Characteristics | |
|---|---|---|
| Intervention | Description | A literacy-based social–emotional learning curriculum for elementary school students |
| | Target population | Kindergarten to grade 5 (aged 6–11 years) |
| | Theory | There are guiding theoretical assumptions around students' social–emotional development. First, students are expected to develop social–emotional competencies by reacting to norms in the school environment that are changed to emphasise more positive social behaviours. Second, they develop social–emotional competencies through environmental consequences of their actions, particularly reinforcement for positive behaviours (e.g. praise or rewards for helping others). Third, they develop these competencies through observing others who model the desired skills and characteristics as well as through practising them in social contexts |

continued

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| | Provider(s) | Teachers |
| | Training | 25 hours of training followed by ongoing coaching of teachers with a minimum of 12 contacts in one school year |
| | Setting | Classrooms |
| | Content | There are two components: (1) a seven-unit, 21–35 lesson literacy-based curriculum in conflict resolution and social–emotional learning for children in primary school (to grade 5) and (2) intensive professional development for teachers |
| | Length/intensity | 21–35 lessons |
| | Control | Usual curriculum |

| Item | Reviewer judgement | Description |
|---|---|---|
| **Risk of bias** | | |
| Sequence generation: was the allocation sequence adequately generated? | Yes | Random number generator used |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | The study had generally low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | No | Study findings presented in an incomplete way at 3-year follow-up |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Yes | Analysis of impact of baseline characteristics carried out |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Nested design addressed by multilevel growth model |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Additional analysis to determine differences between consenting and non-consenting participants carried out |
| Suitable control group: did authors recruit a suitable control group? | Yes | Authors account for differences at baseline in analysis |

| Details | Characteristics | |
|---|---|---|
| **Bullying Literature Project: Couch 2015,[96] Wang 2015[134]** | | |
| Methods | Unit of randomisation | Classroom |
| Details of participants | Country | USA |
| | Sample number | Intervention group: four classrooms, 95 teacher reports, 90 students |
| | | Control group: three classrooms, 55 teacher reports, 42 students |
| | School years included | Years 4 and 5 |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| | Sex | 42.8% female, 57.2% male |
| | Ethnicity | 9.6% African American, 63.3% Hispanic, 9.0% white, 3.0% Asian, 4.2% other, 10.2% did not report |
| | SES | > 50% of students received free or reduced-cost lunch |
| Outcomes | At 1 month post intervention: | |
| | • physical bullying, student report<br>  ○ intervention group, mean 1.20, SD 0.44, $n = 90$<br>  ○ control group, mean 1.14, SD 0.36, $n = 42$; $p = 0.84$ | |
| | • physical bullying, teacher report<br>  ○ intervention group, mean 1.12, SD 0.47, $n = 95$<br>  ○ control group, mean 1.19, SD 0.47, $n = 55$; $p = 0.67$ | |
| | • physical bullying victimisation, student report<br>  ○ intervention group, mean 1.35, SD 0.54, $n = 90$<br>  ○ control group, mean 1.43, SD 0.66, $n = 42$; $p = 0.57$ | |
| | • physical bullying victimisation, teacher report<br>  ○ intervention group, mean 1.04, SD 0.23, $n = 95$<br>  ○ control group, mean 1.04, SD 0.21, $n = 55$; $p = 0.39$ | |
| Details of intervention | Description | This intervention aims to reduce bullying by introducing themes related to bullying through children's literature. It also provides an opportunity for children to role-model practical skills to address or avoid bullying. The Bullying Literature Project integrates themes related to bullying into the children's literature used in a standard English curriculum |
| | Target population | Grade 4 (aged 10 years) |
| | Theory | None identified, although the use of bibliotherapy in the context of this programme (with explicit instruction and modelling of behavioural strategies) is critical for skill development |
| | Provider(s) | Teachers |
| | Training | Not stated |
| | Setting | Classrooms |
| | Content | Lessons target normative beliefs about bullying and aggression, social–emotional skills, strategies to cope with bullying and bystander responses to bullying |
| | Length/intensity | Five 35- to 45-minute sessions over 5 weeks |
| | Control | Not stated |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Item | Reviewer judgement | Description |
|---|---|---|
| **Risk of bias** | | |
| Sequence generation: was the allocation sequence adequately generated? | Not clear | Method of randomisation not indicated |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Not clear | Randomised design presumably accounts for confounding, but not stated |
| Clustering: was clustering of participants accounted for in the analysis? | Not stated | |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Not stated | |
| Suitable control group: did authors recruit a suitable control group? | Not clear | No baseline equivalencies reported |
| **Details** | **Characteristics** | |
| **Bullying Literature Project–Moral Disengagement: Wang, 2017[136]** | | |
| Methods | Unit of randomisation | Classroom |
| Details of participants | Country | USA |
| | Sample number | Intervention group: two classrooms, 42 students |
| | | Control group: two classrooms, 42 students |
| | School years included | Year 4 |
| | Sex | 53.6% female, 46.4% male |
| | Ethnicity | 2.4% Asian, 3.6% white, 94% Hispanic |
| | SES | Not reported |
| Outcomes | At 1 week post-intervention: | |

At 1 week post-intervention:

- physical and emotional bullying, student report – $F_{1,80} = 0.83$; $p = 0.431$
- physical and emotional bullying victimisation, student report
    - intervention group, mean 1.76, SD 0.81 to mean 1.60, SD 0.66; $n = 42$
    - control group, mean 1.23, SD 0.38 to mean 1.38, SD 0.53; $n = 42$

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| Details of intervention | Description | This intervention aims to reduce bullying by introducing themes related to bullying through children's literature. It also provides an opportunity for children to role-model practical skills to address or avoid bullying. The Bullying Literature Project integrates themes related to bullying into the children's literature used within a standard English curriculum. Additionally, the version including moral disengagement discussed the role of moral disengagement in each lesson as well |
| | Target population | Grade 3 (aged 9 years) |
| | Theory | None identified, although the use of bibliotherapy within the context of this programme (with explicit instruction and modelling of behavioural strategies) is critical for skill development |
| | Provider(s) | Teachers |
| | Training | Not stated |
| | Setting | Classrooms |
| | Content | Lessons target normative beliefs about bullying and aggression, social–emotional skills, strategies to cope with bullying, and bystander responses to bullying, as well as moral disengagement |
| | Length/intensity | Five 35- to 45-minute sessions over 5 weeks |
| | Control | Not stated |

| Item | Reviewer judgement | Description |
|---|---|---|
| **Risk of bias** | | |
| Sequence generation: was the allocation sequence adequately generated? | Not stated | |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Yes | All outcomes reported as expected |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Yes | Balance described |
| Clustering: was clustering of participants accounted for in the analysis? | No | Classroom clustering not included |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | |
| Suitable control group: did authors recruit a suitable control group? | Yes | Baseline equivalency established |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| *I-LST training:[a] Smith 2004,[128] Vicary 2006[132]* | | |
| Methods | Unit of randomisation | School |
| Details of participants | Country | USA |
| | Sample number | Intervention group: three schools, 297 students |
| | | Control group: three schools, 201 students |
| | School years included | Starting year 8 |
| | Sex | 45.6% female, 54.4% male |
| | Ethnicity | 96.6% white |
| | SES | 32.5% qualify for free/reduced lunch |
| Outcomes | End of first intervention year: | |

End of first intervention year:

- alcohol (males), MD on frequency score –0.07, $t = -0.61$
- alcohol (females), MD on frequency score –0.23, $t = -1.54$
- binge drinking (males), MD on frequency score 0, $t = -0.01$
- binge drinking (females), MD on frequency score –0.25, $t = -2.11$
- drunkenness (males), MD on frequency score –0.08, $t = -0.82$
- drunkenness (females), MD on frequency score –0.08, $t = -0.72$
- cigarettes (males), MD on frequency score 0.16, $t = 1.08$
- cigarettes (females), MD on frequency score –0.4, $t = -2.01$
- smokeless tobacco (males), MD on frequency score 0.13, $t = 0.73$
- marijuana (males), MD on frequency score 0, $t = 0$
- marijuana (females), MD on frequency score –0.26, $t = -2.18$
- inhalants (males), MD on frequency score –0.09, $t = -1.28$
- inhalants (females), MD on frequency score –0.05, $t = -0.75$

End of second intervention year:

- alcohol (males), MD on frequency score 0.03, $t = 0.2$
- alcohol (females), MD on frequency score –0.13, $t = -0.94$
- binge drinking (males), MD on frequency score 0.04, $t = 0.32$
- binge drinking (females), MD on frequency score –0.09, $t = -1.01$
- drunkenness (males), MD on frequency score 0.08, $t = 0.79$
- drunkenness (females), MD on frequency score –0.07, $t = -0.69$
- cigarettes (males), MD on frequency score –0.17, $t = -0.72$
- cigarettes (females), MD on frequency score –0.61, $t = -2.39$
- smokeless tobacco (males), MD on frequency score 0.07, $t = 0.35$
- marijuana (males), MD on frequency score 0.05, $t = 0.31$
- marijuana (females), MD on frequency score –0.17, $t = -1.71$
- inhalants (males), MD on frequency score 0.09, $t = 1.86$
- inhalants (females), MD on frequency score 0.08, $t = 1.88$

End of third intervention year:

- alcohol (males), MD on frequency score 0.28, SE 0.22
- alcohol (females), MD on frequency score 0.02, SE 0.22
- binge drinking (males), MD on frequency score 0.08, SE 0.18
- binge drinking (females), MD on frequency score 0.08, SE 0.16
- drunkenness (males), MD on frequency score 0.07, SE 0.18
- drunkenness (females), MD on frequency score 0.08, SE 0.18
- cigarettes (males), MD on frequency score 0.2, SE 0.38
- cigarettes (females), MD on frequency score –0.18, SE 0.47
- marijuana (males), MD on frequency score 0.04, SE 0.23
- marijuana (females), MD on frequency score 0.08, SE 0.20

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| Details of intervention | Description | A substance abuse prevention and competency curriculum. It is integrated into the existing subject curriculum by the classroom teachers. Curriculum focuses on social and psychological protective factors affecting substance use |
| | Target population | Middle/junior high school (aged 13–15 years) |
| | Theory | Constructivist approach to learning where natural links among subjects are identified. Learning is expected to occur when information is embedded in meaningful contexts, with repetition across disciplines and opportunities for application are provided |
| | Provider(s) | Teachers supported by external specialists |
| | Training | Teachers were trained in the regular life skills curriculum with additional information on integration and embedding lessons in different disciplines |
| | | Teachers supported one another in multidisciplinary teams |
| | Setting | Classrooms |
| | Content | Topics: self-image and self-improvement; decision-making; smoking, marijuana and alcohol myths and realities; smoking and biofeedback; advertising awareness; coping with anxiety; communication skills; social skills; and assertiveness |
| | Length/intensity | No set number of lessons as material is expected to be integrated wherever appropriate |
| | Control | Standard LST programme or no programme at all |

| Item | Reviewer judgement | Description |
|---|---|---|
| ***Risk of bias*** | | |
| Sequence generation: was the allocation sequence adequately generated? | Not clear | Method of randomisation not indicated |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition and multiple imputation used |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol to confirm |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Yes | Models created separately to assess for differences in sex |

continued

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Item | Reviewer judgement | Description |
|---|---|---|
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Authors note that accounting for clustering did not affect analysis |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Authors used multiple imputations, including substantive and non-substantive variables, to enhance multiple imputation model |
| Suitable control group: did authors recruit a suitable control group? | Yes | No significant differences at baseline |

| Details | Characteristics | |
|---|---|---|
| **KAT: Segrott 2015[127]** | | |
| Methods | Unit of randomisation | School |
| Details of participants | Country | Wales |
| | Sample number | Intervention group: five schools, of which three remained in trial, 172 students |
| | | Control group: four schools, 152 students |
| | School years included | Years 5 and 6 (some year 4 owing to mixed-year classrooms) |
| | Sex | 51.9% female, 47.8% male (0.3% missing) |
| | Ethnicity | 81.3% white, 2.2% black or black British, 4.2% mixed race, 7.3% Asian or Asian British |
| | SES (family affluence scale) | 6.4% low, 33.0% medium, 52.8% high, 7.8% missing |
| Outcomes | | Post intervention (at end of first intervention year): |
| | | <ul><li>drunk in last 30 days, OR 1.5 (95% CI 0.4 to 5.8)</li><li>ever been drunk, OR 1.7 (95% CI 0.5 to 6.8)</li><li>ever had an alcoholic drink, OR 5.3 (95% CI 1.2 to 23.9)</li><li>had drink in last 30 days, OR 0.7 (95% CI 0.2 to 2.5)</li></ul> |
| Details of intervention | Description | The intervention aims to reduce drinking and antisocial behaviours in young people through a classroom curriculum, a parent evening and follow-up family activities |
| | Target population | Grades 5 and 6 (aged 11 and 12 years) |
| | Theory | The social influence model, which suggests that students will be less likely to partake in risky behaviours if they have strong bonds with pro-social other |
| | Provider(s) | Teachers, supported by the KAT organiser |
| | Training | Not stated |
| | Setting | Classrooms, schools and children's homes |
| | Content | A classroom component in which students prepare presentations for the parent evening by developing skills across the normal curriculum around alcohol use effects and consequences, a 'fun' evening with parents and a follow-up DVD for children and parents to watch together |
| | Length/intensity | Not stated |
| | Control | Usual curriculum |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Item | Reviewer judgement | Description |
|---|---|---|
| ***Risk of bias*** | | |
| Sequence generation: was the allocation sequence adequately generated? | Yes | Optimal allocation used to determine randomisation sequence |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol to confirm |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Not clear | Randomised design presumably accounts for confounding, but not stated |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Analysis used multilevel models |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Data collection instruments were validated |
| Suitable control group: did authors recruit a suitable control group? | Not clear | Significant differences in baseline indicated, but unclear if or how these were accounted for |
| **Details** | **Characteristics** | |
| ***Learning to Read in a Healing Classroom:*** [b] *Torrente 2015,* [137] *Aber 2017* [138] | | |
| Methods | Unit of randomisation | School |
| Details of participants | Country | Democratic Republic of the Congo |
| | Sample number | 20 districts, 33 schools (intervention group); 19 districts, 30 schools (control group); 3857 students overall |
| | School years included | Years 3–5 |
| | Sex | 48% female, 52% male |
| | Ethnicity | Not reported |
| | SES | Not reported |
| Outcomes | End of first intervention year: | |
| | ● student-report victimisation (physical and emotional), $d = -0.01$, SE 0.06 | |
| Details of intervention | Description | Teachers are supported to integrate social–emotional learning in literacy lessons, supported by a bank of lesson plans relating to reading and writing. Teachers additionally received substantial professional development, including 'teacher learning circles', and developed strategies to improve the learning environment |

<div align="right">continued</div>

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| | Target population | Years 3–5 |
| | Theory | The programme is led by the theory that altering school ecologies to teach social–emotional learning will lead to improvements in both academics and mental health |
| | Provider(s) | Teachers |
| | Training | Providers had continuous training throughout the year, including with 'learning circles' |
| | Setting | Classrooms |
| | Content | Social–emotional learning, literacy |
| | Length/intensity | Lessons delivered throughout the year |
| | Control | Waitlist |

| Item | Reviewer judgement | Description |
|---|---|---|
| **Risk of bias** | | |
| Sequence generation: was the allocation sequence adequately generated? | Yes | Drew allocation from a hat |
| Allocation concealment: was the allocation adequately concealed? | Yes | Allocation undertaken independently |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | No | High attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Yes | All expected outcomes reported |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Yes | Authors account for covariates |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Clustering included via multilevel model |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | |
| Suitable control group: did authors recruit a suitable control group? | Yes | Waitlist control, suitably balanced |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| **LIFT: Reid 1999,[125] DeGarmo 2009,[98] Eddy 2000,[100] Eddy 2003,[101] Stoolmiller 2000[131]** | | |
| Methods | Unit of randomisation | School |
| Details of participants | Country | USA |
| | Sample number | Intervention group: three schools, 214 students |
| | | Control group: three schools, 147 students |
| | School years included | Year 6 |
| | Sex | 49% female, 51% male |
| | Ethnicity | 86% white, 14% ethnic minority |
| | SES | 12% mother did not graduate from high school, 8% father did not graduate from high school; 36% mother unemployed, 10% father unemployed; 22% single-parent families; 18% receiving benefits; 20% < $15,000/year in early 1990s |
| Outcomes | End of first intervention year: | |
| | ● change in observed playground aggression, $d = -0.14$; $p < 0.05$ | |
| | End of fourth year: | |
| | ● initiation of patterned alcohol use, OR 0.67; $p < 0.01$ <br> ● initiation of smoking, OR 0.90 (not significant) <br> ● initiation of marijuana use, OR 0.81 (not significant) | |
| | End of eighth year: | |
| | ● initiation of alcohol, OR 0.93; $p < 0.05$ <br> ● initiation of smoking, OR 0.90; $p < 0.01$ <br> ● initiation of illicit drug use, OR 0.91; $p < 0.10$ | |
| Details of intervention | Description | Classroom instruction and discussion on specific social and problem-solving skills followed by skills practice, reinforced during free play using a group co-operation game with review of behaviour and presentation of daily rewards. There is also a parent evening to engage families |
| | Target population | Grades 1–5 |
| | Theory | The programme follows the assumption that if young people have relationships with pro-social peers and adults, they will be less likely to engage in risky behaviour |
| | Provider(s) | Teachers |
| | Training | Providers had 40 hours of training |
| | Setting | Classrooms, playgrounds |
| | Content | Social and problem-solving skills |
| | Length/intensity | 20 1-hour sessions spread across a 10-week period; parental night once per week for 6 weeks |
| | Control | Not stated |

*continued*

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Item | Reviewer judgement | Description |
|---|---|---|
| **Risk of bias** | | |
| Sequence generation: was the allocation sequence adequately generated? | Yes | Drew allocation from a hat |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Yes | Assessment staff blind to allocation |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol to confirm |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Yes | Authors account for grade effect |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Clustering included via multilevel model |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Observation included as outcome |
| Suitable control group: did authors recruit a suitable control group? | Not clear | Unclear how differences at baseline were accounted for |
| **Details** | **Characteristics** | |
| ***Positive Action (Hawaii site):[c] Beets 2009,[87] Snyder 2010,[129] Snyder 2013[130]*** | | |
| Methods | Unit of randomisation | School |
| Details of participants | Country | USA |
| | Sample number | Intervention group: 10 schools, 976 students |
| | | Control group: 10 schools, 738 students |
| | School years included | Years 2 or 3 |
| | Sex | 50% female, 50% male |
| | Ethnicity | 26.1% Hawaiian, 22.6% mixed, 8.6% white, 1.6% African American, 1.7% American Indian, 4.7% other Pacific Islander, 4.6% Japanese, 20.6% other Asian, 7.8% other, 1.6% unknown |
| | SES | Control schools had on average 55% free/reduced lunch students, whereas intervention schools had on average 56% free/reduced lunch students |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics |
|---|---|
| Outcomes | End of fourth or fifth intervention year:<br><br>● count of substances (SR), IRR 0.41, 90% CI (0.25 to 0.66)<br>● count of substances (TR), IRR 0.66, 90% CI (0.30 to 1.45)<br>● alcohol (SR), OR 0.48, 90% CI (0.34 to 0.68)<br>● alcohol (TR), OR 0.81, 90% CI (0.41 to 1.58)<br>● drunk (SR), OR 0.30, 90% CI (0.15 to 0.57)<br>● smoked (SR), OR 0.52, 90% CI (0.31 to 0.88)<br>● smoked (TR), OR 0.54, 90% CI (0.28 to 1.02)<br>● high (SR), OR 0.20, 90% CI (0.09 to 0.44)<br>● illegal use (SR), OR 0.28, 90% CI (0.14 to 0.54)<br>● illegal use (TR), OR 0.27, 90% CI (0.1 to 0.72)<br>● count of violence behaviours (SR), IRR 0.42, 90% CI (0.24 to 0.73)<br>● count of violence behaviours (TR), IRR 0.54, 90% CI (0.30 to 0.77)<br>● cut or stabbed (SR), OR 0.29, 90% CI (0.16 to 0.52)<br>● shot someone to hurt them (SR), OR 0.24, 90% CI (0.14 to 0.40)<br>● hurts others (TR), OR 0.61, 90% CI (0.38 to 0.97)<br>● lots of fights (TR), OR 0.63, 90% CI (0.47 to 0.84)<br><br>End of fourth intervention year:<br><br>● school-level math national test (% at/above average), SMD 0.5; $p = 0.495$<br>● school-level math state test (% at/above average), SMD 0.69; $p = 0.04$<br>● school-level reading national test (% at/above average), SMD 0.58; $p = 0.108$<br>● school-level reading state test (% at/above average), SMD 0.72; $p = 0.029$<br><br>End of fifth intervention year:<br><br>● school-level math national test (% at/above average), SMD 0.52; $p = 0.291$<br>● school-level math state test (% at/above average), SMD 1.1; $p = 0.006$<br>● school-level reading national test (% at/above average), SMD 0.54; $p = 0.028$<br>● school-level reading state test (% at/above average), SMD 0.65; $p = 0.043$ |
| Details of intervention | See Positive Action programme details above |

| Item | Reviewer judgement | Description |
|---|---|---|
| ***Risk of bias*** | | |
| Sequence generation: was the allocation sequence adequately generated? | Not clear | Method of randomisation not indicated |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol to confirm |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Item | Reviewer judgement | Description |
|---|---|---|
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Not clear | Randomised design presumably accounts for confounding, but not stated |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Clustering was addressed via multilevel models |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Sensitivity and paired analyses increase robustness of findings |
| Suitable control group: did authors recruit a suitable control group? | Yes | No differences at baseline |

| Details | Characteristics | |
|---|---|---|
| *Positive Action (Chicago site):[d] Li 2011,[116] Bavarian 2013,[83] Lewis 2012,[113] Lewis 2012,[114] Lewis 2013[115]* | | |
| Methods | Unit of randomisation | School |
| Details of participants | Country | USA |
| | Sample number | Intervention group: seven schools, ≈240 students |
| | | Control group: seven schools, ≈260 students |
| | School years included | Year 4 |
| | Sex | ≈48% female, ≈52% male |
| | Ethnicity | 55% African American, 32% Hispanic, 9% white non-Hispanic, 4% Asian, 5% other or mixed |
| | SES | 83% receiving free lunch |
| Outcomes | End of third intervention year: | |

End of third intervention year:

- bullying behaviour, IRR 0.59, 95% CI (0.37 to 0.92)
- serious violence-related behaviour, IRR 0.63, 95% CI (0.45 to 0.88)
- substance use, IRR 0.69, 95% CI (0.5 to 0.97)

End of sixth intervention year:

- count of substances used, SMD –0.29; $p < 0.01$
- frequency of substances used, SMD –0.27; $p < 0.01$
- alcohol ever, SMD –0.35; $p < 0.05$
- alcohol more than once, SMD –0.35; $p < 0.05$
- drunk ever, SMD –0.29; $p < 0.01$
- drunk more than once, SMD –0.22; $p < 0.05$
- cigarette ever, SMD –0.21; $p < 0.05$
- cigarette more than once, SMD 0.03, not significant
- marijuana ever, SMD –0.23; $p < 0.05$
- marijuana more than once, SMD –0.17; $p < 0.05$
- bullying behaviours, student report, SMD –0.39
- bullying behaviours, parent report, SMD –0.31
- violence-related behaviours, IRR 0.38, 95% CI (0.18 to 0.81)
- math (school-level), $d = 0.38$
- reading (school-level), $d = 0.22$
- reading for students in wave 8, $d = 0.83$
- math for students in wave 8, reported as non-significant
- self-reported grades, $d = 0.02$
- teacher rating of academic ability, $d = 0.14$

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| Details of intervention | Description | Positive Action is a social–emotional and character development intervention aimed at encouraging positive behaviours through positive thoughts and actions |
| | Target population | Kindergarten to grade 12 (aged 6–18 years) |
| | Theory | Purkey's[179] theory of self-concept (suggests that positive self-worth leads to positive action, which reinforces self-worth) and Flay *et al.*'s[180] theory of triadic influence (suggests that positive school climate and good relationships with teachers and parents will develop positive self-concept in students) |
| | Provider(s) | Teachers |
| | Training | 3–4 hours of teacher training in the first year and 1–2 hours in subsequent years with booster sessions and mini-conferences (with other participating schools) each year |
| | Setting | Classrooms |
| | Content | Lessons cover six units: self-concept, positive actions for mind and body, positive social–emotional actions, managing oneself, being honest with oneself, and continually improving oneself |
| | Length/intensity | 140 lessons per grade per year |
| | Control | Usual curriculum |

| Item | Reviewer judgement | Description |
|---|---|---|
| ***Risk of bias*** | | |
| Sequence generation: was the allocation sequence adequately generated? | Yes | Random number generator |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol to confirm |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Yes | Differences tested by sex |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Analysis undertaken via multilevel models |

<div align="right">continued</div>

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Item | Reviewer judgement | Description |
|---|---|---|
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Sensitivity analyses were used to estimate conservative effects; missing data handled via full information maximum likelihood; non-normal variables analysed via bootstrap estimates |
| Suitable control group: did authors recruit a suitable control group? | Yes | No differences at baseline |

| Details | Characteristics | |
|---|---|---|
| ***PATHS: Crean 2013[97]*** | | |
| Methods | Unit of randomisation | School |
| Details of participants | Country | USA |
| | Sample number | Intervention group: seven schools, 422 students |
| | | Control group: seven schools, 357 students |
| | School years included | Year 4 |
| | Sex | 57% female, 43% male |
| | Ethnicity | 51% white, 38% African American, 10% other, 17% Hispanic |
| | SES | 33% from single parent homes; 39% families with income < $20,000/year, 43% below the federal poverty line; 11% no parent with high school diploma |
| Outcomes | | End of first intervention year: |
| | | • aggression (teacher report), SMD 0.036, 95% CI (–0.105 to 0.178) |
| | | • aggression (student report), SMD –0.048, 95% CI (–0.189 to 0.092) |
| | | • victimisation (student report), SMD 0.044, 95% CI (–0.098 to 0.185) |
| | | Start of second intervention year: |
| | | • aggression (teacher report), SMD 0.035, 95% CI (–0.107 to 0.176) |
| | | • aggression (student report), SMD –0.064, 95% CI (–0.205 to 0.076) |
| | | • victimisation (student report), SMD 0.074, 95% CI (–0.067 to 0.216) |
| | | End of second intervention year: |
| | | • aggression (teacher report), SMD –0.005, 95% CI (–0.146 to 0.136) |
| | | • aggression (student report), SMD –0.048, 95% CI (–0.188 to 0.093) |
| | | • victimisation (student report), SMD 0.092, 95% CI (–0.05 to 0.234) |
| | | End of third intervention year: |
| | | • aggression (teacher report), SMD –0.199, 95% CI (–0.338 to –0.06) |
| | | • aggression (student report), SMD 0.082, 95% CI (–0.06 to 0.224) |
| | | • victimisation (student report), SMD 0.089, 95% CI (–0.053 to 0.231) |
| Details of intervention | Description | An intervention to reduce conflict by improving students' social–emotional and thinking skills through a curriculum, the establishment of a positive classroom environment and generalised positive social norms throughout the school environment |
| | Target population | Kindergarten to grade 5 (aged 6–11 years) |
| | Theory | Not stated |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| | Provider(s) | Teacher |
| | Training | Ongoing mentoring and coaching (weekly in first year of implementation, bi-weekly thereafter) following initial programme training |
| | Setting | Classrooms |
| | Content | Lessons are designed to improve students' social–emotional and thinking skills and to facilitate a positive classroom environment |
| | Length/intensity | 2–3 lessons per week |
| | Control | Not stated |

| Item | Reviewer judgement | Description |
|---|---|---|
| ***Risk of bias*** | | |
| Sequence generation: was the allocation sequence adequately generated? | Not clear | Method of randomisation not indicated |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol to confirm |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Not clear | Randomised design presumably accounts for confounding, but not stated |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Multilevel model used to nest students in schools |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Additional sensitivity analyses to examine assumptions |
| Suitable control group: did authors recruit a suitable control group? | Not clear | Unclear how significant differences in baseline are accounted for |

| Details | Characteristics | |
|---|---|---|
| ***RHC:[e] Catalano 2003,[95] Brown 2005,[92]*** | | |
| Methods | Unit of randomisation | School |
| Details of participants | Country | USA |
| | Sample number | Intervention group: five schools, 497 students |
| | | Control group: five schools, 441 students |
| | School years included | Years 2 or 3 |
| | Sex | 46.7% female, 53.3% male |

*continued*

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | | |
|---|---|---|---|
| | Ethnicity | | 81.9% white, 7.4% Pacific Islander, 3.0% black, 3.9% Latinx, 2.2% Native American, 0.4% other |
| | SES | | 28% from low-income families in Brown 2005[92] |
| Outcomes | During sixth intervention year (year 7): | | |
| | ● alcohol use, intervention group intervention group 29% vs. control group 30% | | |
| | During sixth or seventh intervention year (year 8): | | |
| | ● alcohol use, intervention group 33% vs. control group 29% <br> ● cigarette use, intervention group 9% vs. control group 8% <br> ● marijuana use, intervention group 8% vs. control group 9% | | |
| | During seventh or eighth year (year 9): | | |
| | ● alcohol use, intervention group 37% vs. control group 40% <br> ● cigarette use, intervention group 14% vs. control group 13% <br> ● marijuana use, intervention group 16% vs. control group 18% | | |
| | During eighth or ninth year (year 10): | | |
| | ● alcohol use, intervention group 46% vs. control group 48% <br> ● cigarette use, intervention group 16% vs. control group 17% <br> ● marijuana use, intervention group 25% vs. control group 28% | | |
| | During ninth year (year 11): | | |
| | ● alcohol use, intervention group 52% vs. control group 50% <br> ● alcohol frequency of use, −0.40; $p < 0.05$ <br> ● cigarette use, intervention group 16% vs. control group 20% <br> ● marijuana use, intervention group 30% vs. control group 31% <br> ● marijuana frequency of use, −0.57; $p < 0.05$ | | |
| Details of intervention | Description | | The programme aims to increase academic success and reduce antisocial behaviour and chronic mental health problems in adolescents. It influences risk and protective factors in the family, school, peer and individual domains |
| | Target population | | Grades 1 and 2 (aged 6 and 7 years) |
| | Theory | | The premise of this programme is that positive social bonds with others who have normalised positive behaviours will inhibit negative behaviours in children, while the opposite is also true |
| | Provider(s) | | Teachers; school–home co-ordinators (for the family component) |
| | Training | | Teachers are trained in classroom management strategies followed by continuous coaching and booster sessions from programme staff |
| | Setting | | Classrooms, schools, homes |
| | Content | | Reading strategies, interpersonal skills and problem-solving skills; parent component includes family management training |
| | Length/intensity | | Not stated (presumed to be ongoing) |
| | Control | | Usual curriculum |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Item | Reviewer judgement | Description |
|---|---|---|
| **Risk of bias** | | |
| Sequence generation: was the allocation sequence adequately generated? | Not clear | Method of randomisation not stated |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol to confirm |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Yes | Baseline characteristics accounted for in regression models |
| Clustering: was clustering of participants accounted for in the analysis? | Not stated | |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Data collection instruments validated; authors examine robustness of findings to missing data |
| Suitable control group: did authors recruit a suitable control group? | Yes | No significant differences at baseline |
| **Details** | **Characteristics** | |
| **Second Step: Espelage 2013,[139] 2015,[140] 2015[141]** | | |
| Methods | Unit of randomisation | School |
| Details of participants | Country | USA |
| | Sample number | Intervention group: 18 schools, 1940 students |
| | | Control group: 18 schools, 1676 students |
| | School years included | Year 7 |
| | Sex | 48.1% female, 51.9% male |
| | Ethnicity | 26.4% African American, 24.7% white, 34.2% Hispanic, 14.7% bi-racial and all others |
| | SES | 74.1% free or reduced lunch |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| Outcomes | End of first intervention year: | |
| | • physical aggression perpetration, OR 0.70; $p < 0.05$<br>• sexual harassment and violence perpetration, OR 1.04; $p > 0.05$<br>• peer victimisation, OR 1.01; $p > 0.05$<br>• sexual harassment and violence victimisation, $p > 0.05$ | |
| | End of second intervention year: | |
| | • physical aggression perpetration, OR 0.80, 95% CI (0.59 to 1.08)<br>• sexual harassment and violence perpetration (Illinois), OR 0.72, 95% CI (0.54 to 0.95)<br>• sexual harassment and violence perpetration (Kansas), OR 0.99, 95% CI (0.71 to 1.48)<br>• peer victimisation, OR 0.94, 95% CI (0.75 to 1.18)<br>• sexual harassment and violence victimisation, OR 0.91, 95% CI (0.72 to 1.15) | |
| | End of third intervention year: | |
| | • $B = 0.005$, SE 0.012 | |
| Details of intervention | Description | Teachers are supported by professional development training to deliver intervention content, which includes bullying, problem-solving, emotional regulation and empathy, alongside videos. Teachers also receive plans to support integration of Second Step content into academic lessons. Modelling, role play and coaching are included in the intervention. Students receive homework to reinforce skills and use group and collaborative work to practise skills |
| | Target population | Grade 6 |
| | Theory | Explicit theoretical rationale is not discussed |
| | Provider(s) | Teachers |
| | Training | Teachers receive materials to implement, alongside training and workshops |
| | Setting | Classrooms |
| | Content | Problem-solving, emotional management, empathy |
| | Length/intensity | 15 weeks of classroom lessons taught weekly or every 2 weeks throughout the school year for 3 years |
| | Control | Minimal bullying intervention offered to control schools |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Item | Reviewer judgement | Description |
|---|---|---|
| **Risk of bias** | | |
| Sequence generation: was the allocation sequence adequately generated? | Yes | Matched-pair randomisation |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | No | Third-year results are not complete |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Yes | Confounders included and explored |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Analysis undertaken via multilevel models |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Authors considered moderators |
| Suitable control group: did authors recruit a suitable control group? | Yes | Baseline equivalence established |
| **Details** | **Characteristics** | |
| **Steps to Respect I: Frey 2005,[104] Frey 2009[103]** | | |
| Methods | Unit of randomisation: | School |
| Details of participants | Country | USA |
| | Sample number | Three schools (intervention group), three schools (control group); 1126 students total |
| | School years included | Years 4–7 |
| | Sex | 49.4% female, 50.6% male |
| | Ethnicity | 70.0% white, 9% African American, 12.7% Asian, 7.0% Hispanic, 1.3% Native American |
| | SES | Not stated |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| Outcomes | End of first intervention year: | |
| | • bullying change score (observation), mean 0.97, SD 1.71 (intervention group) vs. mean 1.19, SD 2.11 (control group); $F_{1,91.3} = 5.02$; $p < 0.01$<br>• direct aggression (student report), mean 0.48, SD 0.62 (intervention group) vs. mean 0.62, SD 0.71 (control group); $F_{1,68.7} = 2.05$, not significant<br>• victimisation (observation), mean 0.80, SD 1.51 (intervention group) vs. mean 0.86, SD 1.44 (control group); $F < 1$, not significant<br>• victimisation (student report), mean 0.90, SD 0.82 (intervention group) vs. mean 1.01, SD 0.83 (control group); $F_{1,72.4} = 3.74$; $p < 0.10$ | |
| Details of intervention | Description | This is an antibullying intervention with both school-wide and classroom components. The school-wide components create new disciplinary policies for bullying and improve monitoring of and intervention in bullying. Classroom curricula create positive social norms and improve social–emotional skills for better engagement with bullying |
| | Target population | Grades 3 to 6 (aged 9–12 years) |
| | Theory | A social–ecological model of bullying that suggests behaviour is shaped by many factors at the school, peer and individual levels |
| | Provider(s) | Teachers |
| | Training | 1-day on-site teacher and staff training in the curriculum to develop effective disciplinary policies, to improve monitoring of students and to effectively intervene with students involved in bullying situations |
| | Setting | Classrooms and whole schools |
| | Content | Recognition of bullying, empathy, friendship skills, communication skills and appropriate bystander responses to bullying |
| | Length/intensity | 11 lessons throughout the year, plus whole-school approaches throughout |
| | Control | Usual curriculum |

| Item | Reviewer judgement | Description |
|---|---|---|
| *Risk of bias* | | |
| Sequence generation: was the allocation sequence adequately generated? | Not clear | Method of randomisation not stated |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Item | Reviewer judgement | Description |
|---|---|---|
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol to confirm |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Not clear | Randomised design presumably accounts for confounding, but not stated |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Analysis undertaken via multilevel models |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Observation as outcome measure increases trustworthiness of findings |
| Suitable control group: did authors recruit a suitable control group? | Not clear | No baseline equivalence reported |

| Details | Characteristics | |
|---|---|---|
| **Steps to Respect II: Brown 2011[93]** | | |
| Methods | Unit of randomisation | School |
| Details of participants | Country | USA |
| | Sample number | 17 schools (intervention group), 16 schools (control group); 2940 students total |
| | School years included | Years 4 to 6 |
| | Sex | Intervention group: 51% female, 49% male |
| | | Control group: 48% female, 52% male |
| | Ethnicity | Intervention group: 52% white, 7% African American, 6% Asian, 43% Hispanic, 35% other or mixed race |
| | | Control group: 53% white, 6% African American, 6% Asian, 41% Hispanic, 35% other or mixed race |
| | SES | School-level average of 40% on free or reduced-price lunch |
| Outcomes | End of first intervention year: | |

- bullying change score (observation), mean 0.97, SD 1.71 (intervention group) vs. mean 1.19, SD 2.11 (control group); $F_{1,91.3} = 5.02$; $p < 0.01$
- direct aggression (student report), mean 0.48, SD 0.62 (intervention group) vs. mean 0.62, SD 0.71 (control group); $F_{1,68.7} = 2.05$, not significant
- victimisation (observation), mean 0.80, SD 1.51 (intervention group) vs. mean 0.86, SD 1.44 (control group); $F < 1$, not significant
- victimisation (student report), mean 0.90, SD 0.82 (intervention group) vs. mean 1.01, SD 0.83 (control group); $F_{1,72.4} = 3.74$; $p < 0.10$

| | | |
|---|---|---|
| Details of intervention | See programme details above | |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Item | Reviewer judgement | Description |
|---|---|---|
| **_Risk of bias_** | | |
| Sequence generation: was the allocation sequence adequately generated? | Not clear | Method of randomisation not stated |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol to confirm |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Not clear | Randomised design presumably accounts for confounding, but not stated |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Analysis includes multilevel models |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Authors correct for statistical error at multiple levels, examine robustness of findings to missing data |
| Suitable control group: did authors recruit a suitable control group? | Not clear | Authors report statistical differences in 1/24 outcome measures between groups |
| **Details** | **Characteristics** | |
| **_The Gatehouse Project: Bond 2004,[91] Bond 2004,[90] Bond 2010,[88] Patton 2006[60]_** | | |
| Methods | Unit of randomisation | District |
| Details of participants | Country | Australia |
| | Sample number | Longitudinal cohort |
| | | • intervention group: two districts, 12 schools, 1335 students<br>• control group: two districts, 14 schools, 1343 students |
| | | Repeated cross-sections |
| | | • 1999<br> ○ intervention group: 1158 students<br> ○ control group: 1428 students |
| | | • 2001<br> ○ intervention group: 966 students<br> ○ control group: 1497 students |
| | School years included | Year 9 at entry |
| | Longitudinal cohort | |
| | Sex | 53.2% female, 46.8% male |
| | Ethnicity | 87.5% Australian-born |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| | SES | 79.2% from two-parent family; 24.2% speak language other than English at home |
| | 1999 cross-section | |
| | Sex | 55% female, 45% male |
| | Ethnicity | 88% Australian-born |
| | SES | 79.2% from two-parent family; 22% speak language other than English at home |
| | 2001 cross-section | |
| | Sex | 54% female, 46% male |
| | Ethnicity | 90% Australian-born |
| | SES | 78% from two-parent family; 19% speak language other than English at home |
| Outcomes | Longitudinal cohort: | |

Longitudinal cohort:

- after 1 year of intervention
  - bullying victimisation, OR 1.03, 95% CI (0.86 to 1.26)
  - any drinking, OR 0.93, 95% CI (0.71 to 1.21)
  - any smoking, OR 0.89, 95% CI (0.72 to 1.12)
  - regular drinker, OR 1.09, 95% CI (0.77 to 1.57)
  - binge drinking, OR 0.95, 95% CI (0.69 to 1.32)
  - cannabis use in last 6 months, OR 0.98, 95% CI (0.69 to 1.4)
  - regular smoker, OR 0.66, 95% CI (0.46 to 0.95)

- after 2 years of intervention
  - bullying victimisation, OR 1.03, 95% CI (0.78 to 1.34)
  - any drinking, OR 1, 95% CI (0.78 to 1.28)
  - any smoking, OR 0.92, 95% CI (0.63 to 1.33)
  - binge drinking, OR 0.99, 95% CI (0.7 to 1.38)
  - cannabis use in last 6 months, OR 1.06, 95% CI (0.75 to 1.49)
  - regular drinker, OR 1.05, 95% CI (0.70 to 1.57)
  - regular smoker, OR 0.72, 95% CI (0.47 to 1.09)

- after 3 years of intervention
  - bullying victimisation, OR 0.88, 95% CI (0.68 to 1.13)
  - any drinking, OR 0.96, 95% CI (0.69 to 1.33)
  - any smoking, OR 0.91, 95% CI (0.67 to 1.24)
  - binge drinking, OR 1.02, 95% CI (0.71 to 1.46)
  - cannabis use in last 6 months, OR 0.81, 95% CI (0.57 to 1.16)
  - regular drinker, OR 1.13, 95% CI (0.77 to 1.66)
  - regular smoker, OR 0.79, 95% CI (0.58 to 1.07)
  - any cannabis use incidence, OR 0.81, 95% CI (0.56 to 1.15)
  - any cannabis use prevalence, OR 0.80, 95% CI (0.57 to 1.13)
  - weekly cannabis use incidence, OR 0.73, 95% CI (0.43 to 1.25)
  - weekly cannabis use prevalence, OR 0.74, 95% CI (0.45 to 1.2)

Repeated cross-sections [after 1 intervention year, substance use (three categories)]:

- 1997, OR 0.92, 95% CI (0.7 to 1.2)
- 1999, OR 0.84, 95% CI (0.61 to 1.18)
- 2001, OR 0.85, 95% CI (0.65 to 1.12)

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| Details of intervention | Description | Through teaching a curriculum and establishing a school-wide adolescent health team, Gatehouse aims to build a sense of security and trust in students, enhance skills and opportunities for good communication and build a sense of positive regard through participation in school life |
| | Target population | Grade 8 (aged 14 years) |
| | Theory | The Health-Promoting Schools Framework, which suggests the school ethos and environment influence student behaviours |
| | Provider(s) | Not clear, but suggested to be principal or assistant principal, curriculum leader, student welfare co-ordinator, teachers and community representatives |
| | Training | The Centre for Adolescent Health liaison team provided training, resources and mentorship to schools |
| | Setting | Whole school |
| | Content | Classroom curriculum unclear |
| | Length/intensity | Not stated |
| | Control | Not stated |

| Item | Reviewer judgement | Description |
|---|---|---|
| *Risk of bias* | | |
| Sequence generation: was the allocation sequence adequately generated? | Not clear | Method of randomisation not indicated |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol available |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Yes | Multivariate analysis accounts for confounders |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | 'Information sandwich' robust standard errors account for clustering within schools |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Yes | Tests for interaction between group and stratification variables performed |
| Suitable control group: did authors recruit a suitable control group? | Yes | Authors account for differences at baseline |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| *Youth Matters: Jenson 2007,[109] 2010,[110] 2013[111]* | | |
| Methods | Unit of randomisation | School |
| Details of participants | Country | USA |
| | Sample number | Intervention group: 14 schools, 702 students |
| | | Control group: 14 schools, 462 students |
| | School years included | Year 5 |
| | Sex | 50.6% female, 49.4% male |
| | Ethnicity | 59.1% Latinx, 14.7% African American, 16.8% American Indian, Asian American or mixed, 9.3% white |
| | SES | Not stated |
| Outcomes | Main analysis: | |
| | • end of second intervention year | |
| | ○ bullying perpetration, OR 0.85; $p = 0.565$ | |
| | ○ bullying victimisation (dichotomous), OR 0.61; $p = 0.098$ | |
| | ○ bullying victimisation (continuous log-transformed score), difference −0.171; $p = 0.049$ | |
| | • end of third intervention year | |
| | ○ bullying victimisation (continuous log-transformed score), difference −0.123; $p = 0.08$ | |
| | Exploratory latent class modelling: | |
| | • end of first intervention year | |
| | ○ bully and bully–victim, 21% (intervention group) vs. 22% (control group) | |
| | ○ victim and bully–victim, 47% (intervention group) vs. 46% (control group) | |
| | • end of second intervention year | |
| | ○ bully and bully–victim, 19% (intervention group) vs. 23% (control group) | |
| | ○ victim and bully–victim, 45% (intervention group) vs. 47% (control group) | |
| | • end of third intervention year | |
| | ○ bully and bully–victim, 16% (intervention group) vs. 16% (control group) | |
| | ○ victim and bully–victim, 36% (intervention group) vs. 45% (control group) | |
| Details of intervention | Description | Youth Matters promotes the development of healthy relationships and social competency and the development of social resistance. Classroom discussions around social issues promote positive social norms |
| | Target population | Grades 4 and 5 (aged 9 and 10 years) |
| | Theory | The Social Development Model, which suggests that the following inhibit antisocial behaviours in children: 1. bonding; 2. shared values and norms; 3. external barriers to antisocial behaviour; 4. social, cognitive and emotional skills |

**TABLE 25** Summary of outcome evaluation characteristics and risk of bias by intervention (*continued*)

| Details | Characteristics | |
|---|---|---|
| | Provider(s) | Educational specialists (not teachers) |
| | Training | Not stated |
| | Setting | Classrooms |
| | Content | Skill modules teach social competency and social resistance, issues modules teach around bullying from the perspective of bullies and victims |
| | Length/intensity | There are four curriculum models, each with 10 lessons, delivered across 2 years |
| | Control | Not stated |

| Item | Reviewer judgement | Description |
|---|---|---|
| **Risk of bias** | | |
| Sequence generation: was the allocation sequence adequately generated? | Not clear | Method of randomisation not stated |
| Allocation concealment: was the allocation adequately concealed? | Not stated | |
| Blinding: was knowledge of the allocation intervention adequately prevented during the study? | Not stated | |
| Complete outcome data: were complete data for each outcome reported, and, if not were adequate reasons for incomplete outcome data provided? | Yes | Low attrition |
| No selective outcome reporting: were the findings of the study free of selective outcome reporting? | Not clear | No protocol to confirm |
| Key confounders: were differences in non-random studies in key confounders (e.g. SES, sex, age) adequately controlled for? | Yes | Analysis begins by looking at effect of baseline characteristics on outcomes, considers effect modification as well |
| Clustering: was clustering of participants accounted for in the analysis? | Yes | Multilevel models used to account for clustering |
| Reduced other source of bias: did authors take steps to reduce any other sources of bias that might affect the results of the study? | Not stated | |
| Suitable control group: did authors recruit a suitable control group? | Yes | Authors account for differences at baseline in analysis |

SMD, standardised mean difference; SR, student report; TS, teacher report.
a   Included three arms, of which one was not relevant to this study (LST without integration). Participant details are reported for the entire study.
b   Included a cohort recruited in year 2, but this cohort received an intervention that did not integrate academic and health education. Information relating to this cohort is not presented here.
c   Because this evaluation enrolled new students as they entered and left the school, participant characteristics presented are at the final year of the evaluation.
d   Because this evaluation enrolled new students as they entered and left the school, data presented are at baseline from 2004.
e   Students were analysed by grade level; thus, students in some school years span two possible intervention durations.