Healthy Minds

Evaluation Report

July 2022

Lucy Stokes, Johnny Runge, David Wilkinson, Heather Rolfe, Cinzia Rienzo, Jonathan Buzzeo and Matthew Bursnall

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education. Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

Jonathan Kay
Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

This evaluation of the impact of the Healthy Minds programme on academic attainment was carried out by a team at the National Institute of Economic and Social Research (NIESR): Lucy Stokes, Heather Rolfe, David Wilkinson, Johnny Runge, Cinzia Rienzo, Matthew Bursnall, and Jonathan Buzzeo.

Contact:
Lucy Stokes, email: **l.stokes@niesr.ac.uk**

Acknowledgements

The authors would like to thank all the schools, staff, and pupils who participated in this research and without whom this evaluation would not have been possible. We would also like to thank the advisory group, chaired by Lord (Professor) Richard Layard, for their guidance throughout the project. Particular thanks to Lucy Bailey at Bounce Forward for working with us and for being extremely helpful and responsive to any evaluation requests and to Dr Grace Lordan and Professor Alistair McGuire for their advice and insights. Finally, thank you to the EEF for funding this research, and especially to the evaluation and programmes team managers at the EEF for their support and patience throughout the delivery of this research.

Executive Summary

The project

Healthy Minds is a personal, social, and health education curriculum for secondary school pupils (Years 7 to 10), which aims to improve pupils' wellbeing and health-related outcomes. The programme, developed and delivered by Bounce Forward, comprises a set of 14 modules developed from components of different health promotion interventions identified to be effective through previous research. Using principles from cognitive behavioural therapy and theory from positive psychology, it helps students improve their resilience and understand the link between thoughts, feelings and behaviour so they understand themselves better and have more empathy for others.

The lessons, which are taught directly by trained teachers or learning support assistants, take about one hour per week and either replace schools' existing PSHE lessons or can be built into the school week at other times. The programme is intended to be delivered to pupils over four years, from Years 7 to 10 (covering ages 11 to 15).

This was an efficacy randomised controlled trial with a waitlist design. It ran from September 2013 until July 2018 and involved 34 schools and more than 9,000 pupils. Schools were recruited in two phases, with 13 schools recruited in phase one and 21 recruited in phase two. It explored whether the programme has an impact on health related outcomes (results published in 2019) and on academic attainment as measured by Attainment 8 scores, which capture GCSE attainment at the end of Key Stage 4 (when pupils are aged 16). This evaluation also explored impacts on absence and exclusions and on attainment in reading and maths at end Year 7 and end Year 8 as secondary outcomes. The evaluation also included an implementation and process evaluation (IPE) with qualitative school case studies, staff interviews, and pupil focus groups. This evaluation report should be read alongside the report on health related outcomes that was published in 2019.

Table 1: Key conclusions

Key conclusions
There is no evidence that Healthy Minds had an impact on pupils' outcomes as measured by Attainment 8 scores, on average. This result has a low to moderate security rating.
Pupils receiving the Healthy Minds programme did not show any additional progress on reading and maths scores at the end of Year 7 or Year 8. These results should be interpreted with caution considering the challenges in administering the tests and related high levels of attrition. There was some evidence of lower levels of absence, especially in Year 7 pupils, and some indication of fewer exclusions among pupils eligible for free school meals (FSM pupils) in schools that received the Healthy Minds programme compared with similar pupils in schools that did not.
There was no evidence to suggest that the programme had a differential impact on the Attainment 8 scores of FSM pupils or according to their prior levels of attainment.
Interviews with participating teachers indicated that the Healthy Minds programme was generally well-received by practitioners. Teachers reported relatively few practical difficulties in delivery with the main issues related to timetabling (particularly in Year 10) and teacher turnover (with new teachers needing to be trained).
Many of the teachers interviewed felt that in order to maximise the impact of the programme it needed to be reinforced in other areas of teaching and learning and in whole-school improvement activities and messaging.

EEF security rating

These findings have a low to moderate security rating. The trial was a well-designed cluster randomised controlled trial, however, there were some important differences in prior attainment between the pupils in Healthy Minds schools and those in the comparison schools, which were controlled for in the analysis. Twelve percent of the pupils who started the trial were not included in the final analysis, mainly because two schools that withdrew did not provide pupil data. Threat of contamination of internal control groups was identified by the evaluation team as a risk to the validity of findings as the evaluation is unable to account for this in analyses.

Additional findings

The primary aim of the Healthy Minds programme is to improve health-related outcomes. The previous evaluation of health-related outcomes (Lordan and McGuire, 2019) demonstrated a positive impact on the primary outcome, self-assessed general health, as measured by the Child Health Questionnaire CHQ-CF87 (with an effect size of 0.25 standard deviations after four years of the programme). Positive effects were also observed for some secondary outcomes, especially for physical health and externalising behaviour. The current evaluation of academic outcomes shows no impact on academic attainment, as measured by Attainment 8. This result is robust to a number of sensitivity analyses and is also unchanged when considering results among those schools that were considered to have delivered the intervention in full. This evaluation found no positive or negative impact on academic attainment, however, the programme may represent a promising approach for improving some health-related outcomes without adversely affecting academic achievement, despite potentially diverting time away from traditional lessons. This programme resulted in many participating schools having a significant increase in timetabled PSHE provision compared to previous practice.

The evaluation also provides some evidence of lower levels of absence, especially in Year 7 pupils, and some indication of fewer exclusions among FSM pupils in schools that received the Healthy Minds programme compared with similar pupils in schools that did not. Across all outcomes, there was little indication of differences in impact according to whether pupils had higher or lower levels of prior attainment, as measured by attainment at the end of Key Stage 2.

These findings are consistent with the perceptions of teachers who, in general, felt that the programme would have greater impacts on wellbeing than on academic attainment and, in some cases, that it may take longer for any such impacts to become measurable. Teachers did, however, believe there were improvements in pupils' thinking skills, such as reasoning, and that the programme had potential to improve pupils' strategies for dealing with stress, such as in relation to exams. The findings are also consistent with reports by teachers in some schools of improved behaviour among their Healthy Minds cohort and pupils themselves described how they sometimes did things differently because of what they had learned through the programme.

Cost

Costs were estimated as part of the separate evaluation of the impact of the Healthy Minds programme on health outcomes (Lordan and McGuire, 2019). This estimated the costs of providing training and resources to deliver Healthy Minds at £23.50 per pupil per year.

Impact

Table 2: Summary of impact on primary outcome

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	P-value	EEF cost rating
Attainment 8	0.00 (-0.19, 0.20)	0	2	9,410	0.97	£ £ £ £ £

Introduction

Background

The Healthy Minds programme aims to improve pupils' wellbeing and health-related outcomes. The intervention forms a personal, social, and health education curriculum for Year 7 to Year 10 pupils. This comprises a set of 14 modules, bringing together individual elements that have been identified to be effective through previous research (Coleman et al., 2011). The programme uses the principles of cognitive behavioural therapy and theory from positive psychology to help students improve their resilience, understand the link between thoughts feelings and behaviour, so they understand themselves better and have more empathy for others. The resilience skills form the basis and thread through the curriculum into the full range of personal, social, and health topics. Lessons are taught by school staff (teachers or learning support assistants) who receive in-depth training in each element.

Concerns over the quality of PSHE provision in schools have previously been raised; in 2013, a review by Ofsted found that PSHE teaching required improvement in 40% of schools (Ofsted, 2013). It has been argued that in order to raise standards, it is necessary for PSHE to become a statutory requirement. Some elements have since become statutory: since autumn 2020, health education has been compulsory in all schools, along with relationships and sex education in secondary schools and relationships education in primary schools (PSHE Association, 2021).[1]

The approach is motivated by the growing recognition of the importance of non-cognitive skills in determining a range of outcomes in life (for example, Heckman et al., 2006; Heckman et al., 2013). While evidence existed on effectiveness of individual elements that are included within the Healthy Minds programme, it was not known how an integrated course combining these elements—and that was suitable for use in UKschools—may impact on pupils' outcomes (for both cognitive and non-cognitive outcomes).

Lordan and McGuire (2019) analysed the impact of the programme on health-related outcomes through a two-arm cluster randomised trial, finding positive effects. Pupils in schools receiving the programme had higher average self-assessed general health than pupils in schools not receiving the programme (an effect size, ES, of 0.25 standard deviations after four years, as measured by the Child Health Questionnaire-CF87). Positive effects were also found for many secondary outcomes capturing physical health and externalising behaviour as measured by the subscales of CHQ-CF87, including for physical difficulties (ES = 0.30), general behaviour (ES = 0.15), general health (ES = 0.15), and family activities (ES = 0.12). Effects on measures of internalised emotions (emotional difficulties, self-esteem, and mental health) were close to zero after four years of the programme (ES = 0.05, -0.04, and 0.07 respectively), but in some cases were negative when measured after two years of the programme (ES = -0.13, -0.21, and -0.19 respectively, and only effects for self-esteem and mental health were statistically significant at the 10% level after applying corrections for multiple comparison testing). Favourable effects were found for three of the six items on the Child Anxiety Related Disorders Questionnaire pain disorder: ES = -0.09; separation anxiety: ES = -0.06; significant school avoidance: ES = -0.17 and for life satisfaction (ES = 0.18) but no impact was found on the Short Mood and Feelings Questionnaire. Lordan and McGuire propose a number of potential explanations for why positive impacts focus on externalising behaviour and health, with any negative interim impact on internalising behaviour, including: the possibility that the programme may improve externalising behaviour through an initial reduction in internalising behaviour (with, for example, pupils becoming more self-critical in the interim); that the programme may lead them to answer the questions differently to pupils not receiving the programme (for example, if they are more reflective); that there may not be sufficient content to positively impact internalising behaviour; or teachers may be better at teaching components relating to externalising behaviour and health.

The main focus of the current evaluation report is whether the programme has an impact on academic attainment. It builds on the Lordan and McGuire (2019) report, and the two evaluations are based on the same randomised trial, but assess different outcomes.[2] The same schools form part of both evaluations although the analysis samples differ, as discussed later in this report. If there is no impact on academic attainment this provides support that the programme improves some health outcomes without adversely affecting academic achievement, despite potentially diverting time

[1] https://pshe-association.org.uk/long-road-statutory-pshe-education-almost
[2] Data was available on health-related outcomes at an earlier point than it was possible to access data on academic outcomes enabling earlier publication of the impacts on health-related outcomes.

away from traditional lessons. In some participating schools, the introduction of regular, timetabled PSHE did represent a significant change in previous PSHE provision (that is, increasing time allocated to PSHE). Furthermore, the skills which the programme aims to foster could potentially have a positive effect on academic results. For example, increased resilience, and developing skills to better manage stress and emotions, may well help students to apply themselves to, and persevere in, their learning and thus perform better academically. A meta-analysis by Durlak et al. (2011) of school-based social and emotional learning programmes found an average effect on academic attainment of 0.27 standard deviations. A more recent review by Corcoran et al. (2018) found similar results, with an average effect size for reading of 0.25 standard deviations, for maths, 0.26, and for science, 0.19, although the authors also found that randomised, larger-scale studies tended to show smaller effects than studies using quasi-experimental designs.

The evaluation comprises a cluster randomised trial with school-level randomisation. The nature of the intervention means that randomising below school level would not be feasible. The primary outcome for the impact evaluation is Attainment 8 (capturing GCSE attainment as measured at the end of Key Stage 4, and thus after receiving the full four-year programme). Secondary outcomes are reading and maths attainment, measured at end Year 7 and end Year 8 (thus after one and two years of receiving the programme respectively), as well as exclusions and absenteeism (again, for Year 7 and Year 8). The impact evaluation is complemented by an implementation and process evaluation to provide greater insight into the quantitative research findings, increasing understanding of implementation, fidelity, and outcomes as perceived by participating teachers and pupils.

Intervention

Using the template for intervention description and replication (TIDieR) checklist (Hoffmann et al., 2014) the key components of the intervention can be described as below. No logic model was developed as this was not a requirement at the time of setting up the evaluation.

Name: Healthy Minds.

Why: the programme aims to improve pupils' wellbeing and health-related outcomes.

Who (recipients): pupils in Years 7 to 10 of secondary school (covering ages 11 to 15 years).

What (materials and procedures): the programme is based on a model curriculum proposed by Coleman et al. (2011) for the delivery of PSHE in British secondary schools. It comprises a set of 14 modules, selected through considering programme content and aims, evidence of effectiveness, and feasibility of implementation. This was informed by a review of existing literature as well as interviews with experts including developers, researchers, and practitioners. The modules included in the curriculum are Penn Resilience Programme, Breathe (Mindfulness), Media Navigator, From School to Life, EU-DAP Unplugged (Part 1 and 2), Media Influences, Resilience Revisited, Sex Ed Sorted (Part 1 and 2), Relationship Smarts Plus, School Health Alcohol Harm Reduction Programme (SHAHRP), Resilient Decisions, Mental Illness Investigated, Parents Under Construction, and Resilient Learners. Full details of the modules, and the underlying evidence of their effectiveness, are provided in Coleman et al. (2011). While evidence of effectiveness was an important consideration in proposing the modules included in the programme, Coleman et al. (2011) describe the challenges in assessing effectiveness and note that the quality of evidence is variable. It should also be noted that such evidence typically focused on non-cognitive outcomes, given the aim of the programme.

The modules listed above were delivered in the following school years:

Year 7	Penn Resilience Programme, Media Navigator, Breathe.
Year 8	From School to Life, EU-DAP Unplugged (a substance use/misuse prevention programme), Media Influences, Sex Education (part 1), Relationships, and a resilience reflection programme ('Review and Connect').
Year 9	Relationships, Alcohol misuse, Sex Education (Part 2), and Resilience ('Resilient Decisions').
Year 10	Mental Illness, Substance Misuse, Relationships ('Parents under Construction'), and Resilience (in exams).

School staff receive in-depth training in delivering the programme along with resources including lesson plans, teacher notes and workbooks. This comprised seven days of training for Year 7 teachers, six in Year 8, two in Year 9, and four days in Year 10.

Who (provider): the intervention was developed and delivered by Bounce Forward, which trains teachers and other school staff to deliver the programme.

How (format): in schools receiving the intervention, the programme either replaced schools' existing PSHE lessons (where this time was allocated) and/or was built into the school week at other times.

Where (location): secondary schools in various locations in England.

When and how much (dosage): one hour session per week over four years (Years 7 to 10). The course comprised 113 hours in total.

Tailoring: schools were able to make some tweaks; we discuss evidence of any adaptations made as part of the implementation and process evaluation.

Implementation challenges

A number of challenges arose during the course of the project.

Recruiting schools to the evaluation proved challenging within the planned timeframes. Recruitment began in the middle of a school year and there were difficulties for schools to align participation in the programme with their usual planning cycles when arrangements for the next school year may already have been made. This meant that recruitment took place in two phases, with the first phase starting the intervention from September 2013 and a second phase starting in September 2014.

It was challenging to both recruit and retain schools for a four-year programme that required committing to a regular timetabled slot over a long period of time. To aid with retention, the original plan had been to use a waitlist control design but the challenges in recruiting meant that only phase one control schools formed a waitlist group while phase two control schools formed a 'pure' control group and did not go on to receive the intervention as part of the trial. We describe the design of the trial and the two phases in more detail within the Methods section.

Academic outcomes were originally to be assessed using tests administered specifically as part of the project assessing pupils' reading and maths skills—the Hodder Education Access Reading and Maths tests ('the Hodder tests'). These were to be administered at the end of each school year, initially at end of Year 7 and end of Year 8 (with continuation beyond that point subject to review by the EEF).[3] However, there were several challenges in administering the tests and this proved very burdensome, particularly for participating schools. Tests were initially used in digital format, up until June 2015. However, due to technical difficulties with the digital version of the tests, from Autumn 2015 paper-based versions of the tests were used instead due to a failure of the Hodder server. The failure of the Hodder server also meant that in some schools the tests due to be carried out in summer 2015 were delayed until the autumn; furthermore, in some schools tests were not completed at all, or not for all intended students.

While these assessments were conducted for some Year 7 and Year 8 pupils, there was considerable missingness and attrition (sample sizes available for analysis are presented later in this report). After the first two years, the decision was taken to stop the use of these assessments and instead to use linked data on pupil attainment at the end of GCSE. This meant there was a full school year between delivery of the intervention ending and pupils' academic attainment being assessed. The update to the trial protocol in 2016 notes that the Year 7 and Year 8 assessments would form secondary outcomes.

Some schools within the treatment group did not fully administer the programme where, for example, they were unable to continue meeting staffing requirements; we discuss this later in this report and while our main analysis is based on an intention-to-treat design, we also undertake a compliance analysis as a sensitivity check.

[3] This may be considered an early point at which to assess attainment given the full programme is designed to run for four years.

Evaluation objectives

A number of changes to the plans for the evaluation were made over time, particularly in respect of the impact evaluation. The original trial protocol does not specify the research questions for the impact evaluation in detail but notes that the main research questions concern whether improving pupils' non-cognitive skills boosts their academic attainment. The trial was originally funded for two years, with impacts on academic outcomes to be assessed through pupils' attainment on reading and maths assessments administered as part of the trial (specifically, the Hodder Education Access Reading and Maths tests, with the former the primary outcome and the latter the secondary outcome). The protocol also noted that a number of subgroup analyses would be conducted: for pupils eligible for free school meals compared with non-eligible pupils, for ethnic minority pupils compared with white pupils, and for children with low attainment scores at pre-test compared with pupils with higher attainment scores.

The protocol also documents the aims of the implementation and process evaluation, which sets out to identify the factors that affect impact and which may explain the findings of the quantitative evaluation, with the aim of assessing evidence in relation to:

- the introduction of the programme in participating schools, including commitment of senior leadership;
- contextual factors, including other activities and initiatives with a resilience component and the exploration of resource issues of relevance to the programme;
- training of teachers in the programme, their understanding of the approach and commitment, and their preparation for the emotional impact on pupils;
- the application of the programme within the PSHE slot, size of groups, and who it is delivered by;
- views on the curriculum materials; and
- collaborative partnerships relevant to the operation and implementation of the programme.

In 2016, an updated trial protocol was published, confirming two additional years of funding (that is, funding in order to continue delivering the programme in Years 9 and 10). In addition, as noted above, the administration of the Hodder tests faced considerable challenges and the 2016 protocol update documents the decision to no longer collect attainment data directly but to explore impacts on academic outcomes by assessing impact on GCSE attainment as the primary outcome instead. The Hodder reading and maths assessments conducted in Year 7 and 8 would instead form secondary outcomes. In addition, the evaluation of academic outcomes would consider impacts on absence and exclusions, with health and wellbeing outcomes explored through the separate evaluation conducted by the LSE team. The subgroup analysis for ethnic minority children was removed due to the necessary data not being available for analysis. The process evaluation was also extended to cover the final two years of the programme. In summary, the research questions to be explored were summarised in the 2016 update to the protocol, with the evaluation aiming to explore (1) whether the programme boosts academic achievement, (2) whether the programme improves pupils' wellbeing, (3) whether any impact of the programme on academic attainment was moderated by its impact on wellbeing, and (4) whether there is a differential impact on disadvantaged pupils (as defined by eligibility for free school meals).

In 2018, further updates to the protocol were made, the key change being that research question (3) would no longer be analysed (whether any impact on academic attainment is mediated by an impact on wellbeing), due to data-sharing between teams not being feasible given commitments made in the letters to parents regarding the study (and a reissuing of letters being discarded as too burdensome and costly).

The resulting research questions following all the changes noted above are documented in the Statistical Analysis Plan for the analysis of academic outcomes and are as follows.

The primary research question to be addressed in this trial is:

- Does the Healthy Minds programme have an impact on academic attainment, as measured by pupil attainment at the end of Key Stage 4?[4]

In addition, the evaluation explores the following secondary research questions:

- Does the Healthy Minds programme have a differential impact on academic attainment (at the end of Key Stage 4) for disadvantaged pupils (as defined by eligibility for free school meals)?
- Does the Healthy Minds programme have a differential impact on academic attainment (at the end of Key Stage 4) according to pupils' prior academic achievement (as defined by attainment at the end of Key Stage 2)?
- Does the Healthy Minds programme have an impact on academic attainment (in maths and English) at the end of Year 7 and the end of Year 8?
- Does the Healthy Minds programme have an impact on the likelihood of exclusion and absenteeism by the end of Year 7 and the end of Year 8?

The original protocol and subsequent amendments, along with the Statistical Analysis Plan (SAP), are all available on the EEF website project page here: **https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/developing-healthy-minds-in-teenagers**

Ethics and trial registration

Ethical review for the project was originally undertaken by the LSE Ethics Committee at the start of the study (see Lordan and McGuire, 2019). No trial registration number (ISRCTN) is available as this registration was not a requirement at the time of setting up the trial. In 2020, given the number of changes that had been made to the plans for the evaluation over time, an ethical review for the analysis of academic outcomes was undertaken by the NIESR Research Ethics Committee (with ethical approval granted 15 July 2020).

The Bounce Forward team made individual contact with each school and provided a full explanation of the evaluation during the scheduled set-up meetings. Headteachers were asked to give signed consent for their school to take part. In signing up for the project, schools were fully aware that they were giving consent for the evaluation to take place and what this would involve.

Schools were provided with letters to be sent to parents asking for parents' permission for pupils to be recruited into the study in the form of a decision to not participate. The letter also provided information about the aims of the study and the handling of data; it included information about the use of data from the National Pupil Database (NPD) and included a privacy notice describing the collection and linking of data in more detail (see Appendix C).

Data protection

This trial commenced prior to the introduction of the General Data Protection Regulations (GDPR), which came into force in 2018.

NIESR's legal basis for processing personal data for the evaluation of academic outcomes is 'legitimate interests' (GDPR Article 6.1(f)). Following the introduction of the GDPR, NIESR undertook a legitimate interests assessment which concluded that NIESR has a legitimate interest in processing the personal data in order to conduct the evaluation, that the processing identified is necessary in order to meet this purpose, and that there are minimal impacts of the processing on the individuals involved (with steps taken to ensure this). At this stage an updated project privacy notice was also added to the project page of the NIESR website (Appendix D). All processing of personal data has remained in line with the assurances made to participants prior to the introduction of the GDPR.

[4] That is, at end Year 11 when pupils are typically 16 years old.

At the start of the project, letters, including a privacy notice, were sent to parents of participating pupils explaining the project and research, including the matching of pupil data to NPD records (Appendix C). Parents could withdraw their child from participating in the research if they wished to do so.

Only the LSE and NIESR evaluation teams had access to personal data for the purposes of conducting the evaluation (with health outcomes data only accessible to the LSE team and academic outcomes data only accessible to the NIESR team). It was also explained that the data used in the evaluation by NIESR would be matched to pupils' records in the NPD.

NIESR will delete any personal data six months following completion of the project. At the end of the project, data (but not pupil names) is submitted to the EEF's data archive and at this point the EEF become data controller.

Project team

The intervention was developed and delivered by the team at Bounce Forward (formerly How to Thrive), led by Lucy Bailey (CEO, Bounce Forward). She has overseen the delivery of the project as a whole, including course development and the training of teachers, as well as all practical aspects of data collection and processing and monitoring, guidance, and support for the schools.

The impact of the intervention on academic attainment was evaluated by a team at NIESR:

Dr Heather Rolfe (formerly Associate Research Director, NIESR) led the implementation and process evaluation from the start of the project. She took over the role of PI on the project in 2016.

David Wilkinson (NIESR Fellow) was the original PI for the project until 2016. He has also provided expert advice on the impact evaluation during its latter stages.

Dr Cinzia Rienzo worked on the impact analysis and took over leading this element of the project in 2016, including during the period when interim reports on the evaluation were submitted to the EEF (for internal review).

Dr Matthew Bursnall (formerly Senior Economist, NIESR) replaced Cinzia Rienzo as the lead for the analysis of impact on academic outcomes in 2018.

Lucy Stokes (Principal Economist, NIESR) took over leading the impact analysis of academic attainment from 2019 and has managed the evaluation of academic outcomes through the final analysis and reporting stages. She had also previously contributed to the impact analysis for the second interim report while Cinzia Rienzo was on maternity leave.

Johnny Runge (Principal Social Researcher, NIESR) has contributed to the implementation and process evaluation, primarily in the reporting stages.

Jonathan Buzzeo (formerly Social Researcher, NIESR) contributed to the implementation and process evaluation during his time at NIESR.

The impact of the evaluation on health-related outcomes was undertaken by Professor Alistair McGuire and Dr Grace Lordan at LSE. The findings from this evaluation were published in a separate report.

The project as a whole has been overseen by an advisory group, chaired by Lord (Professor) Richard Layard.

Methods

Trial design

Table 3 provides an overview of the key aspects of the trial design.

Table 3: Trial design

Trial design, including number of arms		Two-arm, cluster randomised
Unit of randomisation		School
Stratification variable (s) (if applicable)		Proportion of pupils eligible for FSM; GCSE attainment; single or mixed sex school
Primary outcome	Variable	GCSE attainment
	Measure (instrument, scale, source)	Attainment 8 (range 0–90, NPD)
Secondary outcome(s)	Variable(s)	Maths and English attainment Exclusions Absenteeism
	Measure(s) (instrument, scale, source)	Maths and English attainment, end Year 7 and end Year 8 (Hodder Education Access Reading and Maths Tests, age-standardised scores, range 70–130) Exclusions, Year 7 and Year 8, total number of fixed exclusions for academic year (NPD) Absenteeism, Year 7 and Year 8, number of absence sessions (overall absence) during the academic year (NPD)
Baseline for primary outcome	Variable	Attainment at end KS2
	Measure (instrument, scale, source)	KS2 average total points score (range 0–39, NPD)
Baseline for secondary outcome(s)	Variable	Maths: KS2 maths points score English: KS2 reading score* Exclusions, end KS2 Absenteeism, end KS2
	Measure (instrument, scale, source)	Maths: KS2 maths points score (range,0–51, NPD) English: KS2 reading score (range 0–50, NPD) Exclusions at end KS2, number of fixed exclusions for academic year (NPD) Absenteeism at end KS2, number of absence sessions during the academic year (NPD)

* In the SAP we stated that we would use KS1 point scores for reading due to a change in the KS2 reading score variables available within the NPD over time; however, ultimately we were able to use a consistent KS2 measure across all analysis years and therefore it was no longer necessary to use data from KS1 instead.

This evaluation was an efficacy stage trial, using a cluster randomised trial design, with school-level randomisation. There were two trial arms, with schools randomised to treatment or control. As noted earlier, recruitment proved challenging, particularly given the timing of the start of the study and the length of time required to engage and recruit schools, and thus recruitment took place in two phases. For schools in phase one, the control group forms a waitlist control group continuing with business as usual for the cohort of pupils beginning Year 7 in September 2013 and then implementing the programme with the cohort of pupils beginning Year 7 in September 2014 (the following school year). For schools in phase two, the control group continues with business as usual and does not receive the intervention.[5] No incentive payments were provided.

The original intention was to recruit around 30 schools. Ultimately, thirteen schools were recruited to the project in the academic year 2012/2013, forming phase one. Phase one schools assigned to the treatment group (N = 7) implemented the intervention with their Year 7 pupils in the academic year 2013/2014, while those schools assigned to the control group (N = 6) did not implement the intervention in this year. Instead, the phase one schools allocated to the control group implemented the intervention with their Year 7 pupils in the following academic year, 2014/2015 (the waitlist control group). Thus, these phase one control schools become part of the treatment group for the cohort of Year 7 pupils starting in 2014/2015.

Twenty-one phase two schools entered the project in the academic year 2013/2014. Phase two schools assigned to the treatment group (N = 11) implemented the intervention with their Year 7 pupils in the academic year 2014/2015, while phase two schools assigned to the control group (N = 10) did not implement the intervention.

In addition, for both phases, those pupils who were in Year 7 in the year preceding the implementation of the intervention also form part of the control group in order to boost the sample size for analysis (it was always the intention to include this additional year group in the evaluation from the start of the project). For phase one schools, this is pupils who were in Year 7 in 2012/2013, while for phase two schools this is pupils who were in Year 7 in 2013/2014. This is summarised in Table 4. The design of the trial, which includes both treatment and control cohorts within some of the same schools, does introduce some risk of contamination, which, if present, would reduce the ability of the evaluation to detect an impact. This possibility was discussed in the early stages of the trial design but it was not considered to be a significant risk.

Table 4: Summary of treatment and control groups

Intervention group	**N school year groups (N school year groups after dropout)**
Phase one schools allocated to intervention Cohort 2013—pupils in Year 7 in 2013/2014	7
Phase one schools (previously waitlist control group)* Cohort 2014—pupils in Year 7 in 2014/2015	5
Phase two schools allocated to intervention Cohort 2014—pupils in Year 7 in 2014/2015	11
Control group	
Phase one schools allocated to control Cohort 2013—pupils in Year 7 in 2013/2014	6 (5)
Phase one schools (all) Cohort 2012—pupils in Year 7 in 2012/13	13 (12)

[5] If a waitlist control group had also been utilised in phase two, this would have meant waiting an additional year for attainment data if this group were subsequently to be included in the evaluation as a further treatment group.

Phase two schools allocated to control Cohort 2014—pupils in Year 7 in 2014/2015	10 (9)
Phase two schools (all) Cohort 2013—pupils in Year 7 in 2013/2014	21 (20)

*One of the schools in the waitlist control group withdrew and thus ultimately never became part of the treatment group.

In total, therefore, 34 schools were initially recruited to be part of the study and were randomised, comprising 73 school year groups as shown in Table 4. In practice, two of these schools dropped out (after randomisation); the resulting numbers of school year groups are shown in parentheses in the table. Thus 32 schools, forming a total of 69 school year groups, are included in the analysis (as discussed within the participant flow section of this report).

The primary outcome for the trial is Attainment 8, which captures GCSE attainment at the end of Key Stage 4. Secondary outcomes include performance on the Hodder tests, as assessed at the end of Year 7 and at the end of Year 8, as well as exclusions and absenteeism in Year 7 and Year 8. As discussed earlier, the Hodder assessments were initially intended to form the outcomes for the trial but given the considerable challenges faced in collecting these measures, and the burdens this was placing on participating schools, the decision was taken to stop these assessments after Year 8 and instead use a measure of GCSE attainment as the primary outcome measure (retaining the Year 7 and Year 8 assessments as secondary outcomes). The choice to use Attainment 8 was determined at the point of drafting the SAP. Further details of outcomes are given in the Outcome measures section below.

Participant selection

The process of identifying participating schools is described in detail in Lordan and McGuire (2019). This involved first identifying a list of all state-funded secondary schools in 42 local authorities, mainly in the South Eastern area of England. Recruitment aimed to target schools with poor levels of attainment or above average levels of deprivation such that schools were scored based on existing measures of attainment and deprivation. The schools were scored from one to ten based on the decile in which they were located for each of the following indicators (using 2012 GCSE and school census data from the Department for Education):

- percentage of pupils making expected progress in English;
- percentage of pupils making expected progress in maths;
- percentage of pupils gaining at least five GCSEs at grade C or better including English and maths; and
- percentage of FSM pupils.

A school, therefore, scored 40 if in the lowest (worst) decile for progress and attainment at GCSE and in the highest decile for the percentage of FSM pupils. Some schools were excluded due to missing data or because they were involved in similar projects, leaving 174 schools scoring 22 and above, Schools with a score above this threshold were sent a letter inviting them to participate and those schools expressing interest were sent further information about the project and evaluation requirements. The aim was to recruit schools with poor attainment serving pupils with above-average levels of deprivation.

A total of 34 schools were recruited comprising 73 school year groups of which 24 were allocated to the treatment group and 49 to the control group.[6] As noted above, recruitment took place over two phases with 13 schools recruited in phase one and 21 in phase two. Schools were mainly located in the South East and the Midlands.

[6] This larger number of control year groups relative to treatment year groups is in large part driven by the fact that the design includes preceding year groups as a further control in all participating schools.

The programme began with the cohort of Year 7 pupils in the year that schools were assigned to begin the intervention, and then continued with this cohort over the following four years of school (that is, the programme is delivered to this cohort of pupils over Years 7 to 10). All pupils within the relevant year group were eligible to participate.

The Bounce Forward team were responsible for recruiting schools to take part in the study.

Outcome measures

Primary outcome

Our primary outcome measure is Attainment 8 (this is measured using variable KS4_ATT8 from the NPD, range 0–90). Attainment 8 is one of the headline measures used by the Department for Education to assess school performance at the end of Key Stage 4. It measures the achievement of a pupil across eight qualifications including mathematics (double weighted) and English (double weighted), three further qualifications that count in the English Baccalaureate (EBacc) measure, and three further qualifications that can be GCSE qualifications (including EBacc subjects) or any other non-GCSE qualifications on the DfE approved list. Each individual grade a pupil achieves is assigned a point score, which is then used to calculate a pupil's Attainment 8 score (see DfE, 2016, for further details of how Attainment 8 is calculated).

As noted, GCSE attainment was not originally the intended outcome measure; after challenges in administration of the Hodder assessments at the end of Year 7 and Year 8, and concerns over the burdens this was placing on schools, it was decided (as documented in the 2016 updated trial protocol) to instead use GCSE attainment as the primary outcome. The exact measure was, however, not recorded at this time. Attainment 8 was chosen following discussions with both the EEF, the delivery team, and steering group, with the decision documented in the SAP. This measure was chosen on the grounds that it captured attainment across a broader range of subjects (rather than, for example, just focusing on performance in GCSE English or maths), was a measure that is meaningful to schools, and also as it has been used in other EEF trials.

As schools were recruited in two phases, and with multiple cohorts within schools, it is important to be aware of changes in the form of GCSE examinations across years. The analysis involves pupils who sat their GCSEs in the summer of 2017, 2018, and 2019. In summer 2017, the first reformed GCSEs were introduced for English language, English literature, and maths and graded on a new scale running from 9 to 1; in 2018, reformed GCSEs were introduced for a further 20 subjects and by 2019, almost all new GCSEs were in place with points awarded on a 9 to 1 scale. An Attainment 8 score is available from the NPD for all three years but it is important to be aware of these underlying changes in GCSE examinations. In order to inform whether this may have impacted on findings, we run a sensitivity analysis where models are estimated separately for the different years (further details are provided within the statistical analysis section). As noted earlier, the move to using GCSE attainment as the primary outcome measure also meant that there was a full school year between the point at which pupils stopped receiving the intervention and the point at which attainment was assessed.

The data was accessed through the NPD and linked to the pupil data collected in the trial through matching based on a combination of pupil name, UPN, date of birth, school name, and reference number (not all details were necessarily available for all pupils). The required NPD variables were made available to the evaluation team, matched to a unique identifier, within the ONS Secure Research Service (SRS) environment. The evaluation team then used this unique identifier to link to additional data collected as part of the trial (including, for example, treatment allocation).

For a small number of pupils, more than one Attainment 8 measure was recorded in the NPD; we keep the results from the year in which they would have been in Year 11. Where this still resulted in multiple records per pupil, we retain the record with the highest score for use in the analysis.

Secondary outcomes

Our secondary outcomes are:

Reading and maths skills—Year 7 and Year 8

Reading and maths skills in Year 7 and Year 8 were measured using the Hodder tests.[7] These were administered specifically for the project by the evaluation team. They are commercial tests and so the instruments cannot be included within the evaluation report.

Pupils were randomly assigned to take either the reading or maths test, so half of each year group in each school took each of the tests, with the aim of reducing the testing burden in schools. Randomisation was undertaken using a random number generator (the –runiform()- command in Stata).

The reading test consisted of ten sections (with a total of 60 questions) with each aiming to assess four different aspects of reading: literal comprehension, vocabulary, inference, and analysis. The final (raw) score was given by the sum of all four subset scores. The maths test comprised 45 questions aiming to test seven different aspects of maths: using and applying mathematics, counting and understanding numbers, algebra, calculating, understanding shape, measuring, and handling data. The final (raw) score was given by the sum of all seven subset scores.

The analysis of both the reading and maths scores uses the standardised scores, which have a range of 70 to 130.

For the reading test, a time limit of 30 minutes was set and for the maths test, a time limit of 45 minutes was set. Until June 2015, tests were in digital format, however, due to technical difficulties with the digital version of the tests, from autumn 2015 paper based versions of the tests were used. Marking was conducted by external markers. All tests were completed on a whole-class basis under exam conditions. In terms of the psychometric properties of the Hodder tests, the **EEF attainment measures database** rates construct validity and reliability of the reading test at two out of four, with high internal consistency (Cronbach alpha of around 0.95 and 0.94 for Forms A and B, respectively, which are used with the lower secondary age group). No evidence was available on criterion validity. The same ratings are given for the maths test.

At the time of preparing the original trial protocol, the reading measure formed the primary outcome for the trial and maths formed a secondary outcome. As noted earlier, the administration of these assessments proved challenging and was very burdensome for schools; the decision was therefore taken to stop assessments after Year 8 and to instead make use of administrative data from the NPD on attainment at the end of KS4 as the primary outcome. This change is documented in the updated trial protocol in 2016 and in the Statistical Analysis Plan for the evaluation. Reading and maths were chosen as key attainment variables of interest and the Hodder assessments in particular chosen as the measure for largely pragmatic reasons, being suitable for use with the relevant age groups and thought to be practical to administer in schools by the evaluation team.

Year 7 and Year 8 absence

Year 7 and Year 8 absence is measured using the number of overall absences in the relevant school year, as available from the NPD (measured using the variable OverallAbsence_5HalfTerms_ab). Overall absence is defined as the sum of both authorised and unauthorised absences (DfE, 2019). We use the measure that is based on five half terms (rather than the six half term measure, which also includes the final half term of the school year), as this version of the variable is available for all years required for our analysis and so enables consistency over time. Prior to the school year 2012/2013, absence information was only collected for the first five half terms of the school year (DfE, 2019). This level of detail was not apparent at the time of writing the SAP and prior to receiving the NPD data; thus, while the SAP specifies that the overall absence measure will be used, it does not specify whether this is the five or six half term measure.

Year 7 and Year 8 exclusions

[7] https://www.hoddereducation.co.uk/

Year 7 and Year 8 exclusions is measured using the number of total fixed exclusions per pupil, per school year, as available from the NPD (measured using the variable TotalFixedExclusions_ex). This data covers full school years (that is, there is not the same issue regarding the number of half terms covered as applies for absence data).

Absence and exclusions were not specified as outcomes in the original trial protocol but were included as additional secondary outcomes in the update to the trial protocol in 2016. The focus was on outcomes for Year 7 and Year 8, in line with the timepoints for the reading and maths assessments. The inclusion of absence and exclusions allows the exploration of additional outcomes that do not measure attainment but are important outcomes for pupils in their own right, and ones on which the programme may reasonably be expected to have an impact. Data on absence and exclusions from the NPD was obtained and matched in the same way as described for Attainment 8 above.

Baseline measures

For the primary analysis, the baseline measure is attainment at the end of KS2 as measured by the variable ks2_ks2apsfg (average point score – fine grading) available from the NPD. This is based on KS2 reading and maths results. We had originally intended to use the standard average point score (without fine grading, variable KS2_APS) but this was not available in the NPD for all years of our analysis. Details of the calculation of the fine grading measure are provided in DfE, 2020, but in summary, the fine grades use the underlying test marks achieved at KS2 to obtain a finer (more detailed) measure. The fine grade measure is thus more detailed and should capture more variation than the average point score without fine grading.

Alternative baseline measures are used for each of the secondary outcomes, as documented earlier in Table 3, but in summary, the baseline measures are:

- for absence: overall absence in Year 6 (five half terms measure);
- for exclusions: total fixed exclusions in Year 6;
- for reading: KS2 reading score; and
- for maths: KS2 maths score.

Sample size

At the point of preparing the original trial protocol, the power calculations assumed 160 pupils per year group per school, based on analysis of year group size in the preceding year in the selected schools. The initial plan was to assess outcomes using standardised tests of maths and English skills such that 80 pupils per school would take the reading test and 80 would take the maths test. The protocol power calculations were therefore based on these numbers. These also assumed an ICC of 0.13, a significance level of 0.05, and power of 0.8. Assumptions around the inclusion of a pre-test are not specified in the protocol. Based on these figures, the required number of schools to detect an effect size of 0.3 standard deviations is 23, while to detect an effect size of 0.25 requires 32 schools (the scenario presented in Table 8 in the participant flow section later in this report). Meta analysis of similar programmes (Durlak et al., 2011) indicates an average effect size of 0.27. Power calculations were not reported for the FSM subgroup in the trial protocol (and the trial was not designed to be powered to detect an effect on the FSM subgroup) as this was not required to be stated at the time of writing the protocol. Although this did not inform the power calculations as these were produced based on the number of schools rather than the number of school year groups, the original design would have implied 24 school year groups assigned to the treatment condition and 32 assigned to the control condition.[8]

Ultimately, 34 schools were initially recruited to the trial across phase one and phase two. In addition, a special school was also recruited (and randomised) but it was decided that this school would be excluded from the analysis of academic attainment: this is not specified in the protocol but follows the approach taken for the evaluation of health outcomes by Lordan and McGuire (2019) and is documented in the SAP, although they are still included within the IPE. Thirty-two schools are included in the final analysis (see Participant Flow section) as two schools dropped out (and never provided the pupil data required to be able to be included in the academic attainment analysis). As noted above, following

[8] With the greater number of control year groups resulting from the inclusion of preceding year groups in all participating schools.

concerns around measurement burden and attrition resulting from testing burdens in the first two years of the trial, it was decided that attainment data would no longer be collected directly from schools and that instead the primary outcome would be KS4 attainment taken from the National Pupil Database. As data on KS4 outcomes should be available for (almost) all pupils per school, a greater number of pupils would be included in the primary analysis than originally anticipated (that is, assuming this data would be available for the initially assumed average of 160 pupils per year group per school).

The power calculations presented in the 'at randomisation' column of Table 8 are therefore based on assuming GCSE outcomes data is available for all pupils in a year group; in practice, this number stood at an average of 147 pupils rather than 160 (based on an estimated 10,685 pupils involved in total across 74 school year groups, giving an average, with rounding, of 147 pupils per year group). The total number of 10,685 has been estimated given that pupil data was missing for the two schools that withdrew; for these two schools, we estimate pupil numbers based on DfE school census figures. As documented in the Participant Flow section, this number differs from that presented in the SAP, which did not include pupil numbers for the two schools where this was missing, but also because some pupils in schools where data was available were later identified as duplicates. As noted above, assumptions around the inclusion of a pre-test are not stated in the protocol, however, for the power calculations at the point of randomisation presented in the SAP—and in Table 8 of this report—we assume a pre-test-post-test correlation of 0.5.[9] (a previous EEF evaluation using GCSE attainment as the primary outcome, although with different measures, found a pre-test post-test correlation of 0.58 (Greaves et al., 2017); we adopted 0.5 as a conservative estimate). On this basis, the MDES—keeping all other assumptions the same—stands at 0.26. The MDES for the FSM subgroup stood at 0.27 (based on the percentage of FSM pupils at the point of analysis (33%), updated from the calculations in the SAP, which assumed 15% of pupils would be eligible for FSM). Throughout, the MDES calculations are based on the number of schools, rather than school year groups, to allow for conservative estimation. Ultimately, as described in the participant flow section, at the point of analysis the final sample of pupils stood at 9,410, equivalent to an average year group size of 136 pupils. The pre-post test correlation was higher than anticipated at randomisation stage, standing at 0.68, and the ICC was slightly lower than assumed at 0.11. The MDES at analysis stage stood at 0.27 (and 0.28 among the FSM subgroup).

The trial is not powered to detect an effect on the FSM subgroup as the primary population of interest.

Randomisation

Randomisation was conducted using minimisation, where schools are allocated in a way that aims to minimise differences in the characteristics of the treatment and control groups (see for example, Torgerson and Torgerson, 2008). The characteristics used in the minimisation were whether the percentage of pupils eligible for Free School Meals (FSM) was less than 13%, between 13% and 25%, or greater than 25%; whether the percentage of pupils with five GCSEs at grades A*-C was below 59% or not; and whether the school was single sex or mixed. Minimisation was used for randomisation given the expected small sample size; the above characteristics were used with the aim of minimising imbalance across these aspects.

Phase one schools were randomised independently by the London School of Hygiene and Tropical Medicine (conducted by a team member previously at LSE but who had left prior to randomisation, see Lordan and McGuire, 2019), while phase two schools were randomised independently by the LSE. As described in Lordan and McGuire (2019), schools were allocated a unique identifier and randomisation undertaken in Excel using a random number generator routine, with schools randomly allocated to either treatment or control groups.

Analysis for the evaluation of impacts on academic outcomes was undertaken by a separate team at NIESR, but analysis has not been conducted blind to treatment allocation. The pre-specification of analysis in the SAP aims to minimise the risk of any bias.

[9] A previous EEF evaluation using GCSE attainment as the primary outcome, although with different measures, found a pre-test post-test correlation of 0.58 (Greaves et al., 2017); we adopted 0.5 as a conservative estimate.

Statistical analysis

Primary analysis

The primary analysis is carried out using multilevel regression models to reflect the clustered nature of randomisation, following the approach set out in the protocol and SAP. Note that the SAP was updated most recently and that not all aspects detailed in the SAP were updated in a further amendment to the protocol.

As far as data allows, the main analyses are conducted on an intention-to-treat basis. As noted, while 34 schools were randomised, only 32 provided the necessary data to be included; our main analysis sample, therefore, comprises these 32 schools, all of which are included regardless of any further drop out or incomplete delivery of the programme. We also conduct a compliance analysis to explore the robustness of the main results to excluding those schools that withdrew from the programme or did not deliver this in full, as discussed further below.

Following the standard approach set out in the EEF guidance (EEF, 2018), the primary analysis model controls for prior attainment, treatment allocation, and the variables used in randomisation (proportion of FSM pupils, single or mixed sex school, prior school-level GCSE attainment—based on the definitions used for randomisation as described earlier). It also includes an indicator for the pupil cohort—that is, whether the pupil was in Year 7 in the school year 2012/2013, 2013/2014, or 2014/2015, with dummy variables included in the model for the latter two years and the first omitted to form the reference category.

Prior attainment is measured based on pupil attainment at the end of Key Stage 2 using the average total points score, also obtained through linkage to the NPD. The fine grading version of the average point score is used as this measure is available for all three cohorts in our sample. Note that this differs from what is stated in the SAP, which just refers to the average point score without fine grading as this variable was ultimately not available within the NPD for all years of our analysis, as noted. This should not have any substantive implications for the analysis.

The equation estimated is:

$$y = X\beta + Z\mu + \epsilon_{ij}$$

where:

y = vector of outcome scores;

X = covariate matrix (treatment group indicator, KS2 scores, variables used in randomisation, recruitment phase, and cohort indicator);

Z = design matrix identifying which school or cluster an individual attended;

μ = vector of school random effects;

β = fixed effect parameters;

ε_{ij} = residual error term for j-th member of cluster (school) i.

The analysis was conducted in Stata 16.

Secondary analysis

Secondary outcomes are analysed using essentially the same approach as for the primary outcome. As noted, the secondary outcomes include performance on the Hodder Education reading and maths tests as assessed at the end of Year 7 and at the end of Year 8. Thus, y in the model specified for the primary analysis is replaced with the score from the reading (or maths) assessment respectively, and the measure of prior attainment uses subject-specific measures of KS2 attainment rather than the KS2 average points score. All other elements of the model remain the same as for the primary analysis. Note that for reading, this represents a deviation from the SAP for the analysis of reading scores, where we had anticipated that only KS1 reading scores would be available within the NPD data provided. In practice,

KS2 reading scores were available and so we opted to use this measure for greater consistency with the approach for maths scores; this should also provide a more recent measure of a pupil's prior attainment.

Additional secondary outcomes are exclusions and absenteeism in Year 7 and Year 8; for these models, rather than prior attainment, the model includes exclusions and absenteeism at end KS2. All other aspects of the model remain the same as for the primary analysis. As discussed, our measure of absence is based on the five half term measure for consistency across all years; this level of detail was not specified in the SAP. The SAP specified that we would adopt the same model as for the primary outcome but the nature of the absence and exclusions outcomes means that a linear model may well be less appropriate. For this reason, we have undertaken an additional analysis applying multilevel negative binomial models for these outcomes. Results are presented in Appendix H and also briefly discussed together with our main absence and exclusions analysis in the Impact Evaluation results section.

Analysis in the presence of non-compliance

An indicator of compliance was constructed based on information collected by the delivery team as to whether the intervention had been delivered to their satisfaction, effectively, and whether they delivered the programme in full. Thus, compliance is defined at school level. The delivery team rated schools as 'red', 'amber', or 'green', with this rating informed by the extent to which schools had complied with staff training requirements, the timetabling and delivery of lessons over the four-year programme, and the extent to which there was senior buy-in within the school.

- A red rating indicated that a school dropped out of delivering the lessons (regardless of reason and at any stage during the four years).
- An amber rating was given where there was inconsistency in delivery or staff turnover, where it was unclear how much time had been allocated to teaching the lessons, or where there was a lack of buy-in from senior leaders.
- A green rating reflected commitment to the project throughout delivery, with lessons taking place as planned, consistency in staffing, and the programme valued even if there may have been other school changes such as a change in school leadership.

It was possible for a rating to change over time, for example, a school could have initially been rated green but then become red if it dropped out. The final ratings are used for the compliance analysis.

This information is used to form a binary indicator, taking a value of one if the school is considered to have delivered the programme to the satisfaction of the Bounce Forward team (that is, rated as green or green/amber) and equal to zero otherwise (as in the compliance analysis conducted for the separate evaluation of health outcomes by Lordan and McGuire, 2019). Thus, this measure was pre-specified prior to analysis. We use this indicator to conduct a complier average causal effect analysis, for the primary outcome only. We estimate the CACE using two stage least squares (2SLS) regression by estimating a (first stage) model of compliance using the binary measure of compliance described above. The predicted values from the first stage are then used in the estimation of a model of our outcome measure. This analysis is conducted using the `ivregress` command in Stata, which allows the necessary adjustments to be made to standard errors in response to the clustered nature of the data. We also run a further sensitivity check, excluding those schools that were not recorded as delivering the programme to the satisfaction of the delivery team from the analysis; that is, we run the same specification as for the primary analysis but excluding this group of schools.

Imbalance at baseline

To explore whether there is imbalance, we report on the characteristics of the sample both at randomisation and at the point of analysis. This includes school characteristics—including Ofsted rating, school type (academy status), urban/rural location, number of pupils, percentage of FSM pupils, percentage of pupils with English as an additional language (EAL), and percentage of pupils with special educational needs (SEN)—and pupil characteristics: age, gender, FSM eligibility, prior attainment, absenteeism, and exclusions. We follow the standard EEF template, with means and standard deviations reported for continuous variables and counts and percentages in each category given for categorical variables. In line with EEF guidance, differences in 'pre-test' variables are reported as effect sizes.

Missing data analysis

We report the number of complete cases (those without missing data). We consider separately missingness in outcome data and missingness in covariates, and report the distribution of missing observations by treatment arm.

The SAP specified that in the event of greater than 5% missing data at either cluster or individual level, or a significant difference in missingness between treatment and control arms, we would conduct further investigation into the mechanisms of missingness. In practice, if missingness is considered as a percentage of those schools that provided pupil data, then missingness for the primary outcome did not exceed this threshold. If, however, we consider missingness as a percentage of all schools randomised—which, as discussed in the earlier section on sample size calculations, has to be estimated—the level of attrition (and thus missingness) is much higher. However, this missingness is substantively driven by the two schools that did not provide pupil data, and without this data we do not have other pupil characteristics with which to explore associations with missingness. For these reasons we do not explore mechanisms of missingness in further detail, or conduct multiple imputation. We do, however, present the results of some sensitivity analyses. Specifically: first, we repeat the primary analysis on a complete case basis, second, we impute a school-level mean for prior attainment where this is missing, and third, we exclude the measure of prior attainment from the model. None of these analyses were pre-specified in the SAP.

Subgroup analyses

As identified in the trial protocol and SAP, the analysis was also conducted separately for the subgroup of FSM pupils. This subgroup was identified using the variable EVERFSM_6 from the NPD as the EVERFSM_6_P variable usually stipulated for EEF analyses was not available for all years of our analysis (this is not available on the NPD prior to 2015/2016, but this was not apparent at the time of preparing the SAP and thus represents a deviation). While both measures capture FSM eligibility in the previous six years, EVERFSM_6_P includes pupils identified as being eligible at any time in any of the termly or annual censuses while EVERFSM_6 includes those identified as eligible on census day; thus by using EVERFSM_6 we may be understating the number of FSM pupils eligible for FSM compared to EVERFSM_6_P.

We run separate models for the FSM and non-FSM subgroups, where we repeat the specifications for the primary and secondary analyses as described above, but for the relevant subgroup of pupils. We also run analyses interacting treatment allocation with FSM status where we repeat the specifications for the primary and secondary analyses but additionally include an interaction term between treatment allocation and FSM status. The effect size for FSM pupils is presented on the basis of the subgroup model.

We also explore differences according to level of prior attainment, interacting treatment allocation with prior attainment. In order to do this we construct a variable which allocates pupils to low, average, or high attainment based on whether they are in the bottom, middle, or top third at KS2, based on the average KS2 point score. We then repeat our models for the primary and secondary analyses but additionally include an interaction term between treatment status and this prior attainment variable.

All subgroup analyses are conducted both for the primary outcome and all secondary outcomes. In presenting the subgroup analyses it should be noted that the trial was designed to detect effects in the full sample rather than for subgroups and thus findings for these subgroups are exploratory.

Additional analyses and robustness checks

We run an additional model for the primary outcome that includes a broader set of control variables. This replicates the specification that was used for the interim analyses: that is, we run the same model as used for our primary analysis, but additionally controlling for pupil age in months, gender, and FSM eligibility as well as the proportion of SEN pupils in

the school and the proportion of EAL pupils.[10] We also run two further analyses controlling for aspects on which the sample appears unbalanced, the first repeating the primary analysis but additionally controlling for the percentage of SEN pupils in the school and FSM eligibility, and the second additionally controlling for the percentage of SEN pupils, FSM eligibility, and school type.

Some pupils will have moved school during the course of the project. As the primary analysis is on an intention-to-treat basis, all pupils are included but, as pre-specified in the SAP, an additional analysis tests the robustness of the results to including only pupils who remain in the same school throughout (that is, this is the same model as for the primary analysis but including only the applicable subsample of pupils). Given the length of the programme (four years), with the primary outcome then assessed one year later, it was deemed important to include such a sensitivity analysis. We construct two measures for this purpose: the first excludes any pupils who we can identify as having left the school in either their first year (Year 7) or before the January of their Year 11 (based on the school spring census date for the relevant year). This will understate pupils who have left, but information on leaving dates was not available for all years. Our second measure compares whether pupils were at the same school (based on the variable LAESTAB) in Year 7 and Year 11. Again, this is not a perfect measure as some schools may have changed LAESTAB over time (and thus it may overstate the extent of pupils leaving). On the basis of the first measure, just under 1% of pupils in the primary analysis sample are excluded from the analysis, while the second measure excludes around 16% of pupils from the analysis.

Finally, due to the multiple cohorts (year groups) involved in the study, we also conduct a sensitivity analysis for the primary outcome to check whether the results are affected if analysis is confined to pupils within the same year group (acknowledging that such an analysis is inevitably based on a smaller sample size). That is, we run the same model used in the primary analysis, but separately for the sample of pupils that form the 2013 cohort, and then separately for those pupils that form part of the 2014 cohort.

Estimation of effect sizes

Effect sizes are calculated using Hedges g, following the standard approach for EEF trials:

$$\boldsymbol{ES} = \frac{(\boldsymbol{Y_T} - \boldsymbol{Y_{C)adjusted}}}{\boldsymbol{S}*}$$

where s* is the pooled unconditional variance of the treatment and control groups. In practice, this is calculated by dividing the coefficient from the regression model by the total unconditional variance obtained from running an empty hierarchical model for the relevant outcome without covariates, in line with Hedges (2007); 95% confidence intervals are calculated following the approach for cluster randomised trials using multilevel models as in equations (19) and (20) specified in Hedges (2007). This is consistent with the approach available for cluster randomised trials within the eefAnalytics package[11].

Estimation of intraclass correlation coefficients

We report the ICCs for the pre-test and post-test at school level using empty hierarchical linear models including school-level random effects as follows:

$$\boldsymbol{Y_{ij}} = \boldsymbol{\beta_0} + \boldsymbol{\eta_j} + \boldsymbol{\varepsilon_{ij}}$$

where $\boldsymbol{Y_{it}}$ is the pre- or post-test of individual $\boldsymbol{i}$ in school $\boldsymbol{j}$, $\boldsymbol{\beta_0}$ is a constant term, $\boldsymbol{\eta_j}$ is a school-level random effect, and $\boldsymbol{\varepsilon_{ij}}$ is an individual-level idiosyncratic error term. The ICC estimate is recovered as follows:

[10] A summary of the findings from the interim reports is provided in Appendix I.
[11] See for example the stata module for eefAnalytics (Vallis *et al*., 2022), available at: **https://ideas.repec.org/c/boc/bocode/s458904.html**

$$ICC = \frac{var(\eta_j)}{var(\eta_j) + var(\varepsilon_{ij})}$$

Longitudinal analysis

No longitudinal analysis is conducted.

Implementation and process evaluation

The principal aims of the process evaluation were to understand teachers' and pupils' perceptions of the programme and identify factors affecting its implementation. The evaluation aimed to identify features that appeared to contribute to successful implementation, including the effectiveness of training, practical issues, and the influence of factors such as staff confidence in the programme and senior management commitment. Qualitative research within the process evaluation also aimed at bringing greater clarity to the quantitative research findings and understanding the reasons for, and mechanisms behind, the estimated impacts.

The process evaluation aimed at understanding participants' views of the training and experiences during implementation and the application of Healthy Minds at classroom level. In addition to understanding teachers' experiences of the programme, we also obtained the perspectives of pupils through focus groups. We did this to gain further insight into its actual and potential impact than would be possible through analysis of test scores alone. The purpose of the focus groups was to explore and assess pupils' recall and perspectives on the programme and its components. They were also aimed at identifying any lessons for future delivery of the programme. Within focus groups for Year 8 pupils in phase one schools, we asked pupils to work in pairs to produce a leaflet for new Year 7s explaining what to expect from the Healthy Minds programme. This method was only used for Year 8s because it was felt that older pupils would be more able to manage and express their views within a discussion-only group.

The implementation and process evaluation (IPE) was carried out at two points in the life of the project:

- in 2015, in the second year of the project for phase one schools when pupils were in Year 8; and
- in 2017, in the third year of the project for phase two schools and fourth year for phase one schools; at this stage, pupils were in Years 9 and 10, respectively.

The process evaluation therefore captured schools', teachers', and pupils' experiences at various stages of the project, with return visits to three of the schools allowing teachers and pupils to reflect on the development and embedding of the project over the course of up to three years of implementation and spanning Years 7 to 10.

The IPE consisted of the following components:

- evaluation of teacher preparation and training through:
  - observation at one of the five days of training in June and July 2013; and
  - framework analysis of training evaluation data collected by the internal evaluation team at Bounce Forward; and

- evaluation of the implementation, its fidelity, and perceived impact through:
  - visits to all seven intervention schools in phase one of the programme during February and March 2015—at this point, schools were in the fifth term of delivering Healthy Minds and were using it in Years 7 and 8, though the focus of the trial in these schools was Year 8; we carried out interviews with seven leads and 12 additional teachers of the programme, a total of 19 teachers and also carried out a pupil focus group in each of the seven schools, with a total of 57 pupils;
  - analysis of anonymised data on teachers' experiences of teaching the Healthy Minds curriculum elements, collected by the Bounce Forward delivery team in order to understand the impact of each element; and
  - visits to eight schools between February and June 2017—three phase one schools visited in 2015 and five phase two schools visited for the first time; we carried out interviews with seven leads and ten additional teachers of the programme, a total of 17 teachers and carried out a pupil focus group in each of the eight

schools, with a total of 81 pupils; in phase one schools these were in year ten and in phase two schools pupils were in Year 9.

Table 5 shows the number of interviews carried out with teachers and pupils across the 12 schools that took part in the process evaluation. Three schools (Schools 1, 3, and 4) took part in both waves of visits.

Table 5: Case study schools characteristics

Wave one evaluation (spring 2015)			
Case study number	**ID**	**Interviewees**	**Pupil focus group**
School 1	School 1	Healthy Minds lead and teacher Healthy Minds teacher x 2	8 pupils from Year 8
School 2	School 2	Healthy Minds lead and teacher	10 pupils from Year 8
School 3	School 3	Healthy Minds lead and teacher Healthy Minds teacher	6 pupils from Year 8
School 4	School 4	Healthy Minds lead Healthy Minds teacher x 2	8 pupils from Year 8
School 5	School 5	Healthy Minds lead Healthy Minds teacher x 3	9 pupils from Year 8
School 6	School 6	Healthy Minds lead Healthy Minds teacher x 2	8 pupils from Year 8
School 7	School 7	Healthy Minds lead Healthy Minds teacher	8 pupils from Year 8
Wave two evaluation (spring 2017)			
Case study number (phase)	**ID**	**Interviewees**	**Pupil focus group**
School 1 (phase 1)	School 1	Healthy Minds Lead Healthy Minds teacher x 2	8 pupils from Year 10
School 3 (phase 1)	School 3	Healthy Minds Lead Healthy Minds teacher	11 pupils from Year 10
School 4 (phase 1)	School 4	Healthy Minds teacher	13 pupils from Year 10
School 8 (phase 2)	School 8	Healthy Minds Lead Healthy Minds teacher	10 pupils from Year 9
School 9 (phase 2)	School 9	Healthy Minds Lead Healthy Minds teacher x 4	8 pupils from Year 9
School 10 (phase 2)	School 10	Healthy Minds Lead	10 pupils from Year 9
School 11 (phase 2)	School 11	Healthy Minds Lead	8 pupils from Year 9
School 12 (phase 2)	School 12	Healthy Minds Lead Learning Assistant	13 pupils from Year 9
Total participants		**36 staff**	**138 pupils**

The process evaluation is based largely on the data collected by the evaluation team, informed by data collected by the delivery team. The Bounce Forward team collected survey data from teachers on their experiences of teaching the Healthy Minds curriculum elements as well as training evaluation surveys, which was then anonymised and sent to NIESR. This data was checked for consistency with our own qualitative and observation findings and no discrepancies were found, which makes us confident in using it for the evaluation. Our own data, collected through interviews with teachers and pupils, was analysed using a framework approach. This involves analysing qualitative data in a written form and is therefore appropriate for the analysis of transcripts of interviews with teachers and pupils as well as research notes taken during observation of training. Qualitative responses to training evaluation surveys were also analysed in this way.

Our data analysis entailed first coding the data into themes and issues; codes were a mixture of predetermined ones, developed during the design of the process evaluation and taking account of the aims of the intervention, as well as those that emerged from the text of transcripts and observations. Codes identified different types of information, for

example, experiences of the training and of putting the approach into practice as well as other aspects such as values and feelings. Themes identified in the analysis of the pupil focus groups were: learning environment, behaviour management, repetition versus building blocks, relevance, resources/differentiation, decision-making, continuity, outside reinforcement, and evidence of impact. Throughout the analysis process, we looked for similarities and differences in the data. In the final stage of analysis we restructured our document to follow the format required by the EEF.

Steps were taken to ensure that the case study schools included a variety of delivery contexts, such as Ofsted rating, proportion of FSM pupils, and location. Overall, we visited and interviewed teachers and pupils in 12 out of 23 treatment schools. In addition, we have included the findings from a special school that we visited but was excluded from the impact analysis. We have clearly flagged when findings are specific to this school. Overall, the findings may not necessarily reflect the views of the wider population of treatment schools. Nevertheless, we believe the qualitative data collected through the visits provide useful insights into the range and diversity of views and the experience of participants in the intervention. The findings of the process evaluation should be considered with these strengths and limitations in mind.

Table 6: IPE methods overview

Research methods	Data collection methods	Participants/ data sources	Data analysis methods	Research questions addressed
Case study visits wave one	Interviews and focus groups	19 teachers; 57 pupils	Framework analysis	Fidelity, implementation, impact
Case study visits wave two	Interviews and focus groups	17 teachers; 81 pupils	Framework analysis	Fidelity, implementation, impact
Observation training day	Observation	One training day	Framework analysis	Fidelity, implementation
Analysis of training evaluation data by delivery team	Survey	Delivery team	Framework analysis	Fidelity, implementation
Analysis of teacher experience data by delivery team	Survey	Delivery team	Framework analysis	Fidelity, implementation, impact

Costs

Costs were not collected separately for this evaluation but are reported in the evaluation of the impact of the Healthy Minds programme on health outcomes (Lordan and McGuire, 2019).

Timeline

Table 7 provides an overview of key points in the timeline for the intervention and evaluation of academic outcomes.

Table 7: Timeline

Dates	Activity	Staff responsible or leading
Jan–Sep 2013	Recruitment of schools	Bounce Forward
Apr–Sep 2013	Allocation to phase one schools (starting September 2013) and phase two schools (starting September 2014)	Randomisation conducted by LSHTM (phase one) and LSE (phase two)

Dates	Activity	Staff responsible or leading
Sep 2013–Jul 2017	HM programme delivery in phase one treatment schools	Bounce Forward and participating schools
Sep 2014–Jul 2018	HM programme delivery in phase one waitlist control schools and phase two treatment schools	Bounce Forward and participating schools
Feb–Mar 2015	Wave one case studies (interviews and focus groups)	Evaluation team
2016	Interim reports on impact on academic outcomes in Years 7 and 8	Evaluation team
Feb–Jun 2017	Wave two case studies (interviews and focus groups)	Evaluation team
Jun 2018	Pupils receiving programme in phase one treatment schools take GCSEs	n/a
Jun 2019	Pupils receiving programme in phase one waitlist control and phase two treatment schools take GCSEs	n/a
2021	Analysis and reporting of academic outcomes	Evaluation team

Impact evaluation results

Participant flow including losses and exclusions

As discussed earlier in the section on participants, a total of 174 schools were sent an invitation to participate in the study. Of these, 42 expressed interest in participating and 37 agreed to participate; three further schools withdrew at this stage leaving 34 schools that were randomised (Figure 1); two of these dropped out meaning that 32 are included in the final analysis. Only these 32 schools provided the necessary pupil data to be included within the analysis of academic attainment. As described earlier in this report, the figure of 10,685 pupils at the point of randomisation represents an estimate as pupil numbers were estimated for the two schools where pupil data was not provided.

The original files supplied for matching to the NPD (from across all of the 32 schools) contained 10,029 pupil records, of which 9,994 were matched to the NPD (99%). Hence the power calculations in the SAP for the point of randomisation refer to 10,029 pupils (this figure did not include estimates for the two schools not providing pupil data). However, some of these records were identified as duplicates and once these were removed, there were 9,825 unique pupil records remaining.

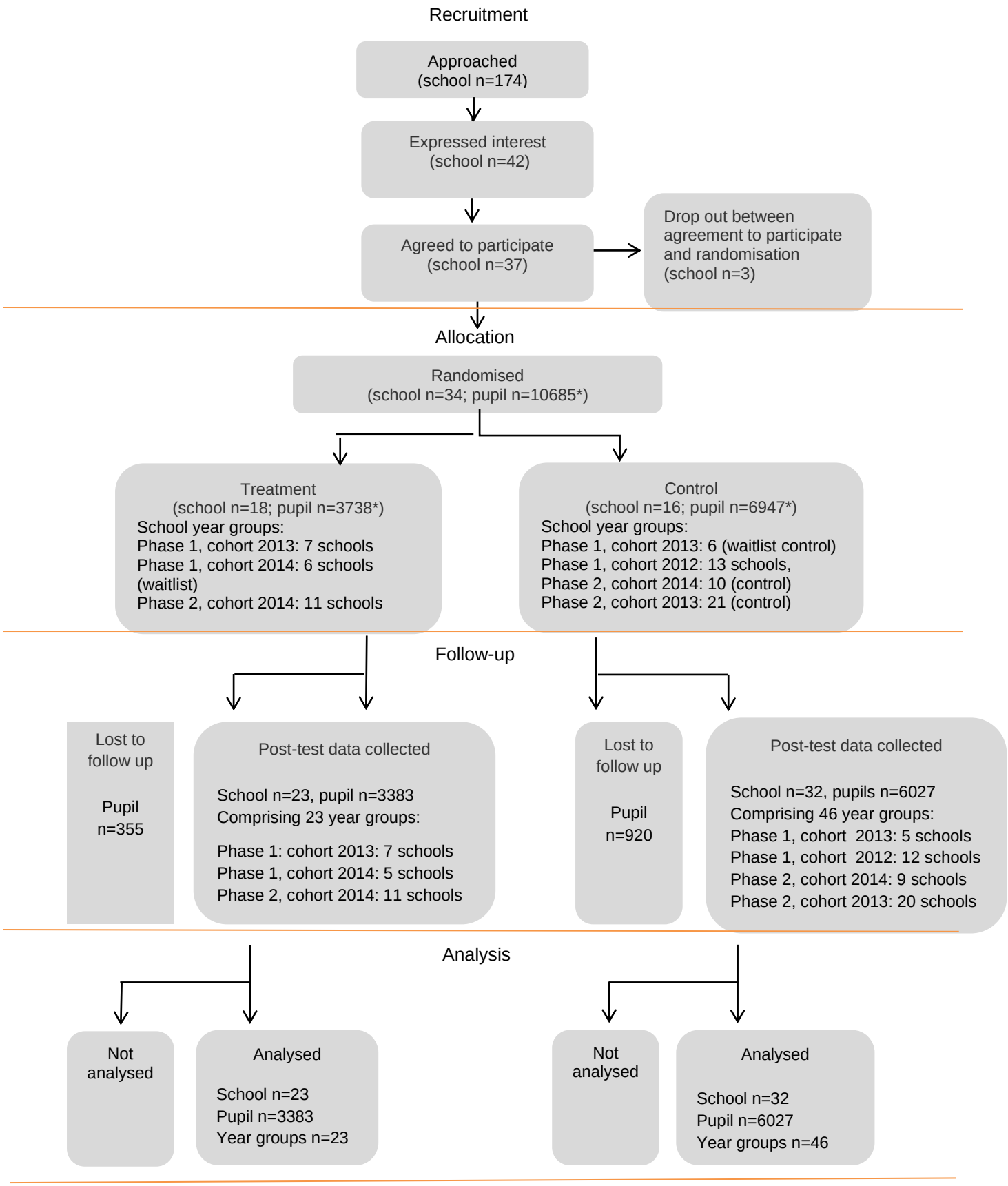
While 9,740 of these records were matched to the NPD, data on our primary outcome was available for a total of 9,410 pupils (this was all pupils with a record in the KS4 attainment files). This forms our sample for the primary analysis.

Schools can potentially contribute to both the treatment and control group for different pupil year groups, or cohorts (as set out earlier in trial design). Of the 34 schools at randomisation, 18 were allocated to the intervention group and 16 to the control group. For phase one schools, the control schools form a waitlist control group (this is six schools at point of randomisation).

All 32 schools forming the analysis sample contribute to the control group, with 23 schools contributing to the treatment group. Hence the figures in the participant flow diagram for follow-up do not sum to the total of 32 as the same schools are contributing to treatment and control groups in different years. In Figure 1 we also show the number of year groups contributing to both the treatment and control groups, by phase.

Figure 1 represents participant flow based on the availability of data for the primary outcome. Availability of data differs for the secondary outcomes considered and these sample sizes are presented within the later section on outcomes and analysis. This particularly affects the reading and maths assessments that were initially collected as academic outcomes. It was always the intention to randomise the completion of these assessments, so that only half of each year group would complete the reading test and half would complete the maths test, but, as noted elsewhere in this report, there were also significant issues with the administration of these tests. Sample sizes are shown within the tables presenting the results of the analysis for these outcomes.

Figure 1: Participant flow diagram—two arms



* Number of pupils randomised is estimated as the two schools which withdrew did not provide pupil data.

Table 8 summarises the MDES at different stages of the trial. At initial protocol stage, the MDES stood at 0.25. At the point of randomisation, 34 schools formed part of the trial and the MDES stood at 0.26; for the FSM subgroup, the MDES stood at 0.27.[12] At the point of analysis, 32 schools formed the analysis sample; data for some pupils was unavailable due to missingness of the primary outcome, but this was a relatively small proportion (see Table 9). This meant that at the point of analysis the MDES stood at 0.27 (and 0.28 for the FSM subgroup).

Table 8: Minimum detectable effect size at different stages

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.25	-	0.26	0.27	0.27	0.28
Pre-test/post-test correlations	Level 1 (pupil)	-	-	0.5	0.5	0.68	0.63
	Level 2 (class)	-	-	-	-	-	-
	Level 3 (school)	-	-	-	-	-	-
Intracluster correlations (ICCs)	Level 2 (class)	-	-	-	-	-	-
	Level 3 (school)	0.13	-	0.13	0.13	0.11	0.09
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		2	-	2	2	2	2
Average cluster size		80	-	147	48	136	46
Number of year groups	Intervention	-	-	24	24	23	23
	Control	-	-	50	50	46	46
	Total	-	-	74	74	69	69
Number of schools	Intervention	16	-	18	18	18	18
	Control	16	-	16	16	14	14
	Total:	32	-	34	34	32	32

[12] The figures for randomisation differ from those presented in the SAP as the figures here have been recalculated after excluding duplicates in the dataset, only found during the process of conducting the analysis, and including estimates for two schools where pupil data was not provided. The FSM figures were also updated to represent the actual percentage of pupils eligible for FSM observed at analysis.

Number of pupils	Intervention	1280	-	3738	1247	3383	1085
	Control	1280	-	6947	2319	6027	2056
	Total:	2560	-	10685	3566	9410	3141

Attrition

Table 9: Pupil-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised*	3738	6947	10685
	Analysed	3383	6027	9410
Pupil attrition (from randomisation to analysis)	Number	355	920	1275
	Percentage	9.5	13.2	11.9

* Number of pupils randomised is estimated as the two schools which withdrew did not provide pupil data.

Table 9 reports the overall rate of attrition from the trial, measured at pupil level. Based on the primary outcome, the final analysis sample comprised 9,410 pupils out of the estimated 10,685 pupils that formed part of the trial at randomisation. This represents an overall attrition rate of 12%. If we measure attrition based only on the 32 schools that had provided pupil data at the point of randomisation, the overall attrition rate stands at 4% as these 32 schools comprised 9,825 pupils in total (with 415 pupils lost from the sample as no KS4 attainment record could be matched for these pupils).

Pupil and school characteristics

Table 10a presents school and pupil characteristics for the intervention and control groups at the point of randomisation. In this trial, schools can potentially form part of either the treatment or control group in different years, depending on the cohort in question. For the purposes of Table 10a, schools are presented as treatment or control based on their original assigned allocation (so, phase one schools assigned to the control group are shown as control schools in Table 10 even though they later receive the intervention as a waitlist control group with their 2014 cohort of pupils).

Focusing first on the characteristics of the 34 schools at randomisation, we see that, compared with the national average at the time, a higher proportion had been rated as ‘requiring improvement’ or ‘inadequate’ by Ofsted at their most recent inspection. Most schools were located in urban areas, reflecting the area of recruitment, and there was a mix of school types, although the majority were academy schools.

The lower half of Table 10a presents the characteristics of pupils at the point of randomisation, by intervention and control group. These are presented on the basis of pupils’ actual assignments to treatment or control groups. Just under half of pupils in both intervention and control groups were female (48% and 47% respectively) while around a third were eligible for free school meals—33% and 36% respectively. These patterns were generally similar among our sample for the primary analysis (Table 10b).

Histograms of KS2 attainment, at both randomisation and analysis, are presented in Appendix E (Figures E.1 and E.2). These show fairly similar distributions across both intervention and control groups. However, in both samples, the mean KS2 score is higher among pupils in the intervention group, with an effect size of around 0.12.

Our analysis of secondary outcomes includes absence and exclusions, with these models including previous measures of absence and exclusions at the end of KS2 (Year 6). On the basis of Year 6 absence and exclusions data, for both we see slightly lower rates in the intervention group compared with the control group (histograms are presented in Appendix E).

Table 10a: Baseline characteristics of groups as randomised

School-level (categorical)	National-level mean	Intervention group		Control group	
		n/N (missing)	Count (%)	n/N (missing)	Count (%)
Ofsted overall effectiveness:[1]					
Outstanding/good	73%	9/16 (2)	9 (56.3%)	8/14(2)	8 (57.2%)
Requires improvement/inadequate	27%	7/16 (2)	7 (43.8%)	6/14(2)	6 (42.9%)
School type:[2]					
Academy converter	39%	5/18 (0)	5 (27.8%)	6/16 (0)	6 (37.5%)
Academy sponsor led/free school	18%	5/18 (0)	5 (27.8%)	4/16 (0)	4 (25.0%)
Community/foundation/voluntary aided/controlled school	38%	8/18 (0)	8 (44.5%)	6/16 (0)	6 (37.6%)
In urban area		15/17 (1)	15 (88.2%)	14/14 (2)	14 (100%)
School-level (continuous) [2]		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)
Number of pupils	968	18 (0)	1001 (352.9)	16 (0)	948 (376.4)
% FSM pupils in past 6 years	28.3%	18 (0)	29.7 (18.1)	16 (0)	34.8 (16.7)
% EAL pupils	13.8%	18 (0)	14.2 (15.2)	16 (0)	18.1 (18.8)
% pupils with SEN statement	5.3%	17 (1)	1.3 (0.82)	15 (1)	2.2 (1.1)
KS4: % achieving 5 A*-C including English and maths	60.6%	17 (1)	60.6 (15.0)	15 (1)	57.6 (15.0)
Pupil-level (categorical)		n/N (missing)	Count (%)	n/N (missing)	Count (%)
Female		1683/3505 (233)	1683 (48.0%)	2926/6208 (739)	2926 (47.1%)

Eligible for FSM		1160/3486 (252)	1160 (33.3%)	2167/6086 (861)	2167 (35.6%)	
2012 cohort		0/3738	0 (0%)	1882/6947	1882 (27.1%)	
2013 cohort		967/3738	967 (25.9%)	3925/6947	3925 (56.5%)	
2014 cohort		2771/3738	2771 (74.1%)	1140/6947	1140 (16.4%)	
Pupil-level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size
Age in months (September of Year 7)		3505 (233)	138.43 (3.50)	6208 (739)	138.48 (3.56)	
KS2 average point score (fine grading)		3423 (315)	28.3 (4.10)	5969 (978)	27.8 (4.40)	0.12
KS2 reading mark		3365 (373)	30.5 (8.9)	5792 (1155)	30.5 (9.8)	-0.00
KS2 maths mark		3375 (363)	69.2 (19.6)	5826 (1121)	67.6 (20.5)	0.08
Absence (end KS2)		3428 (310)	12.11 (14.04)	5986 (961)	12.93 (14.03)	-0.06
Exclusions (end KS2)		3508 (230)	0.012 (0.231)	6232 (715)	0.022 (0.233)	-0.04

[1] Ofsted overall effectiveness ratings as at 31 March 2013.
[2] As reported in DfE Performance Tables, 2013.

Table 10b presents the same characteristics at the point of analysis. As we lose a relatively small proportion of observations in our final analysis sample for the primary outcome, the characteristics presented in Table 10b are similar to those observed at baseline in Table 10a. Attrition has, therefore, not introduced any additional bias into the sample for the purpose of the primary analysis. Other measures obtained from the NPD, that is, absence and exclusions, also show relatively little missingness. However, there is considerable missingness when considering the Hodder assessments of reading and maths available at end Year 8. For analysis of these measures we use prior attainment of KS2; in the full sample at the point of randomisation, the sample appears balanced in terms of KS2 reading scores, although with higher KS2 maths scores among the intervention group compared to the control group (effect size of 0.08, see Table 10a). Within the available sample for analysis, KS2 reading scores were slightly lower among the intervention than control group (effect size of -0.04, see Table 10b) while KS2 maths scores remained higher (effect size of 0.10, Table 10b). These much smaller samples with available test scores for reading and maths, combined with signs that attrition may have impacted the nature of the samples, mean caution should be exercised around the results for these outcomes.

Table 10b: Baseline characteristics of groups as analysed

School-level (categorical)	National-level mean	Intervention group		Control group		
		n/N (missing)	Count (%)	n/N (missing)	Count (%)	
Ofsted overall effectiveness:[1]						

Outstanding/good	73%	9/16 (2)	9 (56.3%)	8/14(2)	8 (57.2%)	
Requires improvement/inadequate	27%	7/16 (2)	7 (43.8%)	5/14(2)	5 (41.7%)	
School type:[2]						
Academy converter	39%	5/18 (0)	5 (27.8%)	5/14 (0)	5 (35.7%)	
Academy sponsor led/free school	18%	5/18 (0)	5 (27.8%)	4/14 (0)	4 (28.6%)	
Community/foundation/voluntary aided/controlled school	38%	8/18 (0)	8 (44.5%)	6/16 (0)	6 (37.6%)	
In urban area		15/17 (1)	15 (88.2%)	13/13 (1)	13 (100%)	
School-level (continuous) [2]		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	
Number of pupils	968	18 (0)	1001 (352.9)	14 (0)	909 (386.4)	
% FSM pupils in past 6 years	28.3%	18 (0)	29.7 (18.1)	14 (0)	37.3 (16.3)	
% EAL pupils	13.8%	18 (0)	14.2 (15.2)	14 (0)	19.9 (19.5)	
% pupils with SEN statement	5.3%	17 (1)	1.3 (0.82)	13 (1)	2.4 (1.05)	
KS4: % achieving 5 A*-C including English and maths	60.6%	17 (1)	60.6 (15.0)	13 (1)	57.6 (16.1)	
Pupil-level (categorical)		n/N (missing)	Count (%)	n/N (missing)	Count (%)	
Female		1635/3383	1635 (48.3%)	2851/6020	2851 (47.4%)	
Eligible for FSM		1085/3364 (19)	1085 (32.3%)	2056/5905 (122)	2056 (34.8%)	
2012 cohort		0/3383	0 (0%)	1659/6027	1659 (27.5%)	
2013 cohort		940/3383	940 (27.8%)	3462/6027	3462 (57.4%)	
2014 cohort		2443/3383	2443 (72.2%)	906/6027	906 (15.0%)	
Pupil-level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size
Age in months (September of Year 7)		3383	138.43 (3.50)	6020	138.47 (3.54)	
KS2 average point score (fine grading)		3307 (76)	28.4 (4.05)	5796 (231)	27.8 (4.35)	0.12
KS2 reading mark[4]		1179 (53)	30.6 (8.8)	1674 (129)	30.9 (9.8)	-0.04

KS2 maths mark[4]		1209 (46)	70.0 (19.5)	1624 (126)	68.1 (20.5)	0.10
Absence (end KS2)[5]		3403 (79)	12.11 (14.03)	5971 (224)	12.93 (14.03)	-0.06
Exclusions (end KS2) [5]		3508 (68)	0.012 (0.231)	6232 (17)	0.022 (0.233)	-0.04

[1] Ofsted overall effectiveness ratings as at 31 March 2013.
[2] As reported in DfE Performance Tables, 2013.
[4] Based on analysis samples for Year 8 reading and maths analysis respectively.
[5] Analysed for both Year 7 and Year 8, thus we report on basis of lowest analysis sample (Year 8 in case of absence), although this is almost identical in the two samples for Year 6 (for exclusions they are identical).

Outcomes and analysis

Primary analysis

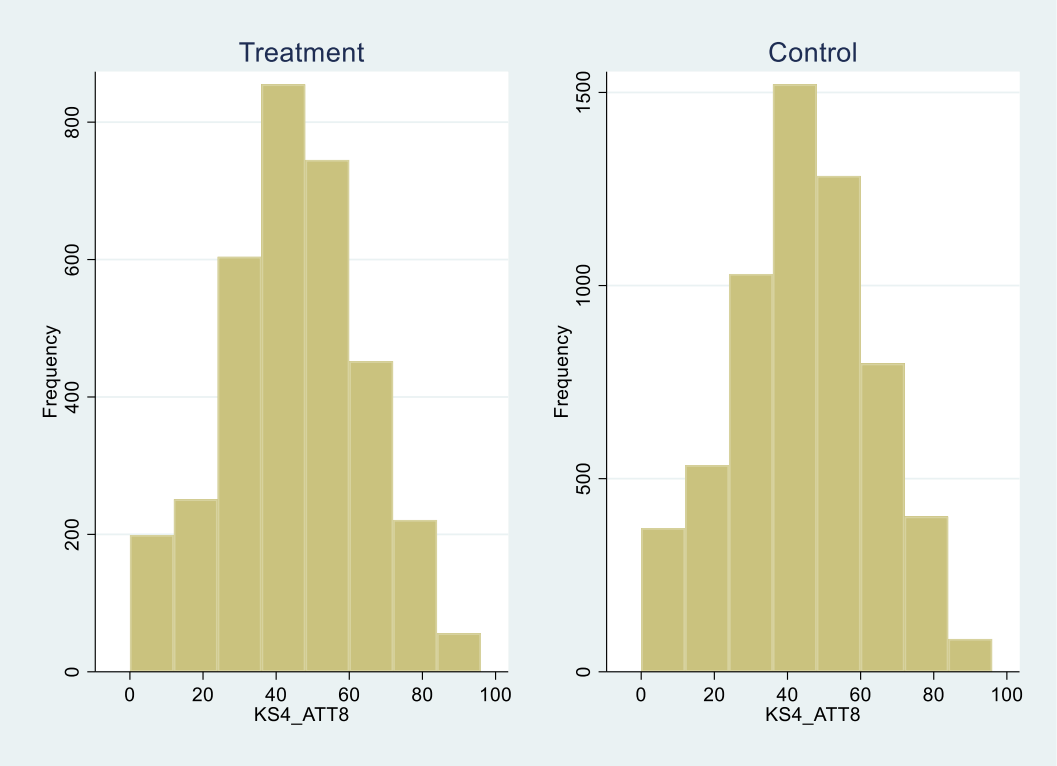
Table 11 presents the results of the primary analysis. The mean Attainment 8 score stood at 44.4 among the intervention group and at 43.8 among the control group. The analysis indicates no impact of the programme on Attainment 8, with an effect size of zero (calculated as Hedges g) and confidence intervals that cross zero. The corresponding variance estimates are provided in Appendix Table 2 (Appendix F).

Table 11: Primary analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Attainment 8	3383 (355)	44.4 (43.7, 45.0)	6027 (920)	43.8 (43.3, 44.3)	9410 (3383; 6027)	0.00 (-0.19, 0.20)	0.97

Figure 2 presents the distribution of Attainment 8 scores for intervention and control groups. This shows similar distributions across both groups.

Figure 2: Histograms, Attainment 8 (primary outcome), by trial arm



Secondary analysis

Table 12 presents the results for all secondary outcomes considered as part of the evaluation. Histograms presenting the distribution of all measures, by intervention and control group, are presented in Appendix G.

Analysis of attainment measures at end Year 7 and end Year 8 present a mixed picture. In Year 7, there is no impact on reading scores, while the effect size for maths stands at -0.11, equivalent to around two months less progress, but confidence intervals are wide, crossing zero, and results not statistically significant. The raw mean scores for reading are similar in the treatment and control groups, while for maths, we see a higher mean among the intervention group. Analysis of attainment measures at end Year 8 indicate no impact on maths, while here the effect size for reading has a negative sign, equivalent to around one month less progress, though again, in both instances confidence intervals cross zero and results are not statistically significant. Confidence intervals are larger than for the main analysis as the sample sizes are much smaller. We also observe substantial missing data on these measures, with scores missing for around 40% of the sample in both Year 7 and Year 8. Thus, it is important to exercise caution in interpreting these results.

The mean number of absence sessions was 12.3 in the intervention group and 12.6 in the control group in Year 7; the equivalent figures for Year 8 were 14.7 and 14.9 sessions respectively (where an absence session is equivalent to half a day). The results indicate that attending a school allocated to receive the programme was associated with a lower level of absence, particularly in Year 7, with an effect size of -0.09. In Year 8, the effect on absence is slightly reduced (effect size of -0.07). There is some indication that there may also be a favourable impact on reducing fixed exclusions, at least in Year 7. The mean number of fixed exclusions (total fixed period exclusions per school year) was 0.05 in the intervention group and 0.09 in the control group in Year 7; and 0.11 and 0.13 in Year 8. However, the uncertainty around these estimates means we cannot draw any conclusions regarding effectiveness.

In Appendix H, we present results from additional analyses for absence and exclusions using multilevel negative binomial models. These provide further support for the impact of the programme on reducing absence, with lower absence among the intervention group in both Year 7 and Year 8 (statistically significant at the 5% level in both year groups). As for the main models, while the results are indicative of fewer exclusions among the intervention group, these estimates are less precisely estimated with wider confidence intervals,

Table 12: Secondary analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Year 7 reading score	1035 (841)	103.90 (102.79, 105.01)	1930 (1520)	103.18 (102.36, 104.00)	2965 (1035; 1930)	-0.01 (-0.28, 0.27)	0.95
Year 7 maths score	997 (865)	97.49 (96.38, 98.59)	1741 (1756)	93.48 (92.66, 94.29)	2738 (997; 1741)	-0.11 (-0.49, 0.28)	0.61
Year 8 reading score	1232 (644)	104.88 (103.77, 106.00)	1803 (1647)	102.18 (101.29, 103.07)	3035 (1232; 1803)	-0.07 (-0.33, 0.18)	0.45
Year 8 maths score	1248 (614)	99.89 (98.89, 100.89)	1749 (1748)	96.04 (95.18, 96.89)	2997 (1248; 1749)	0.00 (-0.30, 0.31)	0.99
Year 7 absence	3495 (243)	12.25 (11.73, 12.76)	6125 (822)	12.61 (12.22, 13.00)	9620 (3495; 6125)	-0.09 (-0.18, -0.00)	0.05
Year 8 absence	3482 (256)	14.69 (14.09, 15.29)	6195 (752)	14.90 (14.45, 15.36)	9677 (3482; 6195)	-0.07 (-0.16, 0.01)	0.13
Year 7 exclusions	3508 (230)	0.05 (0.04,0.07)	6232 (715)	0.09 (0.07, 0.10)	9740 (3508; 6232)	-0.11 (-0.23, 0.00)	0.11
Year 8 exclusions	3508 (230)	0.11 (0.09, 0.13)	6232 (715)	0.13 (0.11, 0.15)	9740 (3508; 6232)	-0.04 (-0.15, 0.06)	0.52

Note: the number of missing observations presented in parentheses for reading and maths scores are best estimates as in around 5% of cases it was not possible to establish with certainty the original assigned subject; furthermore, in the two schools where pupil data was not provided, pupil numbers are estimated and it is assumed that 50% would be assigned to reading and 50% to maths.

Analysis in the presence of non-compliance

For the analysis of compliance we make use of information collected by the delivery team on the extent to which schools were considered to have fully delivered the programme to the satisfaction of the team (see Methods section). This analysis is conducted only for the primary outcome, Attainment 8.

A total of thirteen schools, or school year groups, were considered to have fully delivered the programme. This is out of the maximum of 23 school year groups that were assigned to implement the intervention.

As per the SAP, below we present the results of an IV regression, with Table 13 presenting the first stage results and Table 14 the main regression results. This does not point to evidence of an effect on Attainment 8 among compliers as we do not see a statistically significant association between the indicator of compliance and the primary outcome (as represented by the coefficient and standard error on the compliance indicator in Table 14).

Table 13: First stage, compliance analysis

Treatment	0.549 (0.134)***
Prior KS2 attainment	-0.001 (0.002)
FSM_S	-0.038 (0.050)
Single sex school	0.082 (0.176)
GCSE above/below	-0.047 (0.078)
Cohort13	0.041 (0.093)
Cohort14	0.158 (0.126)
phase one	0.074 (0.107)
N observations	9410

Note: Standard errors are in parentheses. Statistical significance is indicated as *** significant at 1%, ** significant at 5%. Model also includes constant term and dummy for missing KS2 prior attainment.
Results of F test: F(9, 9400) = 6.31, Prob > F = 0.000.

Table 14: Compliance analysis, IV results

Compliance indicator	-0.973 (2.551)
Prior KS2 attainment	2.947 (0.078)***
FSM_S	0.323 (0.656)
Single sex school	-5.599 (1.897)***
GCSE above/below	-3.311 (0.863)***
Cohort13	0.414 (1.331)
Cohort14	-0.746 (2.205)
phase one	-0.653 (1.060)
N observations	9410

Note: Standard errors are in parentheses. Statistical significance is indicated as *** significant at 1%, ** significant at 5%. Model also includes constant term and dummy for missing KS2 prior attainment

As a further check we also drop from the analysis those schools that were not considered to comply; results are presented in Table 15. This makes no substantive difference to the main results, with an effect size of 0.01.

Table 15: Additional compliance analysis, excluding non-compliers

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Excluding non-compliers	2158 (131)	44.4 (43.6, 45.2)	6027 (920)	43.8 (43.3, 44.3)	8185 (2158; 6027)	0.01 (-0.24, 0.25)	0.95

Missing data analysis

For the primary outcome measure, data is missing for around 4% of the sample where pupil data was provided by schools (as discussed earlier in the section on attrition). Pre-test data (KS2 attainment) is also missing for around 4% of this sample and for around 3% of the sample for which the primary outcome is non-missing. As all of these figures fall below the threshold of 5% specified in the SAP, we do not proceed further with investigating the characteristics associated with missingness of either the primary outcome or covariates (in this case, KS2 attainment).

While not pre-specified in the SAP, when the level of missingness in the primary outcome was below the 5% threshold, we do report in Table 16 the results of undertaking some sensitivity analyses around missing data as a further robustness check. The first row reports the results of a complete case analysis; the second row imputes a school-level mean where KS2 data is missing, and the final row drops the measure of prior attainment. None of these scenarios has any substantive impact on the main results, with an effect size of around zero in all specifications.

Table 16: Missing data analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Attainment 8, complete case analysis	3307 (431)	44.3 (43.7, 45.0)	5796 (1151)	43.9 (43.4, 44.4)	9103 (3307; 5796)	-0.01 (-0.21, 0.19)	0.86
Impute school-level mean	3383 (355)	44.4 (43.7, 45.0)	6027 (920)	43.8 (43.3, 44.3)	9410 (3383; 6027)	0.01 (-0.19, 0.20)	0.93
No prior attainment	3383 (355)	44.4 (43.7, 45.0)	6027 (920)	43.8 (43.3, 44.3)	9410 (3383; 6027)	0.00 (-0.19, 0.20)	0.98

Subgroup analyses

Free school meals

Table 17 presents results of conducting analysis for the subgroup of FSM pupils and Table 18 for non-FSM pupils. These have been conducted for both the primary outcome and all secondary outcomes. Note that we use the FSM measure EVERFSM_6 to define this group as EVERFSM_6_P is not available for all years throughout the period of this study. Note also that the trial was not designed to detect effects within the subgroups, thus these findings should be considered exploratory.

The results indicate an effect size of 0.02 for the primary outcome among FSM pupils and -0.02 among non-FSM pupils. Again, this is effectively equivalent to zero months of additional progress. We also run models including an interaction term between treatment and eligibility for FSM (Table 19); this does not show a statistically significant interaction for the primary outcome.

For most secondary outcomes, this interaction term is also not statistically significant. The exception is for both exclusions measures, which suggests there may be some differential impact of the programme among FSM pupils, reducing exclusions among this group (although the average rate of exclusions is higher for FSM than non-FSM pupils).

The subgroup findings also indicate a lower rate of exclusion among FSM pupils in both Year 7 (effect size of -0.19 for FSM pupils compared with -0.03 among non-FSM pupils) and in Year 8 (-0.13 compared with 0.03).[13]

No statistically significant interaction is observed for absence; here, if anything, the separate subgroup analyses in Tables 17 and 18 indicate slightly greater effects for non-FSM pupils than for FSM pupils; and it is among non-FSM pupils in Year 7 where we see a statistically significant impact (although sample sizes are also larger for the non-FSM group).[14]

Findings for reading and maths scores are mixed. In Year 7, the subgroup analysis indicates lower scores for both reading and maths for FSM pupils, and for maths scores only for non-FSM pupils. In Year 8, the impact on reading scores appears similar for both FSM and non-FSM pupils, while in contrast to the Year 7 results, the impact on maths is positive for FSM pupils and negative for non-FSM pupils. In all cases however, confidence intervals are wide and cross zero and estimates are not statistically significant. Interaction terms between treatment and FSM status were not statistically significant at conventional levels (the negative interaction for Year 7 maths scores is statistically significant if evaluated at the 10% level).

Table 17: Subgroup analysis—pupils' eligibility for FSM

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Attainment 8, FSM pupils	1085 (75)	37.6 (36.5, 38.8)	2056 (111)	37.0 (36.2, 37.9)	3141 (1085; 2056)	0.02 (-0.17, 0.20)	0.86
Year 7 reading score	310 (266)	98.66 (96.75, 100.57)	581 (435)	97.00 (95.53, 98.47)	891 (310; 581)	-0.08 (-0.37, 0.21)	0.42
Year 7 maths score	298 (255)	93.15 (91.29, 95.02)	568 (467)	90.42 (89.04, 91.81)	866 (298; 568)	-0.11 (-0.49, 0.27)	0.60
Year 8 reading score	399 (177)	101.32 (99.39, 103.26)	605 (411)	97.49 (95.94, 99.03)	1004 (399; 605)	-0.05 (-0.29, 0.19)	0.67
Year 8 maths score	396 (157)	95.10 (93.33, 96.87)	592 (443)	90.70 (89.36, 92.04)	988 (396, 592)	0.10 (-0.17, 0.38)	0.48
Year 7 absence*	1,160	17.0 (15.9, 18.2)	2,166	16.8 (16.0, 17.7)	3326 (1160; 2166)	-0.08 (-0.19, 0.02)	0.16
Year 8 absence*	1151	19.7 (18.4, 21.0)	2165	19.9 (18.9, 20.9)	3316 (1151, 2165)	-0.06 (-0.17, 0.05)	0.35

[13] When running multilevel negative binomial models (results in Appendix H), the interaction term between treatment and FSM status was only statistically significant for Year 7 exclusions and not for Year 8. However, the impact on exclusions in Year 7 was statistically significant in the subgroup model for pupils eligible for FSM, at the 1 per cent level.
[14] Results from multilevel negative binomial models are broadly in line with these findings (Appendix H).

Year 7 exclusions*	1,160	0.09 (0.06, 0.12)	2,167	0.20 (0.16, 0.24)	3327 (1160; 2167)	-0.19 (-0.35, -0.02)	0.07
Year 8 exclusions*	1,160	0.19 (0.14, 0.23)	2,167	0.26 (0.22, 0.30)	3327 (1160; 2167)	-0.13 (-0.28, 0.03)	0.21

* Indicates that there were either zero or fewer than ten missing observations in each relevant group.
Note that throughout this table, the number reported 'missing' (in parentheses) does not include those pupils we could not link to the NPD (both those where a match could not be made and those in the two schools that did not provide pupil data) as it is not possible to know the FSM status of these pupils. In addition, 'missing' numbers for reading and maths scores represent the minimum number of missing observations as in some cases it was not possible to establish with certainty the original assignment to reading or maths.

Table 18: Subgroup analysis—pupils not eligible for FSM

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Attainment 8, non-FSM pupils	2279 (47)	47.5 (46.8, 48.3)	3849 (70)	47.5 (46.9, 48.1)	6128 (2279; 3849)	-0.02 (-0.23, 0.19)	0.84
Year 7 reading score	716 (412)	106.30 (104.98, 107.63)	1329 (464)	106.02 (105.06, 106.97)	2045 (716; 1329)	0.05 (-0.23, 0.33)	0.59
Year 7 maths score	690 (447)	99.42 (98.07, 100.77)	1159 (637)	94.99 (93.99, 96.00)	1849 (690; 1159)	-0.08 (-0.51, 0.34)	0.71
Year 8 reading score	823 (314)	106.90 (105.56, 108.25)	1155 (638)	105.07 (103.99, 106.14)	1978 (823; 1155)	-0.04 (-0.3, 0.22)	0.72
Year 8 maths score	846 (291)	102.19 (101.01, 103.38)	1103 (693)	98.99 (97.89, 100.08)	1949 (846; 1103)	-0.04 (-0.36, 0.28)	0.81
Year 7 absence*	2,326	9.9 (9.4, 10.4)	3,918	10.4 (10.0, 10.8)	6244 (2326, 3918)	-0.11 (-0.21, -0.02)	0.02
Year 8 absence*	2313	12.3 (11.6, 12.9)	3907	12.3 (11.9, 12.7)	6220 (2313, 3907)	-0.09 (-0.18, 0.00)	0.11
Year 7 exclusions*	2326	0.04 (0.02, 0.05)	3919	0.03 (0.02, 0.04)	6245 (2326; 3919)	-0.03 (-0.12, 0.05)	0.56
Year 8 exclusions*	2326	0.07 (0.05, 0.09)	3919	0.06 (0.05, 0.07)	6245 (2326; 3919)	0.03 (-0.05, 0.12)	0.57

* Indicates that there were either zero or fewer than ten missing observations in each relevant group, or are suppressed.
Note that throughout this table, the number reported 'missing' (in parentheses) does not include those pupils we could not link to the NPD (both those where a match could not be made and those in the two schools that did not provide pupil data) as it is not possible to know the FSM status of these pupils. In addition, 'missing' numbers for reading and maths scores represent the minimum number of missing observations as in some cases it was not possible to establish with certainty the original assignment to reading or maths.

Table 19: Subgroup analysis—interacting treatment and FSM status

	Attainment 8	Year 7 reading score	Year 7 maths score	Year 8 reading score	Year 8 maths score	Year 7 absence	Year 8 absence	Year 7 exclusions	Year 8 exclusions
Treatment	-0.171 (1.397)	0.254 (1.550)	-1.364 (3.756)	-1.375 (1.931)	0.398 (2.741)	-1.485 (0.724)**	-1.095 (0.911)	-0.022 (0.037)	0.002 (0.049)
FSM	-5.949 (0.388)***	-2.003 (0.687)***	-1.075 (0.573)	-2.043 (0.770)***	-3.500 (0.665)***	3.345 (0.356)***	4.664 (0.444)***	0.138 (0.013)***	0.165 (0.018)***
Treatment*FSM	0.211 (0.644)	-0.363 (1.152)	-1.675 (0.970)	0.035 (1.203)	0.117 (1.015)	0.281 (0.583)	-0.359 (0.729)	-0.102 (0.022)***	-0.079 (0.029)***
N observations	9269	2936	2715	2982	2937	9570	9536	9572	9572

Each column shows selected coefficients from a regression of the outcome on treatment arm, eligibility for FSM, treatment*eligibility for FSM, cohort, phase, and blocking variables, plus a dummy for missing pre-test data. Standard errors are in parentheses.
Statistical significance is indicated as ***significant at 1% level, **significant at 5% level

Prior attainment

We also explore whether there is evidence of differences in impact for pupils of 'high' and 'low' prior attainment, where 'high' attainment is defined as being in the top third based on attainment at KS2 and 'low' attainment is defined as being in the bottom third. For each of our primary and secondary outcomes, we run a model interacting treatment allocation with prior attainment (see Statistical Analysis section); the results are presented in Table 20. Interactions between the intervention and high or low prior attainment were generally not statistically significant. The exceptions are a negative and statistically significant interaction between the intervention and being of low prior attainment for the Year 8 maths score and a positive interaction between the intervention and high prior attainment for the Year 7 maths score.

Table 20: Subgroup analysis—interacting treatment and prior and attainment

	Attainment 8	Year 7 reading	Year 7 maths	Year 8 reading score	Year 8 maths score	Year 7 absence	Year 8 absence	Year 7 exclusions	Year 8 exclusions
Treatment	0.264 (1.410)	0.493 (1.512)	-2.714 (3.873)	-1.063 (1.919)	1.005 (2.833)	-1.492 (0.796)	-1.885 (0.979)**	-0.052 (0.038)	-0.011 (0.049)
Low prior attainment (bottom third)	-2.209 (0.535)***	-4.014 (0.846)***	-0.751 (0.786)	-3.704 (0.982)***	-0.299 (0.889)	1.056 (0.403)***	1.702 (0.498)***	0.054 (0.015)***	0.113 (0.020)***
High prior attainment (top third)	7.116 (0.507)***	6.306 (0.812)***	5.243 (0.713)***	5.038 (0.955)***	7.134 (0.822)***	-1.514 (0.423)***	-1.729 (0.522)***	-0.037 (0.016)**	-0.050 (0.021)**
Treatment*Low prior attainment	-0.332 (0.723)	0.392 (1.259)	-1.417 (1.079)	-1.309 (1.339)	-2.325 (1.126)**	0.222 (0.676)	1.208 (0.837)	-0.016 (0.025)	-0.043 (0.033)
Treatment*High prior attainment	-0.981 (0.717)	-1.365 (1.211)	2.807 (1.006)***	0.356 (1.332)	0.457 (1.102)	0.137 (0.676)	0.797 (0.836)	0.014 (0.025)	0.007 (0.033)
N observations	9103	2884	2653	2920	2883	9375	9352	9392	9392

Each column shows selected coefficients from a regression of the outcome on treatment arm, prior attainment (reference category: average prior attainment/middle third), treatment interacted with prior attainment, cohort, phase, and blocking variables, plus a dummy for missing pre-test data. Standard errors are in parentheses.
Statistical significance is indicated as ***significant at 1% level, **significant at 5% level.

Additional analyses and robustness checks

Table 21 presents the results of our additional analyses and robustness checks (for the primary outcome only).

The first three rows present results from including additional control variables in the primary analysis specification. The first row presents results after additionally controlling for pupil age in months, gender, FSM eligibility, as well as the proportion of pupils in the school with special educational needs (SEN) and the proportion of pupils in the school with English as an additional language (EAL). These are the control variables that were included in the interim analyses. The second and third rows repeat the primary analysis with additional controls for aspects on which the analysis sample appears unbalanced: the second row presents results from adding controls for FSM eligibility and percentage of SEN pupils and the third row also controls for school type. In all three specifications, the effect size remains around zero and not statistically significant.

In the fourth and fifth rows, we present results from restricting the models to only pupils who remain in the same school, as described earlier in the Statistical Analysis methods section. The results in the fourth row exclude pupils who could be identified as having left the school in either Year 7 or before the January of Year 11. The fifth row presents results based on pupils who were identified as being at the same school in both Year 7 and Year 11, based on their school identifier. In both cases, again the effect size remains at around zero and not statistically significant.

The final two rows present the results from running separate analyses for the individual cohorts, given that we might be concerned about cohort differences over time and changes in the way that attainment is measured (there are seven cohorts within the treatment group and 25 cohorts within the control group in 2013; these figures stand at 16 and nine cohorts in 2014 respectively). This does highlight the sensitivity of the results in this respect, with a positive effect size for the 2013 cohort, and a negative effect size for the 2014 cohort, although neither estimate is statistically significant and these remain relatively small in magnitude.

Table 21: Additional analyses

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Attainment 8, additional controls[1]	3383 (355)	44.4 (43.7, 45.0)	6027 (920)	43.8 (43.3, 44.3)	9410 (3383; 6027)	0.02 (-0.17, 0.22)	0.76
Attainment 8, additional controls[2]	3383 (355)	44.4 (43.7, 45.0)	6027 (920)	43.8 (43.3, 44.3)	9410 (3383; 6027)	0.01 (-0.18, 0.21)	0.86
Attainment 8, additional controls[3]	3383 (355)	44.4 (43.7, 45.0)	6027 (920)	43.8 (43.3, 44.3)	9410 (3383; 6027)	0.02 (-0.17, 0.22)	0.76
Attainment 8, same schools[4]	3363 (375)	44.5 (43.9, 45.2)	5964 (983)	44.2 (43.7, 44.6)	9327 (3363; 5964)	0.00 (-0.20, 0.19)	0.95
Attainment 8, same schools[5]	2850 (888)	46.2 (45.6, 46.9)	5036 (1911)	45.6 (45.1, 46.2)	7886 (2850; 5036)	-0.03 (-0.27, 0.21)	0.71
Attainment 8, cohort 2013 only	940 (27)	45.82 (44.60, 47.04)	3462 (463)	43.72 (43.07, 44.37)	4402 (940, 3462)	0.03 (-0.29, 0.35)	0.79

Attainment 8, cohort 2014 only	2443 (328)	43.81 (43.05, 44.57)	906 (396)	43.19 (41.94, 44.44)	3349 (2443, 906)	-0.05 (-0.40, 0.31)	0.67

[1] All controls as in interim reports: pupil age in months, gender, FSM eligibility; % EAL pupils in school, and % SEN pupils in school. Dummies included for missingness where applicable.
[2] Controls: % SEN pupils in school, plus eligibility for FSM (and missing dummies for FSM and SEN).
[3] Controls: % SEN pupils in school, school type, plus eligibility for FSM (and missing dummies for FSM and SEN).
[4] Excluding any pupils leaving known to have left school in Year 7 or before January of Year 11.
[5] Based on pupils with same school identifier (LAESTAB) in Year 7 and Year 11.

Estimation of ICC

The overall ICCs at post-test, at school level, are reported earlier in Table 8.

The ICC at pre-test (that is, based on KS2 prior attainment) was 0.11 (based on our 69 school year groups).

Implementation and process evaluation results

This section describes the findings of the implementation and process evaluation around fidelity, implementation, and perceived outcomes. As described in the Methodology section, this section is based on two waves of qualitative research. This included interviews with Healthy Minds leads and teachers, focus groups with pupils, and training day observations conducted by the evaluation team as well as analysis of training evaluation data and teacher experience data collected by the delivery team. This section is, unless specified, based on the primary research conducted by the evaluation team, that is the interviews with PSHE leads and teachers, focus groups with pupils and training day observations. The analysis of data collected by the delivery team has been used to provide context and to cross-check the validity of the findings as well as to identify any gaps in the primary research; however, both data sources supported one another.

Implementation

In this section we look at:

- why schools decided to participate in the programme;
- how PSHE was delivered in the implementation schools before the Healthy Minds programme;
- training and preparation;
- organisational and practical aspects of implementation;
- teacher perspectives on programme design and resources;
- pupil response to the programme and the activities; and
- intentions to use Healthy Minds beyond the life of the project.

Why schools decided to participate in the programme

School leads on Healthy Minds who were interviewed gave a range of reasons for taking part in Healthy Minds, including the need for a structured programme of PSHE within the school, difficulties sourcing curriculum materials for PSHE, variability in the content and delivery of PSHE within and between year groups, the need for staff training in teaching PSHE, and a concern that the school was not meeting its duties in relation to Social, Moral, Spiritual and Cultural Education (SMSC).

Most school leads and teachers had worried that their school's provision of PSHE was inadequate. Most lacked a formal PSHE programme and were either looking for one or were receptive to change. Their priorities varied, but the ones mentioned most frequently included introducing a more complete and formal programme and changing delivery from form tutors to a small team to ensure consistent delivery. A number of schools wanted to move PSHE out of form time into the curriculum (see later) and supported the practice of delivering Healthy Minds through PSHE subject teachers rather than form tutors.

Most school leads and teachers said their school had also lacked good resources for PSHE. A number of schools had developed a package of provision around specific PSHE subjects. Typically, teachers sourced their own materials and shared these between each other on a fairly informal basis. While some good materials had been sourced, schools reported that this had resulted in wide variation in the standard and quality of PSHE teaching and learning and was an issue of concern to those delivering PSHE and senior leaders. Schools were also attracted to participate by the teaching resources and training component. As one PSHE lead explained:

> *'As much as we talk about the student side, from our perspective it was also that there was a training capacity, so we were building capacity of staff to deliver. I think that's been the icing on the cake, that actually we're building capacity of staff to deliver PHSE and high quality PSHE' (School 1, wave one evaluation).*

As well as being motivated to improve provision of PSHE, interviewed school leads had a particular interest in elements of the Healthy Minds programme, including its emphasis on developing resilience.

How PSHE was delivered before Healthy Minds

Prior to their involvement in the Healthy Minds programme, the 12 case study schools had delivered PSHE in a range of ways. In many of the schools, PSHE had been scheduled for an hour a week, but programmes were not always fully

structured and often lacked a clear timetable, topic areas, and associated materials. In some, PSHE was considerably less frequent than weekly and in others the lessons were reported to be often cancelled because of lack of priority to the subject.

A typical model before Healthy Minds involved delivery by form tutors, rotating between different subjects, or combined PSHE/RE lessons with additional 'drop down' days on specific PSHE topics such as sex and relationships, risk taking, cultural diversity, and bullying. For most of the schools, timetabled PSHE of an hour a week therefore represented a significant change.

Involvement in the Healthy Minds programme encouraged schools to move away from distributing responsibility for PSHE across numerous members of staff to a core team of two to four teachers who were responsible for delivering most aspects of the curriculum and had attended all the training sessions.[15] Interviewees spoke of how this approach allowed Healthy Minds teachers to establish trust with pupils in their class and encourage an open and honest discussion about the topics covered.

Importantly, the need for staff to be trained to deliver the Healthy Minds programme also made it unfeasible for schools to operate the common practice of assigning teachers with space in their timetable to deliver PSHE. This approach is likely in itself to have resulted in inconsistency in teaching and learning, particularly where teachers have little interest in, or are uncomfortable delivering, PSHE.

Reasons for not taking part in the programme

Schools that decided not to take part in Healthy Minds gave a range of reasons for this decision, according to data from the delivery team. Many Hertfordshire schools had already taken part in the Penn Resilience Programme training (a component of the Healthy Minds programme) and their participation could therefore contaminate the results. Other reasons included lack of support from senior leadership for a change in existing programmes and delivery models. Some schools adopt different PSHE models for Key Stages 3 and 4 and the programme was seen to disrupt this system. A few schools withdrew from the project after randomisation and during the first year, so that the number of treatment and control schools in phase one of the project was seven and five, respectively, by the time of the process evaluation visits in February and March 2015. A further round of school recruitment took place in 2014, adding a further 22 treatment and control schools to a second cohort of schools which started delivering the project from the autumn term of 2014.

Training and preparation

Each of the two cohorts of teachers in phase one and phase two schools attended an introductory conference followed by five initial days of training on the Penn Resilience Programme, which lays the foundations for subsequent training sessions and components of the programme. A further three days of training was delivered on other modules and components of the Healthy Minds programme. This amounted to a total of eight days of training in 2013/2014 or 2014/2015 and five days in 2014/2015 or 2015/2016.

With the exception of project leads within schools, teachers reported knowing little about the Healthy Minds programme before attending the initial training course covering the Penn Resilience Programme. A few teachers were knowledgeable about related approaches such as neuro-linguistic programming. Most teachers knew a little about resilience and about cognitive behavioural therapy through initial teacher training or through the media, but these had not formed part of their PSHE programmes.

Training evaluation data was collected by the Bounce Forward project team. The team reviewed the evaluation data collected for phase one schools and found high levels of satisfaction with all of the courses. The initial five-day training in the Penn Resilience Programme was highly rated, with an overall score of 4.8 out of 5 (strongly agree) achieved for the measure 'learned valuable skills that can be beneficial in my life and work'. Some of the training sessions were very highly rated, with the media influences training achieving top marks of five for providing useful resources and for the

[15] One school (School 6) had reduced the number of staff involved in delivering PSHE just prior to the beginning of the programme due to the recruitment of two citizenship teachers who were able to take on full responsibility for the subject. However, the school's involvement in Healthy Minds encouraged the continuation of this mode of delivery, even after one of the citizenship teachers left.

skills and knowledge of the trainers. The sex education training received slightly lower ratings than the other courses, but written comments from attendees indicate discomfort with some of the subject matter covered. Trainers were highly rated, particularly for their knowledge about the subject material, organisation, and preparation.

The positive feedback in evaluation by the project team was mirrored in interviews with teachers where the team found general agreement that the training was high quality and had prepared them very well to deliver the programme. Teachers reported very positive experiences of the training delivered by the Bounce Forward team. They rated the quality of the training very highly, describing it using words such as 'outstanding', 'incredible', 'amazing', and 'a fantastic experience'. It was seen as well-structured and interactive. The trainers were seen to believe in the effectiveness of the programme and its underlying ideas and to have conveyed their enthusiasm to participants. Training delivered directly by the Bounce Forward team was singled out as being of particularly high quality, in particular the initial Penn Resilience Programme. The training delivered by third parties, however, was viewed as being more variable.

Teachers also appreciated the contact made with teachers in other schools at the training courses. The only downside was the length of training and the time away from the classroom that this entailed. Some project leads said they had found it difficult to justify this to the school's senior leadership.

More generally, teachers appreciated having training in PSHE and felt that the skills and knowledge base of the programme had given PSHE a greater profile within the school both among their fellow teachers and the pupils receiving the curriculum. As a teacher in a phase one school remarked:

> *'[The training] just gives that little bit more impetus … that makes them think, okay, this is important, it's not just cobbled together wasting time … People delivered this from America, they've come along to train us, and I'm now training you. I think it just ups the ante, which is great from my experience' (Healthy Minds teacher, School 1, wave two evaluation).*

Several of the PSHE teachers interviewed in schools in wave two (schools 3, 6, and 7) disseminated some of the training they had received to other staff within the school who were not involved in delivering Healthy Minds. This was motivated by an interest in staff adopting a common language and terminology for dealing with behavioural issues within the school, which was provided in training and resources on the Penn Resilience Programme. Teachers observed that a whole-school approach would help reinforce the key messages from the programme, and enhance its potential impact on pupils.

Preparedness for delivering the programme

Teachers said they felt well-prepared by the training and many were keen to get started with the programme. Some were also a little apprehensive, particularly those who had been teaching PSHE using programmes they had developed themselves and were familiar with. A number also said they needed time to adapt the materials for use with their pupils, particularly in schools with low levels of literacy.

Training for the delivery of Healthy Minds represented a considerable commitment on the part of schools and teachers. The initial Penn Resilience course included a focus on the individual experiences and behaviours of participating teachers. As such, it was found to be emotionally demanding with one teacher noting, 'I found stuff about myself that I hadn't realised was there' (School 4, wave one evaluation). However, this was seen as a necessary process in order to understand the impact that the programme might have on pupils, as one PSHE lead explained:

> *'Learning about yourself made the next two days very much easier to think about "well how is that relevant to students, how can I teach that to students, and where might students find some of that a bit difficult or become a little uncomfortable?" So I could pre-empt that by some of the bits that I felt uncomfortable with' (School 2, wave one evaluation).*

In relation to pupil response, teachers said they felt well prepared for the emotional aspects of the programme, including pupil responses to some of the scenarios and discussions. This was seen as particularly valuable for delivery of sessions on sex and relationships and for mental health.

Training was valued for the opportunity to practice using the resources and delivering lessons. A PSHE lead explained that this differed significantly from other training they had received, which they descried as focused on theory:

> *'A lot of the training you do is theory. They get you to read the theory for yourself—you can go home and read this, we'll take you through this lesson and how you think you should deliver it' (School 1, wave one evaluation).*

Some teachers also took the opportunity to actively consider how they could adapt content to suit the needs of their students, particularly teachers in schools with low pupil attainment on entry to Year 7 who felt that some adaptation was needed. Ideas for adaptation included more practical activities and small group delivery. Teachers also valued the opportunity to share practice with counterparts in other schools, particularly as the project progressed and they became familiar with teachers in other participating schools.

As described earlier, the programme involves many components delivered over a four-year period. Data was not collected on whether teachers' confidence in delivery may have changed over time, but refresher training is made available by the Bounce Forward team.

Organisational and practical aspects of implementation

Coordinators and teachers reported relatively few practical difficulties organising the delivery of Healthy Minds in their schools. The two organisational issues were timetabling and teacher turnover. On the first of these, scheduling the programme sometimes presented some challenges to timetabling, for example, where PSHE was moved out of form time to the curriculum. After this initial set-up, some schools later had to reschedule Healthy Minds to accommodate other curriculum areas. This pressure arose particularly from Year 10 onwards when a number of schools reduced the lessons from weekly to fortnightly, usually to increase time for maths and English lessons. Teachers saw this change as less than ideal, leading to difficulties covering all topics fully and affecting pupil recall of previous lessons. Not all schools were able to deliver in form groups, with one phase one school having to change delivery to English sets some years into the project. Schools appeared to deliver the programme fully, nonetheless, sometimes by scheduling additional sessions.

The second organisational challenge for some schools was teacher turnover and the need to train new PSHE teachers in the programme. In one phase one school this resulted in delays to implementation and the delivery of the schools' former PSHE lessons for part of the year. While all schools visited had a core team of PSHE teachers who attended all the training sessions and were involved in delivering most aspects of the curriculum, some schools had to engage other teachers in the delivery of PSHE, and these were most often selected from teachers with gaps in their timetable rather than PSHE expertise. This sometimes raised concerns in some schools about the consistency and quality of delivery, where these additional teachers were not fully trained in the programme and less experienced in its delivery. Discontinuity in the PSHE teaching team was also seen as disruptive for pupils, and in particular engagement in personal and sensitive topics. Teacher turnover meant additional time for training, addressed in various ways including twilight training by the Bounce Forward project team and cascade training by teachers.

Class size also presented a challenge to participating schools with few schools visited able to deliver to 15 pupils at a time (or spilt classes of 30) as recommended by Bounce Forward. Full classes of around 30 were more common. While this was not regarded as problematic for delivery, some teachers felt that some pupils did not get the chance to contribute as much to the class discussion as they would have in a smaller group. Teachers delivering to smaller classes said that pupils were able to settle down more quickly and to be more focused on the activities. This was not, however, seen to be specific to the Healthy Minds programme but to PSHE more generally. As one teacher stated:

> *'PSHE is something that in general is better delivered in groups somewhere between 10 and 20. That is what most PSHE teachers would tell you is an optimum number. You've got enough people to do group work and role plays, but you don't have so many people to feel uncomfortable and they don't want to speak out' (School 1, wave one evaluation).*

A colleague added that it was more difficult to both identify and address distress or discomfort among individual children when class sizes are large. More generally, teachers found it important that the right learning environment is created to deliver Healthy Minds, which was one involving discussion, group based work, and creativity.

Schools were supported throughout the project by the Bounce Forward team and advised on how to deal with these and other more minor challenges to delivery. This included delivering one to one training to teachers who had not been able

to attend training days. Teachers also commented on the high quality of communication from the project team at Bounce Forward. This was the case in both waves of interviews, suggesting this communication was sustained throughout the four years of the programme. The willingness of the project team to agree small adaptations to the delivery of the programme was appreciated by teachers needing to fit it in with the curriculum and other school requirements.

Teacher perspectives on programme design and resources

Teachers viewed the programme positively, finding it well-structured with relevant and useful content. It was seen to cover most elements of PSHE. A few teachers identified gaps, for example, financial management, road safety, and personal hygiene. Some pupils in follow-up visits said they would have liked more discussion about life skills and transition to adulthood.

Teachers found the resources, including lesson plans, teacher notes, and workbooks, all useful. The teacher notes were seen as particularly valuable, either used as a guide or as a reference point. Equally teachers were very positive about the booklets for pupils which accompany each unit, which pupils could keep and refer back to. The synergy between the teacher's folder and the student booklets was also appreciated.

Teachers felt that the programme was well-balanced in its content in relation to knowledge and skills, but they also varied in the emphasis they placed on these aspects of the programme. One view was that pupils inevitably forget some of the knowledge but tend to retain the skills and that this retention process was assisted by the way in which skills are conveyed to pupils using the resources (School 5, wave one evaluation). At the same time, some teachers talked about how pupils needed regularly reminding of the skills they had learned within the initial Penn Resilience Programme in order to relate these to the subject matter of later modules.

Adaptations

In general, the materials were found to be very good. Many of the changes made by teachers were relatively small and involved making the language more familiar to pupils.

Some Healthy Minds materials were found easy to use with pupils with only limited adaptation required, for example, the media influence section. Others required additional resources to be used to construct a whole lesson. Some resources and lesson plans were seen as having scope for improvement, in particular through being more concise. Some lesson plans, for example Relationship Smarts, were found to be too lengthy for the proposed time. This created additional preparation time for teachers who had to identify the keys messages for each lesson in order to select the most useful activities.

Some schools had made small adaptations to the Penn Resilience Programme to make it more ‘British’ than American in content, for example, replacing references to basketball with football. Relationship Smarts was seen as in need of similar adaptations. Those in the ‘School to Life’ module were found to need more tailoring and additions to meet the needs of particular pupil cohorts. The EU-DAP Unplugged and SHARP units were seen as needing to include examples relating to UK drug or alcohol use and e-cigarettes. Other modules, for example Sex Ed sorted, were seen as having a high degree of cultural relevance to pupils in the U.K. However, one school had felt the need to add content on homophobia from Stonewall because pupil attitudes within the school required the issue to be addressed in greater depth. At the same time, making adaptations to a PHSE programme was viewed as both unavoidable and positive by teachers since it meant they met the needs of their own pupils, but it did require time.

Teachers in the special school (that was interviewed as part of the IPE but is not part of the impact analysis) had needed to adapt the programme in a number of ways. These included simplifying many of the messages and reducing the reliance on literacy skills. Adaptations had to take account of pupils’ low degree of social engagement, high levels of parental protection, and more limited peer pressure compared to young people in mainstream schools. Focus group discussions found pupils made limited use of social media, had little exposure to alcohol, drugs, and sex—and more generally less exposure to the world outside of school and home. At the same time, this meant that many were very interested in such topics, particularly sex and relationships. The small number with IT skills had made their own searches but these had not necessarily made them better informed. The school had sourced additional materials as visual prompts to assist discussion and understanding of these issues. Healthy Minds lessons had also focused on life after school, where PSHE materials have typically not met the needs of young people with special needs. The ‘school to life’ module was therefore found valuable by teachers, and focus groups found pupils to have good recall of discussions about their

aspirations. These were not always realistic but provided a very useful starting point for pupils to start to envisage their lives once they leave school.

It was also necessary for some mainstream schools to adapt the materials for lower-ability pupils, who were in some cases the majority of their cohort. However, maturity was also identified as a factor which meant some pupils were more able to engage with the course content. As one teacher remarked:

> *'You face the difficulty in PSHE that you don't face in a lot of other subjects: you have to differentiate by maturity and, if you're teaching maths or English or a lot of other subjects, it doesn't really matter to a certain extent, or to a lesser extent, how mature the students are. When you're talking about some of this stuff, maturity is really important … and if you know your group is particularly immature, then some of this stuff needs to be "immaturatised" so that they can understand it on their level' (Healthy Minds teacher, School 1, wave two evaluation).*

To take account of lower levels of ability and maturity, some teachers felt it necessary to reduce the amount of written work in the lessons and increase practical activities and acting out scenarios leading to discussion (Schools 3 and 5, wave one evaluation). Some teachers also found the need to increase the proportion of task-based activities (School 5, wave one evaluation). In one school the project lead had slightly increased the written content of the programme in order for there to be a clearer record of learning outcomes. This was explained to be largely for the purposes of inspection rather than for learning in itself (School 6, wave one evaluation).

Teachers said that pupils liked the illustrations and the characters that feature in the resources pack. One view among teachers was that the Year 7 programme had the feel of materials for slightly younger pupils, with some of the scenarios applying to pupils in Year 6. However, pupils did not express this view to teachers and neither did they to the NIESR evaluators during focus groups with pupils.

Pupil response to implementation and delivery

Teachers reported that pupils responded well to the programme. Levels of engagement were reported to be high but varied between topic areas with some units more popular, such as Media Navigation, and others less so, for example, School to Life. Pupils themselves described most of the Bounce Forward modules as engaging and mostly relevant to their age and current experiences. Overall, confirming teacher reports, sessions on 'Media Navigation' were received positively by pupils and, contradicting teacher reports, 'School to Life' was also received positively with the content seen as relevant and timely. Teachers reported that pupils often had a strong recall of the messages around use of social media and saw these as highly relevant to themselves and their peers. In keeping with this, teachers observed that pupils engaged far more with content that was relevant to their daily lives. As one Healthy Minds teacher stated in relation to the Media Navigator unit:

> *'Any time that you talk about social media they're all on the edge of their seats because they know more than we do as well and they're happy to share their experience'* (Healthy Minds *teacher, School 4, wave two evaluation).*

Some teachers in mainstream schools gave example of lower levels of pupil engagement with some modules, resulting in some cases from lower levels of maturity. For example, in one school, two teachers recounted how some pupils felt emotionally overwhelmed during classes on the Mental Illness Investigated unit in Year 10 and required additional support. Teachers in a few schools felt that pupils found the drugs and alcohol content repetitive. However, teachers across schools reported positive responses to these units, and to sessions on sex education.

Pupils offered mixed reviews of sessions that were more skills-based and dealt with more abstract knowledge, such as the Penn Resilience Programme and Mindfulness. Sessions on mindfulness were viewed very differently by pupils both between and within schools: some said they found the techniques useful and had applied them, while others had not put them into practice, possible due to a lack of maturity.

Pupil response to the learning activities

As well as their content, pupils' responsiveness in lessons was also influenced by the tasks and activities they were set. Teachers spoke of how the Healthy Minds curriculum had encourage them to use a greater diversity of teaching practices than they would typically use in other lessons. This included role play exercises, group work, reflective written tasks,

and drawings. The range of exercises and how they contrasted with other lessons was perceived to have had a positive impact on pupil engagement with the curriculum; it provided them with a different means of learning about particular topics and in the focus groups, pupils independently commented that the novelty of certain activities, compared to how they typically learn in school, helped them to retain certain messages or information better.

Pupils themselves said that the use of videos, case studies, and role play engaged them and reinforced the relevance on their own lives. One pupil reflected the views of others in explaining that:

> *'Someone talking about it isn't that fun. When you watch a video you can visually see what's gonna happen. Sometimes if the teacher is just talking you can miss stuff because you're not paying attention. A video just grabs your attention.'*

Video clips were reported by teachers to offer pupils a more realistic and engaging insight into the issues being discussed, enabling them to relate to the more emotional aspects of the programme. The role-play element was new to some teachers but they had found it worked well, particularly with smaller classes. In general, those schools that actively engaged with the resources by allowing for open discussion and the adjustment of scenarios received the best pupil response. The programme was seen to allow pupils to discuss personal issues in a supportive and trusting environment. Pupils were sometimes reported to be more forthcoming than teachers had expected, sometimes 'over-sharing' which teachers were able to address. In the minority of schools where the pupils described a more 'traditional' learning environment—with more independent writing, teacher direction, and limited use of additional resources—the pupils were notably less engaged and less positive about the programme.

Pupils gave further evidence that the learning environment required for the delivery of the Healthy Minds project had been created in participating schools. Pupils described how the discursive, creative, and group-based elements of the programme kept the class engaged and interested. Pupils also reported that PSHE teachers had created a positive learning environment during the programme where pupils felt confident to participate. Pupils interviewed in wave one and two school visits said they especially enjoyed class discussions. Both teachers and pupils commented on the degree of trust and confidentiality established in Healthy Minds lessons.

One view expressed by pupils was that having the sessions in a familiar peer group and with a consistent teacher helped pupils engage and feel confident to share. One pupil reflected the views of others in stating that:

> *'There's a lot of respect in the lessons for other people's different opinions on things and different stories that people tell, like [name of classmate] said, there's always something to relate to because with different situations, so I definitely think, yeah, we all do listen as a group, we all work together like class discussion, everyone gets involved, everyone listens to each other' (Year 8, wave one).*

Similarly, a pupil in another school compared PSHE to other curriculum subjects:

> *'It's more interactive, it's more enjoyable. Say in maths you just sit there and do the work, in this you interact more with the teacher' (Year 8, wave one).*

Teachers said that pupil behaviour was better than in other curriculum subjects, although this varied between schools. One teacher remarked on the good behaviour of pupils within Healthy Minds lessons, explaining this with reference to pupil engagement in content and delivery (School 5, wave one evaluation). Pupils felt somewhat differently, reporting behaviour management issues which disrupted the Healthy Minds lessons. One view was that the lack of formal assessment in the programme enabled pupils to opt out of constructive participation. However, pupils said that this problem was not confined to PSHE lessons. At the same time, the majority view was that their fellow pupils did actively engage with the lessons and that disruption was sporadic, rather than continuous.

More generally, there was a perception that pupils' engagement in PSHE declined somewhat as they got closer to exam periods. At the same time, pupils commented that in Years 9 and 10 it was very helpful to have a lesson where they are free from the pressure of assignments or assessments.

Although some pupils recognised the importance of building upon previously acquired knowledge and saw value in the iterative nature of the programme, some felt the sessions were at times repetitive. However, it also meant that pupils

were clear about some of the main messages of the programme, for example, making good decisions and not jumping to conclusions. One pupil said of the different modules:

> *'They deal with different things. I wouldn't say that it's the same message over and over again. I'd say that you might get a glimpse of the message in another category, another module' (Year 9, wave two).*

The extent to which the programme was seen as repetitive by pupils also seemed highly dependent on teacher delivery and additional initiatives going on in the schools. For example, pupils in one wave one school described a heavy school-focus on smoking, which negatively impacted upon their opinion of those sessions in the programme. Pupils felt that the emphasis had been too strongly on abstention rather than choice. This module had also raised concerns among some pupils in the special school about their parents' smoking, because one of the key messages for some pupils was that you were basically a bad person if you smoked.

Future plans

An indicator of successful implementation is whether schools were committed to continuing delivering Healthy Minds after the end of the project. Many Healthy Minds teachers were keen to continue the programme since it provided them with materials which could be used 'off the shelf' and which had been found to work well with pupils. They felt there was no similar programme which they could use in its place. Teachers were very positive about the evidence-based content of the curriculum, their preparatory training, and its intended delivery as weekly one-hour sessions. It was felt that this combination of factors had helped to raise the profile of PSHE as a subject and gave it a greater degree of esteem among both pupils and teachers. As such, all interviewees wanted to maintain this model of delivering PSHE and keep teaching units from the Healthy Minds curriculum after the trial ended. However, some schools expressed concerns about whether this was sustainable in light of other teaching priorities and pressures on resources. This particularly applied to Year 10 when PSHE faced being squeezed out of the timetable by GCSE subjects. In some cases, teachers thought they might make adaptations, taking some of the elements of the programme as a whole, reducing some elements, and mixing the materials with their own sessions (School 3, wave one evaluation). In order to continue with the programme beyond the life of the project, some schools would need to train new additional teachers, involving additional costs. This was seen as a decision for schools' senior leadership.

Fidelity

In this section we look at evidence of whether participating schools delivered the intervention as intended, looking in particular at:

- whether schools delivered the programme to plan;
- schools' compliance with administrative requirements;
- adaptations to the programme and any implications for fidelity; and
- additional programmes and activity.

Fidelity in programme delivery

Most of the schools that we visited delivered the weekly Healthy Minds lessons with only minor tweaks, as described in the intervention description earlier. However, some schools dropped out. One school lost its Healthy Minds teacher in the Easter of the first year of implementation and delivered Healthy Minds only partially, mixing it with its existing programme. The school then withdrew from the trial.

The case study schools that kept with the programme largely delivered it according to plan although some had taken more time to deliver some units. For example, one school had found the 'breathe', mindfulness, unit challenging to deliver with its cohort of lower attaining pupils with significant emotional needs and had allocated more time. Some schools delivered a reduced PSHE timetable in Year 10 because of pressure on the timetable from GCSE subjects.

Compliance with the administrative aspects of the project was good. The lead teachers in schools said they did not find it difficult to provide the required information to the project team, such as feedback on the course modules. Teachers found the module questionnaires easy and straightforward to complete. For some, this was also a useful exercise, allowing for reflection on delivery:

'It made me think about how the students responded ... sometimes, on some of the lessons I thought they might be a bit dry and I didn't understand how the students would actually understand and get involved and they have flabbergasted me because they have got so involved with it ... being able to think back to that and feedback that has been really nice' (School 2, wave one evaluation).

Adaptations which could affect fidelity

The main adaptations to the content of the curriculum reported by interviewees were shortening the length of some of the lesson plans because of time constraints caused by pressure from other curriculum subjects. Teachers also made some adjustments to make content more culturally relevant to British students. Some schools had shortened the length of units so that they could fit in additional content not covered by the Healthy Minds curriculum such as citizenship and financial education.

As described above, most of the adaptations made were minor and involved merely making the language and scenarios more 'authentic' to the particular group of pupils. The structure, messages, and learning points remained unchanged. Teachers understood that they were involved in a trial. They were aware of the need for fidelity within the programme and understood instructions from the project team that they could 'tweak' the materials but not alter the content of the programme. This requirement appears to have been followed by treatment schools among the case studies in the IPE so that the adaptations would be more accurately described as 'tailoring'.

Some specific changes were made to the delivery of Sex Ed Sorted. These included delivering the first part of the unit in Year 9 instead of Year 8 on the grounds of age-appropriateness.

A more substantial adaptation was the exclusion of homework activities by some of the schools visited which operated a policy of not setting homework from PSHE. The homework activities within the Healthy Minds programme are intended to allow students to practice what they have learnt on their own as a way to reinforce learning, and also to prepare. Some teachers felt that the homework component of the course was quite substantial and that not setting it at all resulted in delays to the scheduled programme. Teachers in schools that enforced the homework policy strictly found that a small number of activities could not be included, for example, those involving discussion with parents. However, solutions were generally found and these included small exercises to prepare for future sessions. In other cases, pupils were simply asked to complete work they had started in class. Overall, there was variation in practice in relation to homework but this is unlikely to have affected fidelity.

Additional programmes

The evaluation sought to establish whether treatment schools were delivering other initiatives to the trial cohort that focused on improving pupil wellbeing that may also have contributed to any impact observed in this area. None were mentioned in the research interviews. However, one school (School 8, wave two evaluation) stated that since it started delivering the programme, it now provided all pupils with 'mentor time' outside of lessons where PSHE topics, pre-selected on a termly basis, were discussed as a group. Interviewees commented that they wanted to provide pupils with additional PSHE content outside the formal lessons as they recognised the value doing all they could to support young people's non-academic development from their involvement in the programme. The school concerned also observed that they also have an excellent student support network of support officers and counsellors.

A separate school (School 10, wave two evaluation) was also teaching pupils about Growth Mindset approaches to their intelligence in order to help them cope with exam stress and the demands of the curriculum.

Perceived outcomes

In this section we look at:

- teachers' views on the general impact on pupils of the programme;
- pupils' views on the general impact and their understanding of the programme's messages;
- teachers' and pupils' views of the impact of programme on academic performance; and

- the transfer of Healthy Minds messages beyond PSHE.

General impact—teachers' views

We asked teachers about the impact of the programme on pupils and whether it had led to changes in their behaviour, including academic work. During the first wave of visits, many teachers felt it was too early in the life of the programme to be able to identify impact on pupils, particularly in relation to academic performance. They also tended to place greater value on any emotional or behavioural changes which Healthy Minds might bring about. One teacher spoke about how the programme could positively impact on pupils' ways of thinking:

> *'It's thinking in a way that they won't be taught anywhere else. It's thinking about thinking, which is something these children don't do because they've got this bull in a china shop approach to life' (School 1, wave one evaluation).*

Some felt that it might be possible to see the effects on attitudes and behaviour at a later stage of the project. This view was expressed at both the first wave of school visits and the follow-up ones. Another view was that pupils might not fully use the skills acquired through the programme until much later in their lives (School 5, wave one evaluation). Teachers felt, nonetheless, that the programme was making an impact on pupils' thinking and this was demonstrated in classroom discussions during Healthy Minds lessons. Pupils were reported to have shown that they understood the concepts, ways of thinking, and approaches of Healthy Minds. The most important messages were seen as around issues of (not) making assumptions, using evidence to reach conclusions, coping with setbacks, and not giving up. (School 5, wave one evaluation). At the time of the follow-up visits, teachers felt that these themes were still retained, and that the importance of resisting peer pressure was also an enduring message.

In terms of impacts on behaviour, teachers gave some examples of where pupils had used Healthy Minds techniques in school, outside of PSHE lessons. For example, teachers in three schools during the second wave of school visits in 2017 (Schools 4, 8, and 11) observed or had heard of pupils using the breathing exercises and mindfulness techniques in stressful situations such as completing assessments or homework. While the unit on breathing exercises was completed early in the programme's delivery (Year 7), several schools at wave two mentioned that they continued to hold regular mindfulness meditation sessions with pupils—for instance, in form time or assemblies—to rehearse and reinforce these techniques.

Some teachers also reported examples of where pupils said they had used Healthy Minds approaches when facing challenging situations at home. Particular mention was made of relaxation techniques and walking away from conflict. In one school, pupils had told the Healthy Minds teacher that they had used the messages and techniques to resist peer pressure to drink alcohol. In another example, pupils had told their teacher they had used its messages to be more resilient when participating in sport.

Some teachers reported an improvement in pupil behaviour in the school, which they believed could be explained, at least in part, by the Healthy Minds programme. One school had found a reduction in aggressive behaviours in Years 7 and 8, which the project lead believed was partly a result of the Healthy Minds programme, in particular the reduction in repeat offences (School 5, wave one evaluation). The incorporation of Healthy Minds language, strategies, and thinking processes into behaviour management was seen as responsible for this change (School 5, wave one evaluation).

The most widely observed behavioural impact reported by PSHE teachers—owing to its clear visibility in the classroom—was increased confidence and maturity in discussing sensitive topics and a higher degree of emotional literacy. This was seen to stem from the work of teachers to cultivate a safe environment for pupils over a number of years, starting with confidentiality contracts drawn up with pupil input and establishing a trusting classroom environment. This had also enabled teachers to talk about sensitive issues more openly. As one teacher explained:

> *'It's based on trust, you know, like, I just felt like, oh my God, I wouldn't be able to talk about sex education if I had not developed their trust to talk about it without embarrassment' (* Healthy Minds *teacher, School 9, wave two evaluation).*

Some PSHE teachers (for example, in School 3) felt that it was possible to attribute these gains in maturity to pupils' participation in the Healthy Minds programme rather than their natural development over this period.

General impacts—pupils' views

From focus groups with pupils it was possible to identify varying levels of impact. Media Navigation and School to Life appeared to have most impact on pupil's knowledge, attitudes, and behaviour. Pupils reported changing their use of social media and had developed some ideas on how they might approach their future within school and beyond. Evidence of impact from the Penn Resilience Programme and Mindfulness was less common, although some pupils said they had used the approaches explored in these units.

The greatest difference in impact was found not between elements of the programmes but between schools. Pupils' recollection of the programme varied substantially, based on findings from focus group visits conducted as part of the wave one evaluation, with pupils in three of the schools demonstrating a good or strong recollection of the programme and what they had learned. In two of the schools pupils had an incomplete recollection of the programme's elements, activities, and messages, however, these pupils were also less confident and articulate than in the other schools so it is possible that they were just less able to express what they had learned. In the other two schools, pupils' level of recall varied, as did their reports of how it had impacted on their lives in and out of school.

We asked pupils to design a leaflet for new Year 7s explaining the Healthy Minds lessons and what to expect. As explained above, pupils' initial recall of lesson content was clearer in some schools than others. In designing leaflets, pupils were able to remember what each of the main modules covered and some of the key messages. Those which were recorded in their leaflets most frequently were: Being Safe Online, Thinking Before You Act, Keeping Calm, Planning Your Future, Making Important Choices, and dealing with the 'Risks of Life'. The issues of bullying and substance misuse featured strongly in the pupil leaflets. Other comments related to the opportunity for class discussion, for work without tests, and to be yourself. One leaflet designed by pupils expresses some of these points well:

'PSHE: the real you

- *Is more a personal lesson where you learn about your safety and awareness*
- *This lesson can make you learn about drugs, alcohol, photoshop and more*
- *In this lesson there's more active group lesson where you can cooperate*
- *This is a lesson where you can experience in any way you want to*
- *It's a more calm lesson where focus is on personality instead of grades and levels*
- *A time to get creative and personalise what you do - show people the real you*
- *It's a lesson that's a lot more social compared to English, Maths and Science*
- *You can truly express yourself.'*
  (School 5, wave one evaluation).

Pupils talked about how they had changed their behaviour as a result of their Healthy Minds lessons. Typical examples of behavioural change were not jumping to conclusions, walking away from an argument rather than prolonging it, and remaining calm when witnessing family disputes. Many pupils said they had added privacy settings to their online profiles or were more careful about using their mobile phones in public. Pupils also talked about being able to make their own decisions on matters such as smoking and alcohol based on the knowledge they were given within the Healthy Minds programme. There was evidence that it had provided them with the ability to apply reasoning and to take responsibility for their decisions and subsequent actions. While these messages were more difficult for pupils in the special school to comprehend, a number were able to give examples of how they had used Healthy Minds messages in their everyday lives. For example, when pupils in the focus group were asked if they had used self-talk, pupils gave examples of difficult situations where they had:

> *'I've used that when my coat got locked inside and it started raining. I went back upstairs and found someone to open the door for me.'*
>
> *'When I forget my reading books and I leave them at school.'*

In a minority of schools, some pupils said they had not experienced this benefit of the programme. The extent to which pupils, across schools visited as part of the IPE, reported being able to reflect on their own behaviour and change their response seemed to be a consequence of teacher delivery of the programme: in schools where pupils described a more collaborative, relaxed, and in some ways less formal learning environment, pupils enjoyed being accountable for their

opinions and being able to take others into account. Where this style of delivery was practised, pupils appeared to be given more opportunity to evaluate the strength of advice and evidence from sources such as teachers and the media.

In the focus groups (for example, in School 8), pupils commented that in these conditions they had learnt not to feel embarrassed about discussing sensitive topics such as sex and relationships with their peers. Teachers reflected that they had made similar gains themselves.

Academic impacts—the views of teachers and pupils

Teachers were generally of the view that Healthy Minds was more likely to have an impact on wellbeing than on pupils' academic performance. However, they felt that, indirectly, the programme might then impact on pupils' academic work either because wellbeing facilitates learning or, more specifically, improved relationships with other pupils can have the same effect.

Some teachers felt that increased personal resilience was the mechanism through which the programme could have an impact on performance. It was felt that the programme could equip pupils with different strategies to persevere, apply themselves to their studies, and to recover from setbacks, either in their personal or academic lives (for example, School 4, wave one evaluation). The programme was also seen as potentially effective in improving pupil strategies for dealing with stress, particularly at key pressure points such as GCSE exam time, leading to less wasted time and a stronger focus on effective revision and learning. There was also a view that the programme could help pupils make better choices, for example, in GCSE subjects (School 3, wave one evaluation).

There was also a view that it would take a number of years for the programme to become embedded in schools and to bring about a cultural change which could be reflected in academic results. It was also believed that, for the programme to be effective in bringing about this change, Healthy Minds messages would have to be reinforced in other areas of teaching and learning and in school life. Despite the examples given above, some teachers were doubtful that this was happening in their school, and that pupils 'compartmentalised' their Healthy Minds learning to PSHE.

Very few pupils described ways in which Healthy Minds could impact upon their academic attainment. Some pupils drew links between subjects covered in PSHE—for example, the chemicals in tobacco smoke—and science lessons.

Transfer of Healthy Minds messages beyond PSHE

Many teachers believed that Healthy Minds could impact on pupil behaviour to the extent of achieving measurable impacts in wellbeing and academic results if it is reinforced in other curriculum subjects and in whole-school activities and messaging. Some schools were attempting to encourage pupils to actively transfer the knowledge and techniques they had picked up from PSHE to other subjects. A number of Healthy Minds teachers interviewed also taught PE, and a few respondents (for example, in School 4, wave one) stated that as a result they had started to focus more on wellbeing as part of their subject, reinforce the key messages from certain units (for example, around resilience), and were using the mindfulness meditation techniques with pupils before sporting events to good effect. Teachers in a number of the intervention schools visited as part of the IPE also identified scope for more reinforcement of Healthy Minds messages in curriculum subjects because the programme has a good ‘fit’ with the ethos of the school, for example, its emphasis on ‘student voice’ and resilience. Schools with a strong pastoral support structure also felt that the programme complemented this ethos. However, even where this was the case, whole-school adoption required coordination, planning, and training to be put in place. In one school this was hampered by the presence of other initiatives with slightly different messages and language to Healthy Minds and a lack of training for other teachers in the school (School 1, wave one evaluation).

As noted earlier, a number of Healthy Minds teachers interviewed in wave two schools disseminated some of the training they had received to other teachers in the school who were not involved in Healthy Minds. While it is hard to judge, it should be noted that this risks contamination between intervention and control groups within the same school. It was motivated by an interest in adopting a common language within schools around behaviour, but also to reinforce messages and the programme’s impact on pupils. Pupils also indicated the importance of consistency and reinforcement by others members of staff. It was clear from focus groups that pupils were inclined to place more value on issues which

clearly applied either to their school lives in general, for example bullying, or life outside of school, for example drugs and alcohol.

Based on the findings from the schools that were visited as part of the IPE, whether the programme was perceived as a whole-school initiative and whether parents were involved in a meaningful way, therefore, seems key to whether the programme can have a real impact. Creative outputs from the programme (for example, the school to life map), started in the classroom and taken home, were identified as the main gateway through which pupils engaged their parents. Where pupils had discussed issues which arose from Healthy Minds with their parents, their engagement seems to have been valued. For this reason, the decision of some schools not to set PSHE homework may have weakened the programme's impact. It was apparent that there were substantial differences in levels of pupil to parent engagement both between schools and between pupils in schools, with pupils from apparently more advantaged social backgrounds more likely to report discussing PSHE issues with parents and siblings.

Control group activity

Healthy Minds is a structured and systematic programme covering discrete areas of teaching and learning. It is therefore unlikely that control group schools would be delivering a similar programme, though we do not have any evidence of this. It is also known that schools vary in the structure and content of their PSHE programmes. This was certainly true of the treatment schools prior to the programme where most did not have a fully structured PSHE programme consisting of a clear timetable, topic areas, and associated materials (see earlier). Therefore, Healthy Minds represented a significant change in approach for them. This may not have been true of control schools, making it important to find out about their PSHE programmes. Information about the organisation and delivery of PSHE in control schools was gathered by the Bounce Forward project team. This consisted of a brief paragraph, provided by the school lead, about their current PSHE provision. Only very high-level summary information was collected for each school, so it is important not to overstate and put too much emphasis on the analysis of this data. However, on a broad level, it is apparent that the control groups varied in their organisation and delivery of PSHE. Some control groups delivered PSHE fairly regularly, usually for one hour a week or every fortnight. This was usually delivered by form tutors. Some schools said the provision was not structured, but dealt with on an ad hoc basis, and dealt with by teachers if something came up. In other schools it was integrated into the curriculum rather than taught separately: in one case the integrated subjects were RE, citizenship, history, and geography. Although we were not able to examine the content of the control schools' programmes, the descriptions provided by Bounce Forward suggest that none were delivering a similar programme to Healthy Minds.

Conclusion

Table 22: Key conclusions

Key conclusions
There is no evidence that Healthy Minds had an impact on pupils' outcomes as measured by Attainment 8 scores, on average. This result has a low to moderate security rating.
Pupils receiving the Healthy Minds programme did not show any additional progress on reading and maths scores at the end of Year 7 or Year 8. These results should be interpreted with caution considering the challenges in administering the tests and related high levels of attrition. There was some evidence of lower levels of absence, especially in Year 7 pupils, and some indication of fewer exclusions among FSM pupils in schools that received the Healthy Minds programme compared with similar pupils in schools that did not.
There was no evidence to suggest that the programme had a differential impact on the Attainment 8 scores of pupils eligible for free school meals, or according to their prior levels of attainment.
Interviews with participating teachers indicated that the Healthy Minds programme was generally well-received by practitioners. Teachers reported relatively few practical difficulties in delivery with the main issues related to timetabling (particularly in Year 10) and teacher turnover (with new teachers needing to be trained).
Many of the teachers interviewed felt that in order to maximise the impact of the programme, it needed to be reinforced in other areas of teaching and learning and in whole-school improvement activities and messaging.

Impact evaluation and IPE integration

Evidence to support the logic model

The Healthy Minds programme aims at improving health-related outcomes. It uses the principles of cognitive behavioural therapy and theory from positive psychology to help students improve their resilience and understand the link between thoughts, feelings, and behaviour so they understand themselves better and have more empathy for others. The programme comprises a set of 14 modules bringing together individual elements for which previous research has provided evidence of effectiveness (Coleman et al., 2011). Evaluation of evidence on effectiveness of these components was typically focused on non-cognitive outcomes, given the aim of the programme.

Lordan and McGuire (2019) show that the programme had positive impacts on health-related outcomes in line with the intended aims of the programme. On average, pupils in schools receiving the programme had higher self-assessed general health than pupils in schools not receiving the programme (an effect of 0.25 standard deviations after four years, as measured by the Child Health Questionnaire-CF87). Positive effects were also found for many secondary outcomes capturing physical health and externalising behaviour (as measured by the subscales of CHQ-CF87, including for physical difficulties, general behaviour, general health, and family activities). Effects on measures of internalised emotions (emotional difficulties, self-esteem, and mental health) were close to zero and positive after four years of the programme, but in some cases were negative when measured after two years of the programme.

The current evaluation of academic outcomes shows no impact on academic attainment as measured by Attainment 8; thus it can be argued that the programme improves health outcomes without adversely affecting academic achievement, despite potentially diverting time away from traditional lessons. It does not suggest, however, that there was any positive impact on academic attainment at GCSE, at least, not as captured by the measure used within this evaluation. It also does not show an impact on reading and maths scores at end Year 7 or end Year 8. It does show some favourable impacts on rates of absence, and there are indications of lower rates of exclusions among pupils eligible for FSM. The implementation and process evaluation also provides support for the impact of the programme on pupils' wellbeing. We discuss these findings in further detail below.

Interpretation

The results of this evaluation do not indicate an impact of the programme on academic attainment as measured by Attainment 8 scores at the end of Key Stage 4. There were no indications that there were differential impacts for pupils eligible for free school meals or according to pupils' prior academic achievement (as measured by their attainment at the end of Key Stage 2). The evaluation also explored reading and maths scores from assessments administered in Years 7 and 8; however, the administration of these assessments faced significant challenges, resulting in significant

attrition for these outcomes, meaning any results need to be interpreted with caution. Results for reading and maths scores were not statistically significant.

These findings are in accordance with how teachers perceived the potential outcomes from the programme. In terms of impact on school performance and academic outcomes, teachers felt it was too soon in the lives of both the project and the pupils for this to be measurable. Some felt that this might be possible in a few years' time—but possibly not until later. At the same time, teachers believed the programme was impacting on pupils' thinking, and that they had developed skills of reasoning and enquiry. Teachers felt that these skills, combined with increased personal resilience, were the mechanisms through which the programme could improve pupil performance. The programme was also seen as potentially effective in improving pupil strategies for dealing with stress, particularly at key pressure points such as exam time or when choosing GCSE options. Alongside these potential mechanisms for improving attainment, it is also worth acknowledging that in some of the participating schools timetabled PSHE lessons had represented a substantial change to previous PSHE provision (that is, an increase in time allocated to PSHE), and where lessons replaced academic ones (although the extent of this is not known), we may have anticipated a negative impact on attainment outcomes (if this reduced the time available for traditional/academic lessons). It is possible that these impacts could have offset each other resulting in the average effect of no impact on attainment that is observed. Furthermore, not all treatment schools delivered the four-year programme in full; there were 13 schools (or school year groups) which delivered this in full and to the satisfaction of the delivery team (out of the 23 school year groups assigned to implement the intervention). This may also have reduced the likelihood of observing a positive impact on attainment. Findings from the IPE indicate that most schools that had remained with the programme delivered it as intended, that is, with generally minor adaptations (discussed further below), although some had needed more time to deliver some units, and some had needed to deliver a reduced PSHE timetable in Year 10 because of timetable pressures from GCSE subjects.

The current evaluation does indicate lower rates of absence, especially among Year 7 pupils, and among FSM pupils, there are some indications of lower rates of exclusions in Years 7 and 8. These findings provide further support for the effects of the programme on non-cognitive outcomes and are consistent, for example, with some of the measures considered in Lordan and McGuire, for example, the reduction in significant school avoidance. Teachers in some schools reported improved behaviour in their Healthy Minds cohort, which they believed might be an effect of the programme. Pupils described how they sometimes did things differently as a result of what they had learned through the programme. Typical examples of behavioural change were not jumping to conclusions, walking away from an argument rather than prolonging it, and remaining calm when witnessing family disputes. Many pupils said they had added privacy settings to their online profiles or were more careful about using their mobile phones in public. More broadly, it was apparent through focus group discussions that they understood the importance of using evidence, applying reasoning, and of taking responsibility for their decisions and actions.

Findings from the IPE (acknowledging that this only covers a sample of treatment schools) indicate that in terms of implementation, schools were very clear on the programme's requirements, including scheduling of training and delivery of the project modules. Training was very well received by teachers, with both the internal and external evaluation finding it highly rated. Teachers valued the opportunity to practise using the resources and had the opportunity to think about how they might be tailored for use with their own pupils.

A number of schools had needed to make substantial changes to how they organised PSHE to take part in the programme. One motivation for taking part was that schools welcomed the opportunity, which they felt the programme offered, to meet their Spiritual Moral Social and Cultural (SMSC) Education requirements. The previous arrangements for PSHE in some of the schools appeared to be unsatisfactory, with materials and lesson plans often put together by individual teachers. In some schools there were staffing challenges and not all delivered the full four-year programme in its intended form; but a compliance analysis did not suggest any greater impact on Attainment 8 scores among schools that complied in full.

As noted in the introduction, the quality of PSHE teaching in schools has previously raised concern. Teachers and senior leaders interviewed by NIESR welcomed Healthy Minds as a complete programme and package of materials covering key PSHE issues. They reported that such a comprehensive 'off the shelf' package is hard to find. They valued what they saw as the key messages of the programme around making good decisions, healthy choices, and building resilience. Pupils also saw the programme as relevant to them. Some teachers had adapted the materials, principally to meet the needs of lower attaining pupils, and sometimes to reflect a more British context. The adaptations were relatively

minor and could be described more as tweaks or tailoring. Teachers also added their own resources, for example, video clips where they felt these were needed, and had enjoyed making these small additions.

Teachers enjoyed delivering the programme. They rated its content, resources, activities, and messages highly. They reported that pupils responded well to the programme, in particular the opportunity for discussion. Pupils described how the discursive, creative, and group-based elements of the programme kept the class engaged and interested. Many pupils said they welcomed the opportunity within the programme for discussion. Having a lesson where all contributions are valued and not graded was also welcomed by pupils. Both teachers and pupils said the programme allowed personal issues to be discussed in a supportive and trusting environment.

Many teachers believed that, for the programme to have an impact, it needs to be reinforced in other areas of teaching and learning and in whole-school activities and messaging. In most schools this had either not taken place or was in the very early stages at the time the IPE was conducted. Consistency and reinforcement by other members of staff and by family was also identified as important by pupils: they were more inclined to place value on issues which applied either to their school lives in general or life outside school, for example bullying and drugs. Teachers felt that the programme had a good 'fit' with the ethos of their schools and that the programme could be reinforced in other subjects and in whole school activities.

This study adds to the existing evidence base by providing a UK.-based study of the impact of providing an integrated, standardised PSHE course on academic outcomes in secondary schools. The programme has previously been shown to impact positively on health-related outcomes. The current analysis does not indicate a positive impact on academic attainment, at least as captured by the measures used in this study, but taken together, the findings of the two evaluations indicate that the programme improves some health-related outcomes without adversely affecting attainment at end Key Stage 4.

As noted, previous reviews of social and emotional learning programmes have shown positive effects on academic attainment and so it is worth reflecting on why this may not be observed in this case. The review by Corcoran et al. (2018) found positive impacts on reading and maths and smaller, although still positive, impacts on science. They also find, however, that effects, while still positive, are typically smaller in larger scale, randomised studies (such as is the case for this trial) and argue that more evidence is needed in this respect. It is also worth noting that Corcoran et al. focus on outcomes in terms of particular subjects (typically reading and maths) while this evaluation focuses on a measure that combines attainment across a range of subjects. The programmes evaluated also take many different forms and operate in different contexts; while on average positive effects are found, there is inevitably variation across individual studies. Further, in our evaluation, there is a full school year between the intervention ending and the point at which the primary outcome of GCSE attainment is assessed; the time between intervention and testing in the studies included within Corcoran et al.'s meta-analysis is not reported, although the authors acknowledge that a potential limitation of their analysis is that it does not include this period as a potential moderator. Both Corcoran et al., and the earlier review by Durlak et al. (2019), note the smaller number of studies that cover the secondary school phase of education.
The Healthy Minds programme comprises modules that, as far as feasible, were evidence-based, but given the aim of the programme, such evidence was focused on non-cognitive outcomes (and even then, Coleman et al. note that the availability and quality of evidence is variable). While we do not find impacts on academic attainment, we do find some further support for the impact of the programme on non-cognitive outcomes. Furthermore, it is important to bear in mind the limitations of the analysis in assessing impact, which may have affected the ability to capture an effect; these are discussed in more detail in the next section.

Limitations and lessons learned

In interpreting the results on academic attainment, it is important to recognise limitations in the analysis.

The design of the study, where schools can contribute to both treatment and control groups with different year groups, results in some potential challenges. First, with treatment and control groups in the same school, it is possible that such a design risks some contamination across year groups, which may reduce the ability of the evaluation to detect an effect. It is unclear as to what extent this may have happened in practice as data was not collected regarding this, but the fact that some teachers commented on the need for the programme to be reinforced in a whole-school approach tentatively

suggests there may have been potential for this to occur. Second, as the primary outcome measure is collected across three different academic years, it is important to acknowledge that while Attainment 8 scores exist in all three years, as discussed earlier in the report, there are some underlying changes to the measure. This may matter because the cohorts are not balanced across treatment and control groups; the treatment group is formed predominantly of the 2014 cohort (more than 70% of pupils) with the remainder coming from the 2013 cohort. In contrast, around half of the control group are from the 2013 cohort, around 15% from the 2014 cohort and the remainder from the 2012 cohort. It is not clear *a priori* the direction of any bias that this may introduce, and we include an indicator for cohort within our regression models, but it is still important to acknowledge this as a potential limitation. We also ran separate models for each of the 2013 and 2014 cohorts; while effect sizes from both models were small and not statistically significant, we do observe a positive sign on the treatment coefficient in 2013 and a negative sign in 2014 indicating the potential sensitivity of the results in this respect.

There are also indications of some imbalance across treatment and control groups, and average levels of prior attainment (KS2 average point scores) were slightly higher in the treatment group (although as prior attainment is included in the model, this should be accounted for).

The use of outcome measures from administrative data has reduced attrition from the evaluation for these outcomes, however, it should also be acknowledged that Attainment 8 had not been the original intended outcome at the time the study commenced. Nevertheless, no impact is observed for reading and maths scores at end Year 7 or end Year 8. As noted elsewhere, the administration of the maths and reading scores faced considerable challenges and there was significant attrition in this respect (of around 40% in the case of Year 7 scores, for example).

In terms of generalisability of the results, the recruitment targeted schools with particular characteristics, and in particular those with lower levels of attainment or above average deprivation. In practice, average attainment among the final sample of schools analysed is broadly similar to the national average, although the average percentage of pupils eligible for FSM is slightly above the national average. There was a smaller proportion of schools rated ‘good’ and ‘outstanding’ by Ofsted within the analysis sample compared with the picture nationally. This means that we should be cautious in saying that the same findings would apply for schools in general; results may well differ for schools with different characteristics. While the sample includes a mix of schools with differing characteristics, sample sizes do not allow us to delve into greater detail around any potential differential effects for different types of schools.

As described throughout this report, the evaluation faced a number of challenges in implementation.

Recruitment and retention of schools proved challenging, in part, this reflects the difficulties in requiring schools to commit to a project (and evaluation) with a relatively long timeframe. The project was expertly managed by the project team at Bounce Forward but, as noted above, not all schools allocated to the treatment group were able to deliver the programme to its full extent. One broader lesson for evaluation of interventions that are delivered over a relatively long period is the inevitable higher probability of difficulties in keeping schools on board throughout, where, as seen in this study, there can be turnover in key school staff or changes in school priorities.

The requirements of the evaluation around testing placed pressure on schools with some finding the administration of tests problematic. Although there had been measures to try and reduce the burden, such as randomly assigning pupils to complete the reading and maths tests such that pupils did not do both assessments, this process was still difficult for schools. For any evaluation, careful consideration needs to be given to the choice of outcome measures, both conceptually but also in terms of practical implementation. Much has been learnt in this regard across numerous EEF trials since the point at which this study was first designed. Wherever possible, making use of data that is already collected (for example, through linkage to the NPD) is valuable in terms of reducing burden on schools. However, it is not entirely without its own limitations as it can also mean that it takes longer for such data to become available for analysis affecting the timeliness of results.

As discussed earlier, the data collected from treatment schools as part of the process evaluation only represent the views and experiences of a subset of the larger treatment population. While visited schools were selected to include a variety of delivery contexts, the qualitative findings are not necessarily representative but provide an insight into the range and diversity of views and experiences in the treatment population.

Future research and publications

In original plans for the trial it had been intended to explore whether any impact of the programme on academic attainment was mediated by any impact on wellbeing. In practice this could not be explored as part of this trial, as detailed earlier in this report, due to not having the necessary data-sharing permissions. While the findings have not shown an impact on academic attainment in the current trial, it would still be of interest to consider relationships between the impact of the programme on wellbeing and academic outcomes in future research.

Given findings around absence and exclusions in the earlier years of secondary school, it would also be of interest to explore whether such relationships persist into the latter years of secondary school. It would also be of considerable value to understand longer-term outcomes of the programme, beyond compulsory education, for example, whether it may influence progression to further and higher education and, ultimately, how individuals fare in the labour market.

It is the intention of the project and evaluation teams to seek to publish these findings. No further analysis by the evaluation team is anticipated at this stage.

References

Coleman, J., Hale, D. and Layard, R. (2011) 'A Model for the Delivery of Evidence-Based PSHE (Personal Wellbeing) in Secondary Schools', Centre for Economic Performance DP 1071, London: LSE.

Corcoran, R. P., Cheung, A. C. K., Kim, E. and Xie, C. (2018) 'Effective Universal School-Based Social and Emotional Learning Programs for Improving Academic Achievement: A Systematic Review and Meta-Analysis of 50 Years of Research, *Educational Research Review*, 25, pp. 56–72. https://doi.org/10.1016/j.edurev.2017.12.001

DfE (2016) 'Progress 8: How Progress 8 and Attainment 8 Measures Are Calculated', Department for Education: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/561021/Progress_8_and_Attainment_8_how_measures_are_calculated.pdf

DfE (2019) 'A Guide to Absence Statistics', Department for Education: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/787314/Guide_to_absence_statistics_21032019.pdf

DfE (2020) 'Secondary Accountability Measures: Guide for Maintained Secondary Schools, Academies and Free Schools', Department for Education: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/872997/Secondary_accountability_measures_guidance_February_2020_3.pdf

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D. and Schellinger, K. B. (2011) 'The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions', *Child Development*, 82 (1), pp. 405–432.

EEF (2018) 'Statistical Analysis Guidance for EEF Evaluations' London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

Greaves, E., Sianesi, B., Sibieta, L., Amin-Smith, N., Callanan, M. and Hudson, R. (2017) 'Achieve Together: EEF Evaluation Report and Executive Summary, London: Education Endowment Foundation. https://d2tic4wvo1iusb.cloudfront.net/documents/projects/Achieve_Together_Evaluation_Report.pdf?v=1630925981

Hedges, L. V. (2007) 'Effect Sizes in Cluster-Randomized Designs', *Journal of Educational and Behavioral Statistics*, 32 (4), pp. 341–370. **https://doi.org/10.3102/1076998606298043**

Heckman, J. J., Stixrud, J., and Urzua, S. (2006) 'The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior', *Journal of Labor Economics*, 24(3), pp. 411-482.

Heckman, J. J., Pinto, R. and Savelyev, P. A. (2013) 'Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes', *American Economic Review*, 103 (6), pp. 1–35.

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D. et al. (2014) 'Better Reporting of Interventions: Template for Intervention Description and Replication (TIDieR) Checklist and Guide', *BMJ*, 348, g1687. DOI: 10.1136/bmj.g1687

Lordan, G. and McGuire, A. (2019) 'Healthy Minds: Health Outcomes, Evaluation Report and Executive Summary', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Healthy_Minds.pdf

Ofsted (2013) 'Not Yet Good Enough: Personal, Social, Health and Economic Education in Schools': https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/413178/Not_yet_good_enough_personal__social__health_and_economic_education_in_schools.pdf

Torgerson, D. J. and Torgerson, C. J. (2008*) Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction*, London: Palgrave Macmillan.

Vallis, D., Singh, A., Uwimpuhwe, G., Higgins, S., Xiao, Z., De Troyer, E. and Kasim, A. (2021) "EEFANALYTICS: Stata module for Evaluating Educational Interventions using Randomised Controlled Trial Designs," Statistical Software Components S458904, Boston College Department of Economics.

Appendix A: EEF cost rating

Figure 2: Cost Rating

Cost rating	Description
£ £ £ £ £	*Very low:* less than £80 per pupil per year.
£ £ £ £ £	*Low:* up to about £200 per pupil per year.
£ £ £ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix B: Security classification of trial findings

OUTCOME: ***Attainment 8***

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%	4		Adjustment for threats to internal validity **[minus 2 padlocks]**		
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%					
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					2 padlocks
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	High	Imbalance > 0.1 SD in pre-test (KS2 average point score), in favour of intervention group. Note that this is controlled for in the analysis.
Threat 2: Concurrent Interventions	Low	IPE reports some minor examples.
Threat 3: Experimental effects and contamination	Moderate	Contamination of internal control groups is flagged as a risk in the discussion section but the evaluation is unable to assess the extent of this and/or account for it in analyses.
Threat 4: Implementation fidelity and compliance	Moderate	CACE analysis reported; AND, compliance was moderate (13/23 schools were judged as compliers).
Threat 5: Missing Data	Moderate	Total missing data is moderate; AND, analyses accounting for missing data are similar to complete-cases analyses.
Threat 6: Measurement of Outcomes	Low	Outcome test justified in relation to reliability, validity, utility and acceptability with target population; AND, administered and marked blindly to allocation; AND, no ceiling/floor effects found.
Threat 7: Selective reporting	Low	Study not registered as it was set up prior to it being mandatory to register the protocol. Despite changes to the design over the duration of the evaluation, these were all clearly reported and publicly documented. Analyses were pre-specified.

- **Initial padlock score:** 4 Padlocks due to MDES >0.2, Attrition >10%
- **Reason for adjustment for threats to validity**: minus 2 padlocks – three moderate and one high risk rating
- **Final padlock score:** initial score adjusted for threats to validity = 2 Padlocks

Appendix C: Recruitment materials

Sample parent letter, treatment group

School address
telephone number

Tel:
Fax:
Mob:
Email:
My ref:
Date:

Dear Parent

We are writing to inform you that [name of school] have agreed to take part in a research project, called Developing Healthy Minds in Teenagers along with 34 other secondary schools across the UK. The project will run for 5 years with lessons being taught each week starting during September 2013, until the end of year 10. This letter is to inform you about the project and research and give you the option to withdraw your child from the research if you would like to.

The schools involved will be teaching a specific curriculum, one hour per week that will cover the statutory requirement to teach Personal Social Health Education (PSHE). The PSHE curriculum includes lessons on building resilience, learning to use the internet and technology safely, understanding the importance of healthy relationships as well as drug and sex education. Your child will be learning skills during the lessons that will be relevant to them now and transferable beyond school.

Your child will be asked to take part in classroom based tests that will measure their academic levels and their health and emotional wellbeing. The first time will be at the start of term in September 2013 and then each June for the duration of the project.

The academic test

Your child will be randomly assigned to take either a math or english test lasting for no more than a normal lesson. Students are used to taking tests in lessons so this won't feel unusual or different for them. The results will only be used for the research and will not influence your child's place in school, teaching or academic career. Individual results will not be available to the school or anyone else outside the research team.

The test responses will be automatically and securely sent to the National Institute of Economic and Social Research (NIESR), where they will be stored securely and confidentially, in accordance with the Data Protection Act. The data will be used in the evaluation by NIESR and will be matched to pupils' records in the National Pupil Database. The data will be stored without pupils' names.

The health questionnaire

Each student will also be asked to complete a questionnaire taking no more than a normal lesson. The questionnaire will ask health related questions. Your child will be given the option to opt out of completing the questionnaire on the day. This will not impact them in any way other than not to be involved in the research. They will still take part in the PSHE lessons. In order that you and your child are aware of what the questionnaires involve and what your rights are under the Data Protection Act, please see the privacy notice below.

The research will not identify individual students or schools. At the end of the project, in 2018, an anonymised version of the dataset will be made available on the UK Data Archive. This version will contain no characteristics which could be used to identify individual students or schools.

As with any research your child does not have to participate in any of the testing if they do not want to, and this will be made clear to them before they are asked to complete the tests. Parents are also entitled to refuse permission for their children to participate. If you do not want your child to participate, please fill in and return the slip below.

If you would like to find out more about this project, or if you have any questions please contact (lead teacher insert name and contact details).

Yours sincerely

Head or Principal

If you DO NOT wish your child to participate in the research testing, return this form to your child's class teacher.

Child's name: ……

Child's class Teacher: ……..

☐ **I DO NOT wish my child to participate in the research**

Parent name (BLOCK CAPITALS) ……

Parent signature: ……

Date ……

Privacy notice Developing Healthy Minds in Teenagers Project

What is the purpose of the research?
The Personal Social Health Education (PSHE) curriculum that has been designed for the Developing Healthy Minds in Teenagers project is going to be taught in 34 secondary schools in the UK. The aim of the curriculum is to teach evidenced based lessons, taught by teachers who have been given specific training and have the expertise in the subjects. The PSHE curriculum includes lessons on building resilience, learning to use the internet and technology safely, understanding the importance of healthy relationships as well as drug and sex education. Your child will be learning skills during the lessons that will be relevant to them now and transferable beyond school. Other research studies focusing on individual parts of the four year curriculum, suggested that certain skills can help students avoid stress and anxiety, and allow them to make more informed decisions around risk behaviour such as drinking and smoking.

The London School of Economics (LSE) and National Institute of Economic and Social Research (NIESR) will be analysing the data collected through the project to assess the impact and effectiveness of the programme.

What personal data will be collected?
Each participating student will take an academic test (either maths or English) and complete a health questionnaire. They will be asked to do this before they are taught the PSHE curriculum and at the end of each year (in July) for the four years of the project. The health questionnaire will measure general health and wellbeing such as how they feel, their anxiety and mood levels and how they usually respond to particular situations. Access to a copy of the

questionnaire can be made available to view in school if you wish. The academic test will measure their levels in maths and literacy. Students will be randomly assigned to take either a maths or an English test.
These are similar to standard testing used in schools.

How will the personal data be processed?
The National Institute of Economic and Social Research (NIESR) will be processing the academic data and the London School of Economics (LSE) will be processing the health data. Only the LSE and NIESR will have access to the data collected. The LSE or NIESR will not store any data against a student name. They will analyse the anonymised answers given by your child. The analysis will form the basis of a report on the overall effectiveness of the PSHE curriculum. No-one at your child's school will see the answers given by your child and the LSE or NIESR will not store any answers given by your child against their name.

There are 2 situations when it might be necessary for someone from the LSE to discuss the data collected from the health questionnaire only: (1) if a request is received from your child to have access to the information held about them; (2) if the information collected raises concern about your child's safety or if your child's response indicates that they are very distressed. Where this is the case the LSE will contact the designated staff member at your child's school so that you child can receive the appropriate support needed. The school will not see your child's specific answer, they will just know that a concern has been raised.

Will the personal data be disclosed to anyone else?
The LSE and NIESR will have access to the personal data for the purpose of conducting the evaluation of the impact of the PSHE curriculum. The academic data will not be disclosed to anyone else. Personal data from the health questionnaire will not be disclosed to any other parties, except (a) the subject of the data, upon the child's request; or (b) to inform the designated staff member at the school/setting if a child is believed to be unsafe or very distressed, as described above.

Opting out of the research
A full explanation of this will be given to your child before they begin taking part in the academic test or completing the health questionnaire. The opportunity to opt out of the research will be explicit from the onset. Any child can request they are not involved in the research. This would not exclude them from participating in the PSHE lessons.

Sample parent letter, control group

School address
telephone number

Tel:
Fax:
Mob:
Email:
My ref:
Date:

Dear Parent

We are writing to inform you that [name of school] have agreed to take part in a research project, called Developing Healthy Minds in Teenagers along with 34 other secondary schools across the UK. The project will run for 5 years. This letter aims to explain the research and gives you the option to withdraw your child from the research if you would like to.

Some of the schools involved will be teaching a specific Personal Social Health Education (PSHE) curriculum, one hour per week to their students that start in September 2013 through to year 10. In order to measure the impact of the new curriculum, the researchers need to compare the students being taught the curriculum to students already in the school or in similar schools who are not being taught the curriculum. This is known in research as a control group.

Your child is part of the control group and will be asked to take part in classroom based tests that will measure their academic levels and their health and emotional wellbeing. The first time will be at the start of term in September 2013 and then each June for the duration of the project.

The academic test

Your child will be randomly assigned to take either a maths or English test lasting for no more than a normal lesson. Students are used to taking tests in lessons so this won't feel unusual or different for them. The results will only be used for the research and will not influence your child's place in school, teaching or academic career. Individual results will not be available to the school or anyone else outside the research team.

The test responses will be automatically and securely sent to the National Institute of Economic and Social Research (NIESR), where they will be stored securely and confidentially, in accordance with the Data Protection Act. The data will be used in the evaluation by NIESR and will be matched to pupils' records in the National Pupil Database. The data will be stored without pupils' names.

The health questionnaire

Each student will also be asked to complete a questionnaire taking no more than a normal lesson. The questionnaire will ask health related questions. Your child will be given the option to opt out of completing the questionnaire on the day. This will not impact them in any way other than not to be involved in the research. In order that you and your child are aware of what the questionnaires involve and what your rights are under the Data Protection Act, please see the privacy notice below.

The research will not identify individual students or schools. At the end of the project in 2018 an anonymised version of the dataset will be made available on the UK Data Archive. This version will contain no characteristics which could be used to identify individual students or schools.

As with any research your child does not have to participate in any of the testing if they do not want to, and this will be made clear each time before they are asked to complete the tests. Parents are also entitled to refuse permission for their children to participate. If you do not want your child to participate, please fill in and return the slip below.

If you would like to find out more about this project, or if you have any questions please contact (lead teacher insert name and contact details).

Yours sincerely

Head or Principal

If you DO NOT wish your child to participate in the research testing, return this form to your child's class teacher.

Child's name: ……

Child's class Teacher: ……..

☐ **I DO NOT wish my child to participate in the research**

Parent name (BLOCK CAPITALS) ……

Parent signature: ……

Date ……

Privacy notice Developing Healthy Minds in Teenagers Project

What is the purpose of the research?
The Personal Social Health Education (PSHE) curriculum that has been designed for the Developing Healthy Minds in Teenagers project is going to be taught in 34 secondary schools in the UK. The aim of the curriculum is to teach evidenced based lessons by teachers who have been given specific training and have the expertise in the subjects covered in PSHE. The PSHE curriculum includes lessons on building resilience, learning to use the internet and technology safely, understanding the importance of healthy relationships as well as drug and sex education. Other studies into the individual parts of the four year curriculum suggested they can help students avoid stress and anxiety, make more informed decisions around risk behaviour such as drinking and smoking. The research team suggest that the curriculum will show change in behaviour and academic performance and this can only be decided when data is compared with students who are similar, but are not being taught the curriculum. This is referred to in research as a control group.

The London School of Economics (LSE) and National Institute of Economic and Social Research (NIESR) will be analysing the data collected through the project to assess the impact and effectiveness of the programme.

What personal data will be collected?
Each participating student will take an academic test (either maths or English) and complete a health questionnaire. They will be asked to do this in September 2013 and at the end of each year (in July) for the four years of the project. The health questionnaire will measure general health and wellbeing such as how they feel their anxiety and mood levels and how they usually respond in particular situations. Access to a copy of the questionnaire can be made available to view in school if you wish. The academic test will measure their levels in maths and literacy skills. Students will be randomly assigned to take either a maths or an English test. These are similar to standard testing used in schools.

How will the personal data be processed?
The National Institute of Economic and Social Research (NIESR) will be processing the academic data and the London School of Economics (LSE) will be processing the health data. Only the LSE and NIESR will have access to the data collected. The LSE or NIESR will not store any data against a student name. They will analysis the anonymised answers given by your child. The analysis will form the basis of a report on the overall effectiveness of the PSHE curriculum. No-one at your child's school will see the answers given by your child and the LSE or NIESR will not store any answers given by your child against their name.

There are 2 situations when it might be necessary for someone from the LSE to discuss the data collected from the health questionnaire only: (1) if a request is received from your child to have access to the information held about them; (2) if the information collected raises concern about your child's safety or if your child's response indicates that they are very distressed. Where this is the case the LSE will contact the designated staff member at your child's school so that you child can receive the appropriate support needed. The school will not see your child's specific answer just know that a concern has been raised.

Will the personal data be disclosed to anyone else?
The LSE and NIESR will have access to the personal data for the purpose of conducting the evaluation of the impact of the PSHE curriculum. The academic data will not be disclosed to anyone else. Personal data from the health questionnaire will not be disclosed to any other parties, except (a) the subject of the data, upon the child's request; or (b) to inform the designated staff member at the school/setting if a child is believed to be unsafe or very distressed, as described above.

Opting out of the research
A full explanation of this will be given to your child before they begin taking part in the academic test or completing the health questionnaire. The opportunity to opt out of the research will be explicit from the onset. Any child can request not to be involved in the research.

Appendix D: Updated project privacy notice (academic outcomes)

Evaluation of Healthy Minds (academic outcomes): Project Privacy Notice

Why are we collecting this data?
The National Institute of Economic and Social Research (NIESR) are collecting personal data to enable the evaluation of the Healthy Minds programme. The Healthy Minds programme aims to improve pupils' wellbeing and health-related outcomes. The intervention comprises a Personal, Social and Health Education (PSHE) curriculum for Year 7 to Year 10 pupils.

NIESR is evaluating whether the programme has an impact on academic attainment. Through a randomised controlled trial, the evaluation is examining whether pupils who have received the programme show different academic outcomes to those who have not.

Who is this research project sponsored and funded by?
NIESR is undertaking the independent evaluation of academic outcomes, funded by the Education Endowment Foundation (EEF).

The other organisations involved in the Healthy Minds project are:

- Bounce Forward (formerly How to Thrive) - a national charity specialising in practical resilience training and curricular for schools – is the overall lead on the project and on delivering the programme in schools
- A team of academics from LSE are working with Bounce Forward. This team led and conducted the analysis and evaluation of health outcomes and are advising on research design
- The project as a whole is overseen by a steering committee, chaired by Lord Richard Layard, LSE.

What is the legal basis for processing activities?
The legal basis for processing personal data is covered by:

GDPR Article 6 (1) (f) which states that 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interest are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'.

Our legitimate interest for processing the personal data collected in this study is to conduct the evaluation. Our condition for processing special category personal data (gender) is that this is necessary for scientific research purposes and is in the public interest. In particular, these data will allow us to account for potential differences by gender in the analysis.

How will personal data be obtained?
Personal data is collected from information obtained directly from participating schools, and through linkage to the National Pupil Database (NPD). These will be augmented with assessment data from tests administered as part of the project.

What personal data is being collected by this project?
Personal data for the evaluation will include data about pupils and staff in participating schools as described below:

Names and contact details of staff in participating schools were collected in order to facilitate the evaluation (for example, to arrange the visits and interviews that form part of the evaluation).

Participating schools provided the evaluation team with data on pupils, whilst the researcher was on school site. These data include pupil names, date of birth, Unique Pupil Number (UPN) and school name.

NIESR will access and link this pupil data to background and school data held on the National Pupil Database (NPD). The NPD data to be requested will include: whether or not the pupil is eligible for Free School Meals (FSM), gender,

month and year of birth, leaving date (if applicable), and school identifiers, as well as information on attainment, absence and exclusions.

All of the above pupil data will also be matched with data from assessments undertaken as part of the project.

These data will enable the evaluation team to undertake the research and analysis required for the evaluation.

Who will the personal data be shared with?
Personal data will be only be used by the evaluation teams (NIESR and LSE).

At the end of the project, NIESR will submit data to the EEF data archive, which is now managed by the Fischer Family Trust. Anonymised data will also be stored in the UK Data Archive.

Is personal data being transferred outside of the European Economic Area (EEA)?
No personal data is stored or transferred outside of the EEA.

How long will personal data be retained?
NIESR will delete any personal data six months following the completion of the project.

Can I stop my personal data being used?
The evaluation team handles your personal data in accordance with the rights given to individuals under data protection legislation. If at any time you wish to withdraw your data, please contact dataprotectionofficer@niesr.ac.uk

In certain circumstances, data subjects have the right to restrict or object to processing, please contact NIESR's Data Protection Officer (dataprotectionofficer@niesr.ac.uk) . They also have the right to see information held about them. The evaluation team will cooperate fully when a subject access request (SAR) is made.

Who can I contact about this project?
Please contact Lucy Stokes (l.stokes@niesr.ac.uk) for any queries relating to this project.

In certain circumstances, data subjects have the right to restrict or to object to data processing. Please contact NIESR's Data Protection Officer in these circumstances. Individuals also have the right to see information held about them. You can make a subject access request by contacting NIESR.
If you have a concern about the way this project processes personal data, we request that you raise your concern with us in the first instance (see the details above). Alternatively, you can contact the Information Commissioner's Office, the body responsible for enforcing data protection legislation in the UK, at https://ico.org.uk/concerns/.

Appendix E: Histograms, prior attainment measures, by trial arm

Figure E.1: Histograms: KS2 average point score with fine grading, randomisation sample

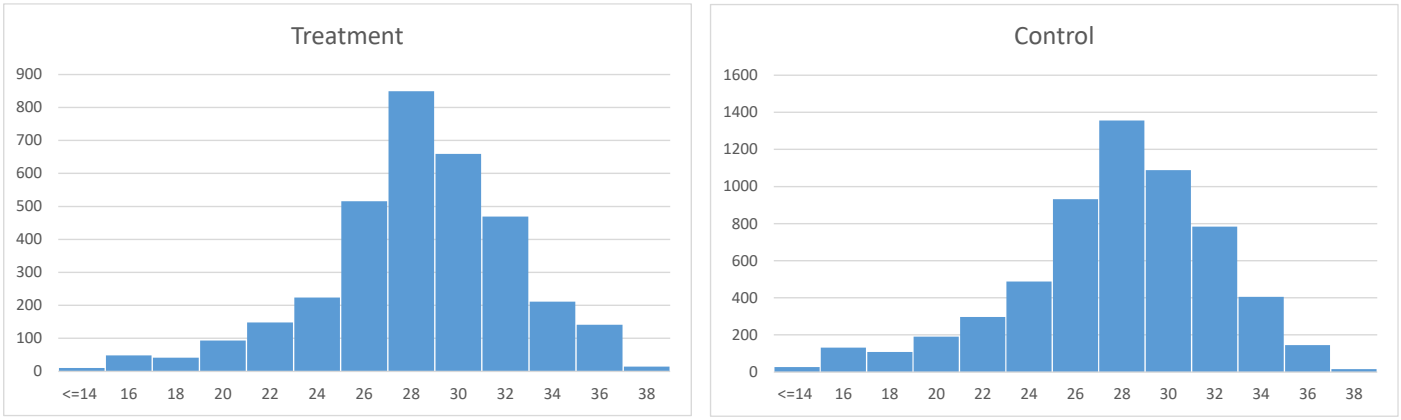


Figure E.2 Histograms: KS2 average point score with fine grading, analysis sample

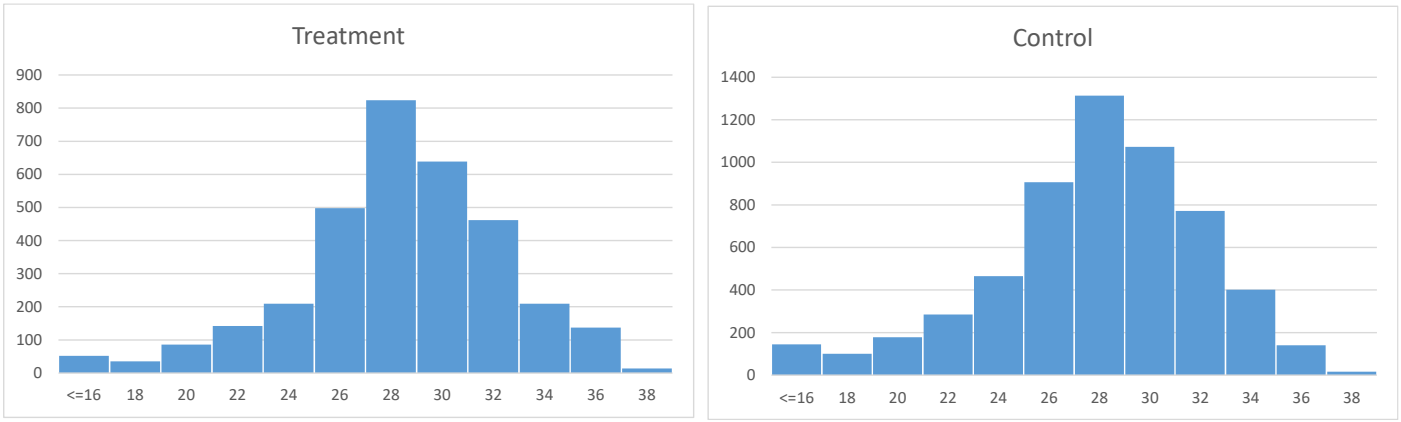


Figure E.3 Histograms: Absence, randomisation sample

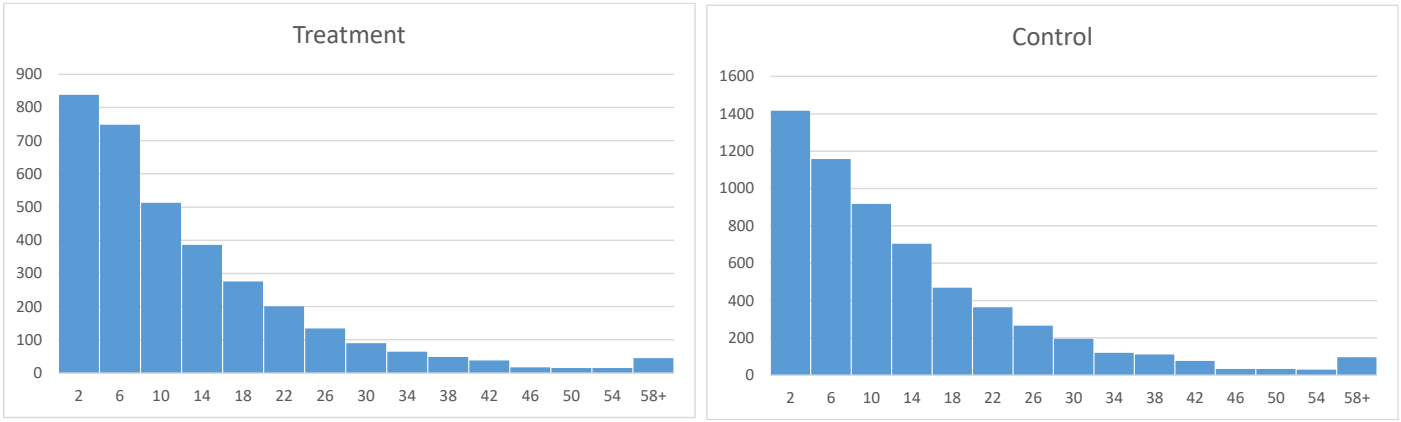


Figure E.4 Histograms: Exclusions, analysis sample

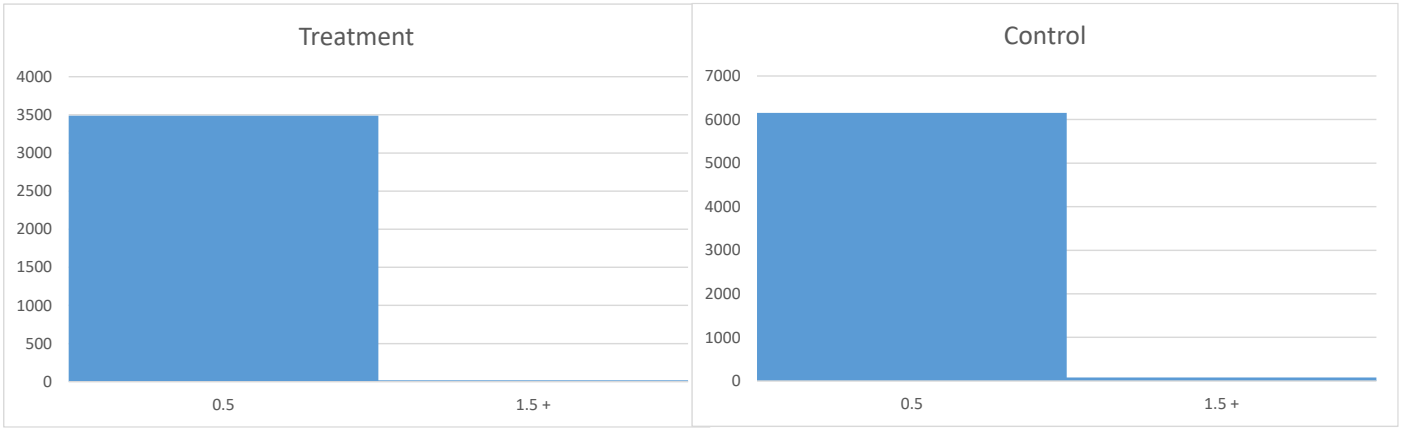


Figure E.5 Histograms: KS2 reading mark, Year 8 analysis sample

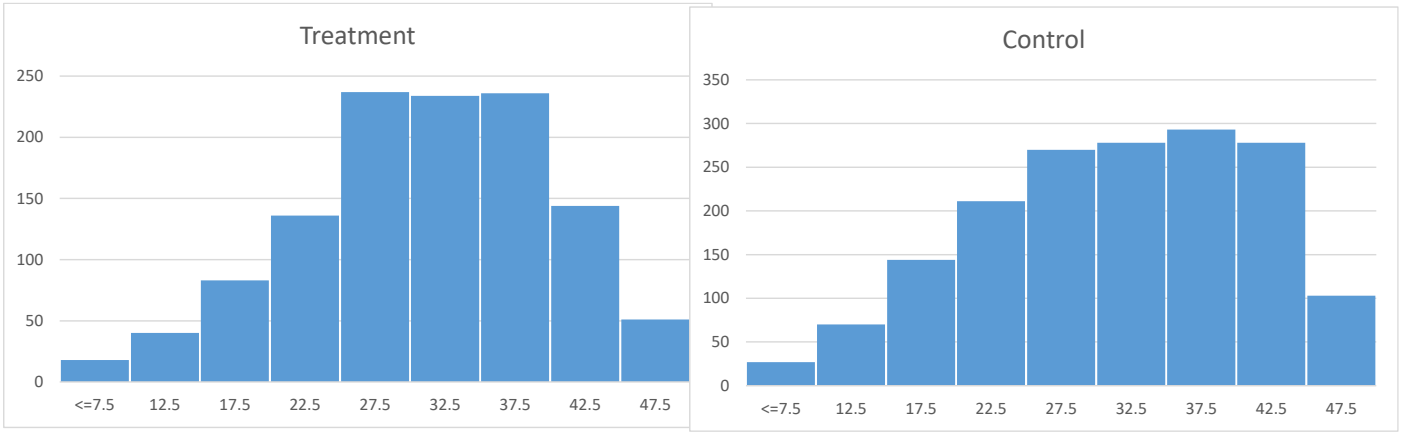
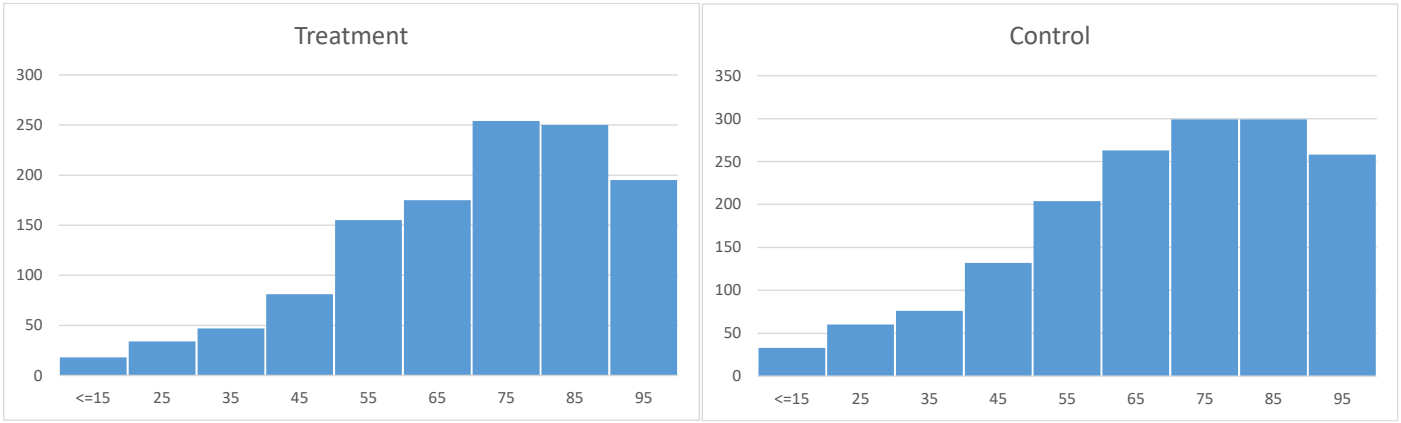


Figure E.6 Histograms: KS2 maths mark, Year 8 analysis sample



Appendix F: Effect size estimation

Appendix table 2: Effect size estimation

			Intervention group		Control group			
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance	Population variance (if applicable)
Attainment 8	0.536	0.058	3383	365.7	6027	375.0	371.7	-
Year 7 reading score	0.71	-0.098	1035	330.9	1930	336.9	334.8	-
Year 7 maths score	4.01	-1.903	997	315.7	1741	302.7	307.5	-
Year 8 reading score	2.71	-1.45	1232	397.2	1803	371.6	382.0	-
Year 8 maths score	3.85	0.45	1248	322.9	1749	334.3	329.5	-
Year 7 absence	-0.36	-1.41	3495	240.9	6125	244.5	243.2	-
Year 8 absence	-0.22	-1.33	3482	328.5	6195	334.1	332.1	-
Year 7 exclusions	-0.03	-0.06	3508	0.14	6232	0.32	0.25	-
Year 8 exclusions	-0.02	-0.03	3508	0.42	6232	0.47	0.45	-

Appendix G: Histograms, secondary outcomes, by trial arm

Figure G.1 Year 7 absence, by trial arm

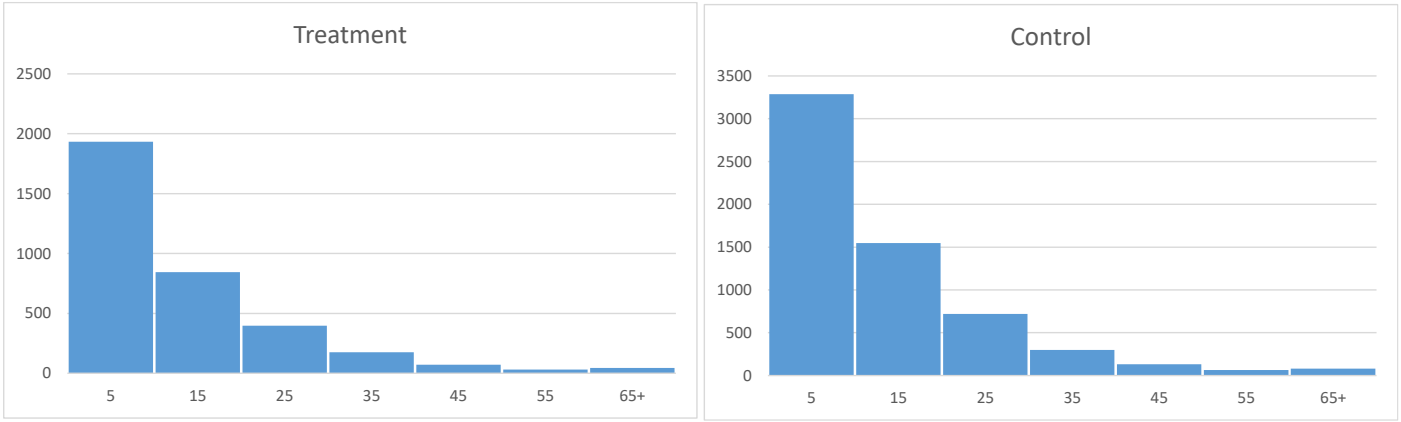


Figure G.2 Year 8 absence, by trial arm

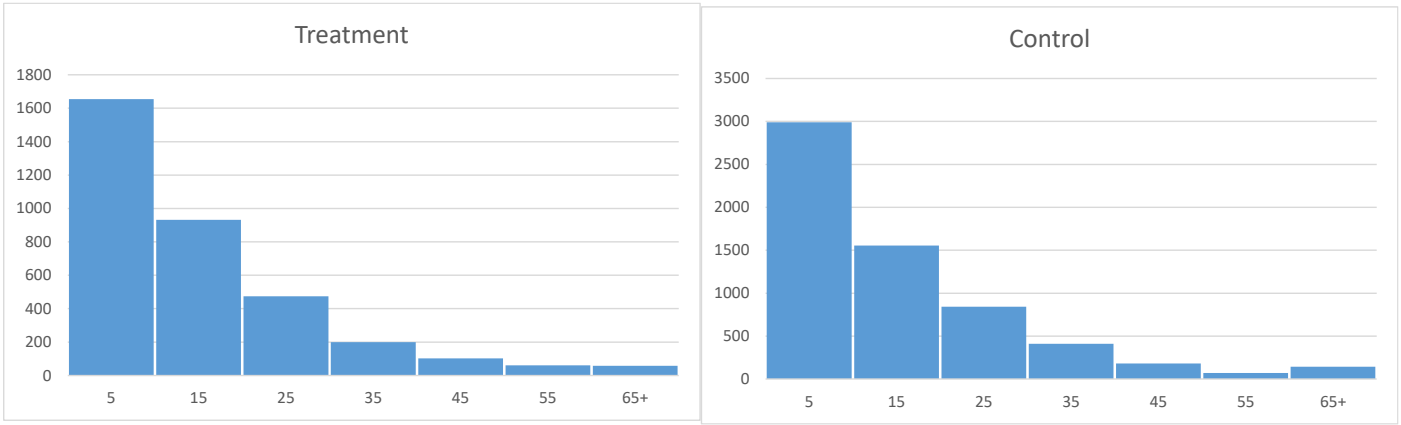


Figure G.3 Year 7 exclusions, by trial arm

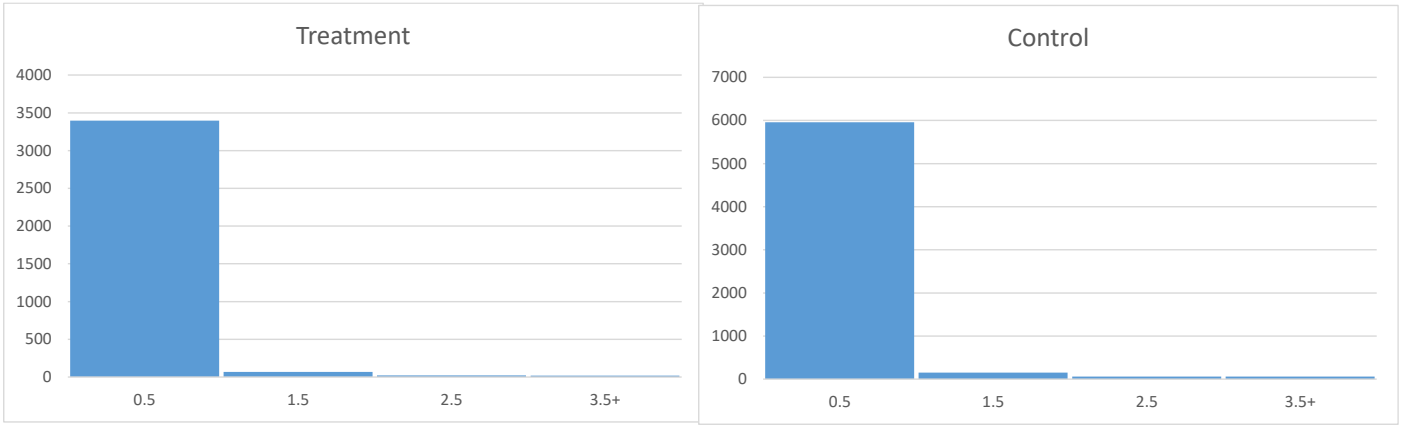


Figure G.4 Year 8 exclusions, by trial arm

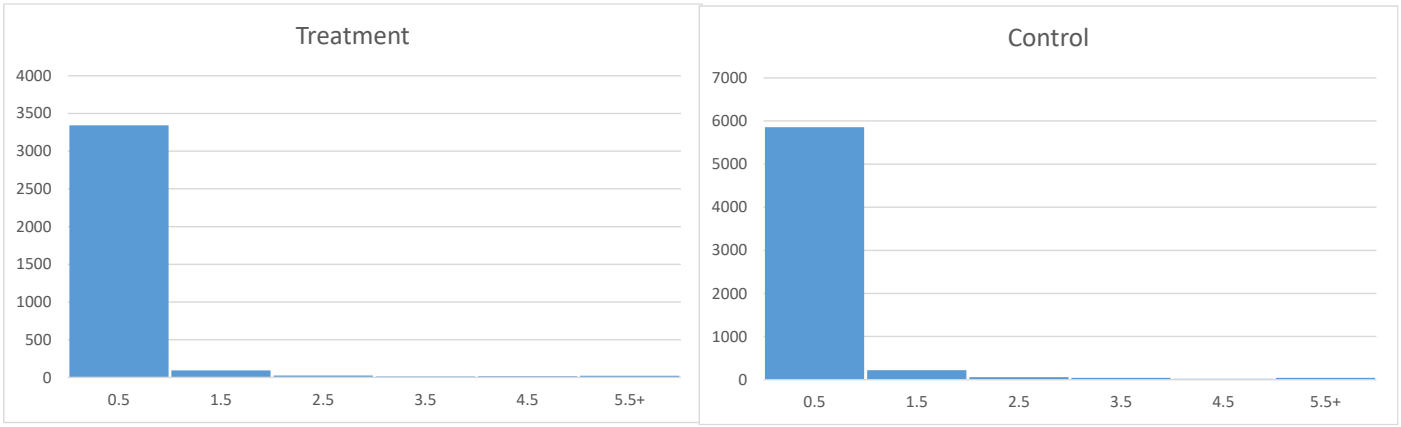


Figure G.5 Year 8 reading score, by trial arm

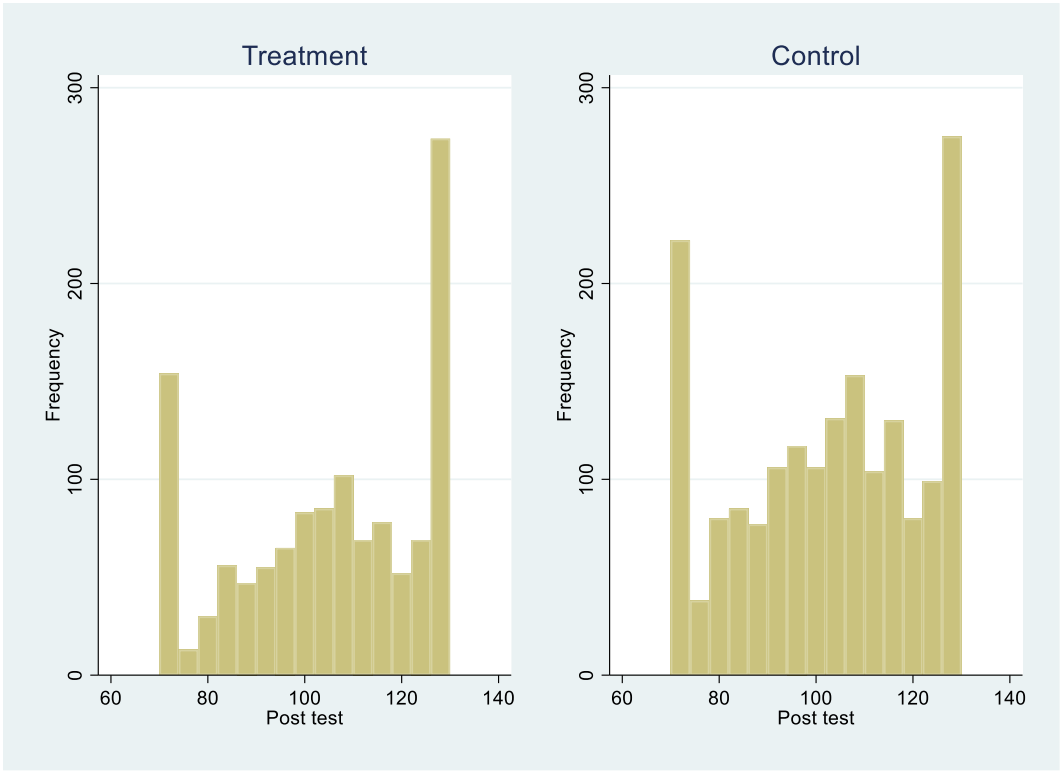


Figure G.6 Year 8 maths score, by trial arm

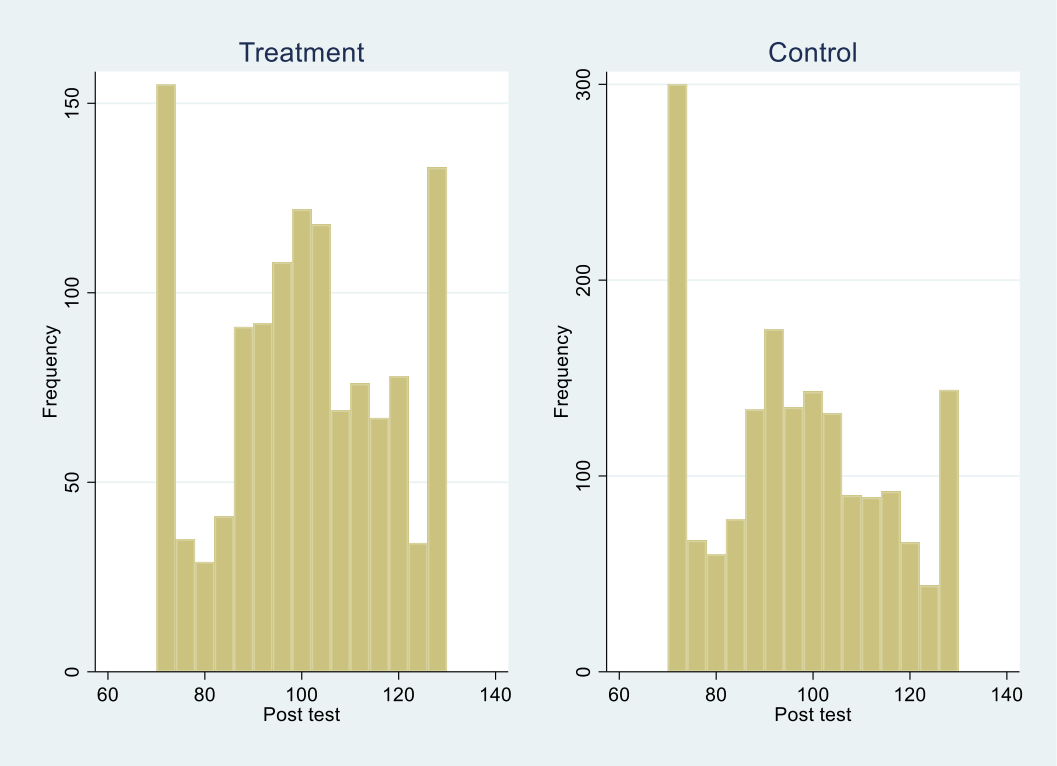


Figure G.7 Year 7 reading score, by trial arm

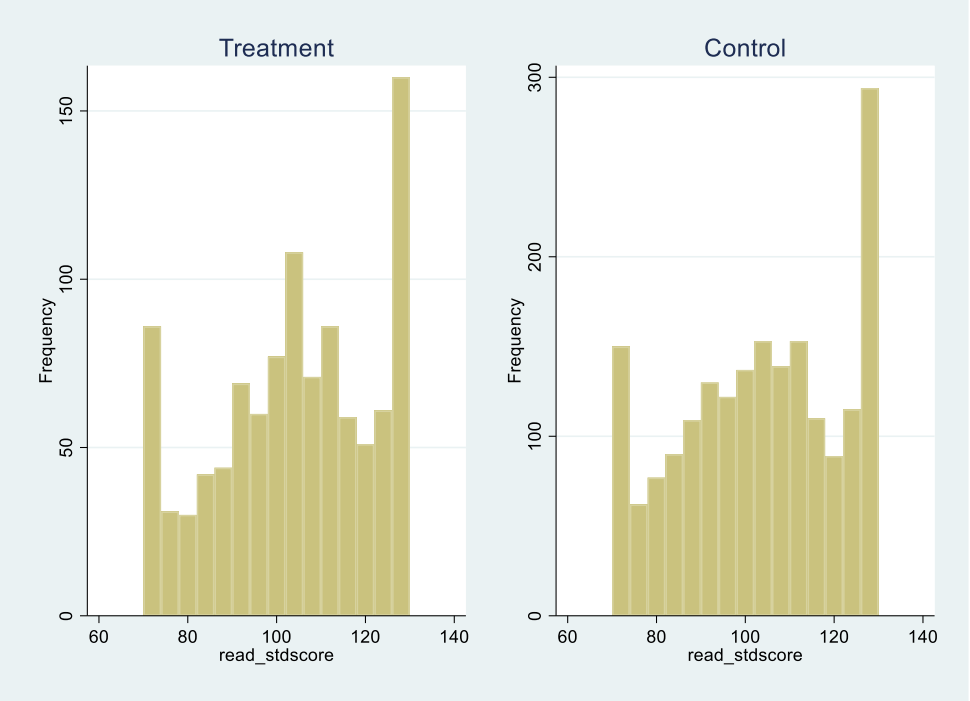
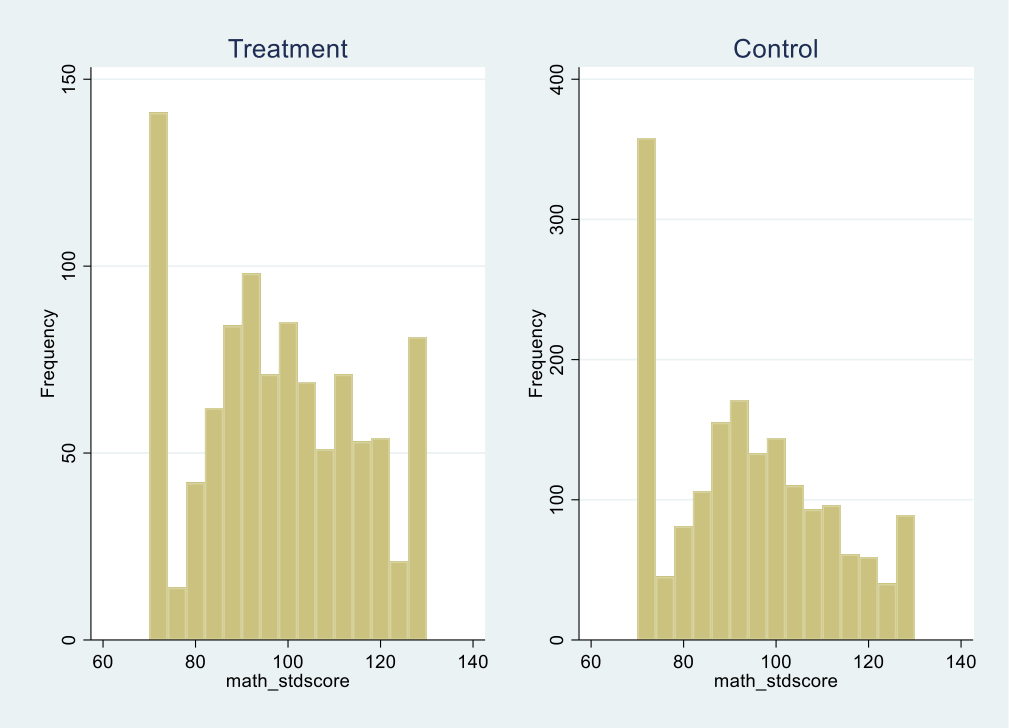


Figure G.8 Year 7 maths score, by trial arm



Appendix H: Absence and exclusions, sensitivity analysis

The results presented within the main report for absence and exclusions are estimated using the same estimation approach as for the attainment outcomes, in line with the approach documented in the statistical analysis plan (that is, using multilevel linear models). However, the distribution of both the absence and exclusions data (see Appendix G) mean that such an approach may be less appropriate, and we therefore report here the results of running multilevel negative binomial regression models as a sensitivity analysis.

Table H.1 presents results for our full sample. This indicates statistically significant impacts for absence in both years, with lower absence among the intervention group. Absence is lower by a factor of 0.87 in Year 7 and by 0.89 in Year 8, compared with the control group. Impacts for exclusions in both Year 7 and Year 8 are not statistically significant at conventional levels.

Table H.1: Absence and exclusions, multilevel negative binomial models

	Unadjusted means				Effect size			
	Intervention group		Control group					
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	IRR* (95% CI)	p-value	Regression coefficient (treatment variable)
Year 7 absence	3495 (243)	12.25 (11.73, 12.76)	6125 (822)	12.61 (12.22, 13.00)	9620 (3495; 6125)	0.87 (0.79, 0.96)	0.00	-0.14
Year 8 absence	3482 (256)	14.69 (14.09, 15.29)	6195 (752)	14.90 (14.45, 15.36)	9677 (3482; 6195)	0.89 (0.80, 0.99)	0.03	-0.11
Year 7 exclusions	3508 (230)	0.05 (0.04,0.07)	6232 (715)	0.09 (0.07, 0.10)	9740 (3508; 6232)	0.49 (0.23, 1.05)	0.07	-0.72
Year 8 exclusions	3508 (230)	0.11 (0.09, 0.13)	6232 (715)	0.13 (0.11, 0.15)	9740 (3508; 6232)	0.75 (0.39, 1.43)	0.38	-0.29

*IRR is the incidence rate ratio, where 1 indicates equivalence in the two groups (i.e. here the intervention and control group)

If we repeat this for the subgroup of pupils eligible for FSM (Table H.2), here we observe a statistically significant impact for exclusions in Year 7, with a lower rate of exclusions in the intervention group compared with the control group. Among pupils not eligible for FSM (Table H.3), we do not see impacts for exclusions in either year group, but find a statistically significant difference for absence in Year 7. If we include an interaction term between receiving the intervention and FSM status, rather than estimating the models for separate subgroups (Table H.4), we see a statistically significant interaction between FSM status and treatment for exclusions in Year 7 (but not in Year 8).

Table H.2: Subgroup analysis, pupils eligible for FSM, multilevel negative binomial models

	Unadjusted means				Effect size			
	Intervention group		Control group					
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	IRR* (95% CI)	p-value	Regression coefficient (treatment variable)

Year 7 absence	1,160	17.0 (15.9, 18.2)	2,166	16.8 (16.0, 17.7)	3326 (1160; 2166)	0.90 (0.78, 1.03)	0.12	-0.11
Year 8 absence	1151	19.7 (18.4, 21.0)	2165	19.9 (18.9, 20.9)	3316 (1151, 2165)	0.90 (0.78, 1.04)	0.17	-0.10
Year 7 exclusions	1,160	0.09 (0.06, 0.12)	2,167	0.20 (0.16, 0.24)	3327 (1160; 2167)	0.28 (0.12, 0.66)	0.00	-1.28
Year 8 exclusions	1,160	0.19 (0.14, 0.23)	2,167	0.26 (0.22, 0.30)	3327 (1160; 2167)	0.56 (0.27, 1.16)	0.12	-0.58

Table H.3: Subgroup analysis, pupils not eligible for FSM, multilevel negative binomial models

	Unadjusted means				Effect size			
	Intervention group		Control group					
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	IRR* (95% CI)	p-value	Regression coefficient (treatment variable)
Year 7 absence	2,326	9.9 (9.4, 10.4)	3,918	10.4 (10.0, 10.8)	6244 (2326, 3918)	0.84 (0.79, 0.93)	0.00	-0.17
Year 8 absence	2313	12.3 (11.6, 12.9)	3907	12.3 (11.9, 12.7)	6220 (2313, 3907)	0.89 (0.79, 1.01)	0.07	-0.11
Year 7 exclusions	2326	0.04 (0.02, 0.05)	3919	0.03 (0.02, 0.04)	6245 (2326; 3919)	0.84 (0.35, 2.01)	0.70	-0.17
Year 8 exclusions	2326	0.07 (0.05, 0.09)	3919	0.06 (0.05, 0.07)	6245 (2326; 3919)	1.06 (0.48, 2.34)	0.89	0.06

Table H.4: Subgroup analysis, interacting treatment and FSM status, multilevel negative binomial models

	Year 7 absence	Year 8 absence	Year 7 exclusions	Year 8 exclusions
Treatment	-0.157 (0.049)***	-0.104 (0.053)**	-0.109 (0.405)	-0.023 (0.343)
FSM	0.275 (0.027)***	0.317 (0.027)***	1.673 (0.157)***	1.285 (0.139)***
Treatment*FSM	0.047 (0.045)	-0.008 (0.044)	-1.009 (0.265)***	-0.357 (0.231)
N observations	9570	9536	9572	9572

Note: Each column shows selected coefficients from a regression of the outcome on treatment arm, eligibility for FSM, treatment*eligibility for FSM, cohort, phase and blocking variables, plus a dummy for missing pre-test data. Standard errors are in parentheses. Statistical significance is indicated as *** significant at 1%, ** significant at 5%

Finally we also repeat our subgroup analysis for prior attainment, interacting the treatment variable with prior attainment (Table H.5). None of the interaction terms are statistically significant, in line with the findings reported within the main report using linear models.

Table H.5: Subgroup analysis, interacting treatment and prior attainment, multilevel negative binomial models

	Year 7 absence	Year 8 absence	Year 7 exclusions	Year 8 exclusions
Treatment	-0.159 (0.056)***	-0.130 (0.057)**	-0.629 (0.421)	-0.148 (0.350)
Low prior attainment (bottom third)	0.081 (0.031)***	0.119 (0.030)***	0.770 (0.165)***	0.838 (0.151)***
High prior attainment (top third)	-0.175 (0.032)***	-0.185 (0.032)***	-0.860 (0.235)***	-0.886 (0.204)***
Treatment*Low prior attainment	0.020 (0.051)	0.004 (0.051)	-0.144 (0.292)	-0.315 (0.258)
Treatment*High prior attainment	0.016 (0.052)	0.058 (0.052)	-0.054 (0.397)	0.076 (0.320)
N observations	9375	9352	9392	9392

Note: Each column shows selected coefficients from a regression of the outcome on treatment arm, prior attainment (reference category: average prior attainment/middle third), treatment interacted with prior attainment, cohort, phase and blocking variables, plus a dummy for missing pre-test data. Standard errors are in parentheses. Statistical significance is indicated as *** significant at 1%, ** significant at 5%

Appendix I: Summary of interim findings

Two interim reports were produced as part of the evaluation of academic outcomes.

The first interim report considered outcomes at the end of Year 7 for all pupils, as well as outcomes at the end of Year 8 for pupils in Phase 1 schools only. Results are shown for reading in Table I.1 and for maths in Table I.2.

Table I.1: Primary Analysis, Reading, Year 7 and Year 8

	Raw means					Effect size		
	Intervention group		Control group					
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Interclass correlation	Hedges g (95% CI)	p-value
A. Year 7								
Reading	935 (173)	102.8 [101.6, 104.0]	1,571 (83)	102.9 [102.1, 103.9]	2,506 (339)	0.03	-0.02 [-0.13, 0.08]	0.66
FSM	150 (36)	95.2 [92.4, 98.1]	247 (16)	95.1 [92.9, 97.2]	397 (52)	0.000	-0.08 [-0.30, 0.14]	0.51
B. Year 8								
Reading	519 (16)	103.3 [101.7, 104.9]	455(0)	103.8 [102.1, 105.5]	975 (16)	0.030	0.03 [-0.31, 0.8]	0.85
FSM	61 (1)	102.4 [97.4, 107.3]	100 (0)	99.0 [95.2, 102.89]	161 (1)	0.000	0.06 [-0.47, 0.58]	0.84

Table I.2: Secondary Analysis, Maths, Year 7 and Year 8

	Raw means					Effect size		
	Intervention group		Control group					
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Interclass correlation	Hedges g (95% CI)	p-value
A. Year 7								
Math	902 (173)	96.5 [95.4, 97.7]	1,423 (188)	92.2 [91.3, 93.1]	2,325 (361)	0.27	-0.04 [-0.18, 0.10]	0.56
FSM	148 (34)	92.3 [89.6, 95.0]	207 (35)	96.5 [95.4, 97.7]	419 (69)	0.20	0.14 [-0.19, 0.47]	0.42
B. Year 8								
Math	564(1)	94.5 [93.1, 96.0]	410 (1)	95.7 [93.9, 97.4]	976 (2)	0.27	0.09 [-0.73, 0.90]	0.83
FSM	69 (1)	90.4 [86.4, 94.4]	89 (1)	89.7[85.9, 93.4]	158 (2)	0.09	0.10 [-0.50, 0.70]	0.75

The second interim report investigated outcomes at the end of Year 8, for pupils in Phase 1 and Phase 2 schools; Table I.3 reports results for reading and Table I.4 for maths.

Table I.3 Primary Analysis, Reading, Year 8

	Raw means						Effect size	
	Intervention group		Control group					
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	ICC	Hedges g (95% CI)	p-value
Year 8								
Reading	1365 (0)	104.4 [103.4, 105.5]	1843 (0)	102.2 [101.3, 103.1]	3208 (1365; 1843)	0.056	-0.08 (-0.22,0.06)	0.27
FSM	200 (0)	99.0 [96.2, 101.9]	278 (0)	95.7 [93.4, 97.9]	478 (200; 278)	0.063	-0.01 (-0.32, 0.31)	0.97

Table I.4 Secondary Analysis, Maths, Year 8

	Raw means						Effect size	
	Intervention group		Control group					
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	ICC	Hedges g (95% CI)	p-value
Year 8								
Maths	1395 (8)	99.3 [98.3, 100.2]	1775 (1)	96.0 [95.2, 96.9]	3170 (1395; 1775)	0.128	0.34 (0.18, 0.50)	0.00
FSM	202 (1)	92.5 [90.1, 95.0]	276 (0)	88.6 [86.6, 90.6]	478 (202; 276)	0.058	0.06 (-0.26, 0.37)	0.73

This document is available for download at https://educationendowmentfoundation.org.uk

Education
Endowment
Foundation

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk
@EducEndowFoundn
Facebook.com/EducEndowFoundn