Education Endowment Foundation

Fit to Study

Evaluation Report

September 2019

Fatima Husain, Vainius Bartasevicius, Lydia Marshall, Sandy Chidley, and Emma Forsyth

NatCen Social Research

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 (formerly Impetus Trust) 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.

This project was funded as part of the Education and Neuroscience scheme, which was jointly funded by Wellcome and Education Endowment Foundation and launched in January 2014. The aim of the scheme was to provide funding for collaborative projects between educators and neuroscientists to develop evidence-based interventions for use in the classroom, or to rigorously test existing tools and practices.

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

The project was independently evaluated by a team from NatCen Social Research: Fatima Husain, Lydia Marshall, Vainius Bartasevicius, Phoebe Averill, Tom Chadwick, Sandy Chidley, and Emma Forsyth, Nico Jabin authored the original analysis plan.

The lead evaluator was Fatima Husain. Lydia Marshall led the evaluation until May 2019)

Contact details:

Fatima Husain
NatCen Social Research
35 Northampton Square
London EC1V 0AX
Tel:020 7549 8508
Email: fatima.husain@natcen.ac.uk

Executive summary

The project

Fit to Study aimed to increase the amount of physical activity undertaken by Year 8 children in PE lessons. Previous studies had suggested that increasing students' physical activity could lead to improved cognition and academic attainment. However, these studies had not tested interventions that could work at scale. This project was designed to test whether Fit to Study could support a large number of schools to increase physical activity in order to benefit academic attainment. It was developed and delivered by a team of academics based at the University of Oxford and Oxford Brookes University.

The intervention required teachers to include two key activities in every PE lesson:

1. Four minutes of vigorous physical activity during the warm-up at the start of each lesson
2. Three 'fitness infusions' throughout the lesson which consist of two minutes of intense physical activity

The delivery team provided a two-hour, face-to-face training session for teachers on how to deliver the intervention. Teachers who were unable to attend were invited to a live webinar covering the same content. Teachers who attended a training session were expected to pass on what they had learned to other PE teachers at their school. The delivery team provided ongoing support throughout the school year through email and phone communications, and school visits.

Fit to study was evaluated using a randomised controlled trial involving 104 schools and 8,707 pupils. The primary outcome of interest for the trial was maths attainment, measured using the Progress Test in Mathematics. Alongside the impact evaluation, a mixed-methods process study explored implementation and teachers' responses to the programme. The process study included observations of the initial training delivered to PE teachers, observations of PE lessons, interviews with staff, and a teacher survey. The intervention was delivered throughout the academic year 2017/2018, with teacher training beginning in July 2017 and the post-test in June and July 2018.

Fit to Study was funded as part of the Education and Neuroscience scheme, a collaboration between the EEF and Wellcome Trust to provide funding for collaborative projects between educators and neuroscientists.

Key conclusions	
1.	There is no evidence that Fit to Study had an impact on Year 8 pupils' maths outcomes. This result has a low security rating.
2.	There is no evidence that Fit to Study had an impact on the maths outcomes of pupils eligible for free school meals.
3.	Generally, PE teachers struggled to implement Fit to Study as intended in each PE lesson as required. Implementation was difficult in lessons where skill development was a priority or large amounts of equipment were in use.
4.	Attendance at the initial face-to-face training was poor. Furthermore, training participants felt that the content of the training should focus less on theoretical aspects and more on how the intervention should be delivered in PE lessons and how delivery challenges can be addressed. Teachers reported that there were instances where Fit to Study disrupted lesson flow.
5.	Despite the implementation issues experienced by teachers, the majority of schools said they would recommend Fit to Study as a way to promote physical exercise.

EEF security rating

These findings have a low security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions. It was a well-designed, two-armed randomised controlled trial. The trial was not as well-powered as other EEF trials because recruiting a larger number of schools could have affected the quality of implementation under ideal developer-led conditions. 44% of the pupils who started the trial were not included the final analysis because of, problems around the implementation of the trial and difficulties with testing, and in some instances, due to external pressures.

Additional findings

Additional analyses did not suggest that Fit to Study had a differential effect according to gender or prior maths attainment.

The findings of this study are not consistent with previous research that has suggested that physical exercise interventions had a positive impact on pupils' cognitive functions. However, these findings are consistent with previous research that looked at the impact on academic attainment, which also failed to demonstrate impact, even with interventions using higher doses. An important difference between the current study and prior research is that while most prior studies have delivered longer sessions of moderate activity, the current study delivered brief bursts of vigorous activity. Therefore, it is important to note that the amount of vigorous activity specified for the intervention by the developers was an estimate to accommodate the structure of PE lessons in English state secondary schools.

The process evaluation suggested additional possible explanations for the lack of positive impact:

- Some PE teachers felt that bringing pupils together for the physical activity infusions required making regular stops to the lesson which resulted in the increase of 'dead' time. In some cases, this could lead to Fit to Study lessons being less active overall than regular PE lessons.
- There was inconsistent implementation and intervention fidelity was low. Only a small majority (53%) of the treatment schools that conducted the post-intervention testing indicated that more than half of all Year 8 PE lessons had been delivered as planned.

Although there is no evidence for positive effect on academic attainment in this trial, the majority of teachers felt that the intervention could be useful for increasing physical activity and fitness. This will be explored in the future by the developers through analysis of fitness data they collected.

Cost

The cost of Fit to Study is very low, between £4.56 and £4.80 per pupil per year over 3 years, depending on whether a school opted for face-to-face or online training. Participating teachers attended an initial two-hour training session.

Table 1: Summary of impact on primary outcome

Outcome/ Group	Effect size (95% confidence Interval)	Estimated months' progress	No. of pupils	P value	EEF security rating	EEF cost rating
Maths	-0.008 (-0.06, 0.05)	0	4,845	0.661	🔒🔒🔒🔒🔒	£ £ £ £ £
Maths FSM pupils	-0.01 (-0.12, 0.09)	0	1,448	0.709	N/A	£ £ £ £ £

Introduction

Background evidence

The role of regular physical activity in people's lives has been the subject of research in a number of different fields; most obviously in medicine, focusing on the benefits of following healthy lifestyles at both the individual and societal level (Ortega et al 2018; Warburton et al 2006; However, there is increasing focus on the benefits that participation in sport or PE provision that requires Moderate to Vigorous Physical activity (MVPA) can have on other aspects of peoples' lives, including mental well-being, social development (such as leadership skills and team working), social interactions, cognition, and educational attainment (Booth et al 2014; Roberts et al. 2010) .

Participation in physical activity has been the focus of policy makers for some time. A 2013 cross-government initiative in the UK set out a priority to increase participation in sport amongst people of all ages, but with a focus on engagement amongst children and young people (DCMS and DfE, 2015). The initiative set out funding through the PE and Sport Premium, to expand the School Games Programme. It included a commitment to support Sport England to help 14-15 year olds to keep playing sports throughout their lives. An Ofsted review found that whilst there had been an increase in Good and Outstanding PE teaching in schools, there was still significant progress to be made for PE in secondary schools to consistently contribute to pupils' fitness. The review suggested increasing time given to teaching core PE each week and providing weekly opportunities to attain high standards of performance and engage in regular, high-intensity vigorous activity for sustained periods of time.

In addition to general concerns about participation in physical activity, the level of activity and engagement during PE lessons may vary by pupil characteristics, such as gender. A Government review of participation in sport among girls highlighted low levels of engagement as a concern. The drop in girls' participation levels begins in primary school and becomes more pronounced in years 8 and 9 of secondary school (between 12-14 years of age) (DfE, 2013). This trend continues as girls' experiences of PE and sports at school are likely to affect attitudes and engagement in sport later in life. Moreover, girls' and women's participation in sport is significantly lower among lower socio-economic groups (House of Commons, 2014)1. Schools, particularly those in deprived communities, have an important role to play in motivating all pupils to increase participation in PE and sports through effective PE provision and teaching.

Recent reviews of studies looking at the impact of increasing MVPA2 in children on academic achievement, found that heart rate increases during PE lessons had either a positive relationship with attainment or none at all with virtually no studies reporting a negative effect3(CDCP, 2010)., Evidence from the field of neuroscience has shown that MVPA increases brain volume, blood flow and connectivity, particularly in the hippocampus (Erikson et al, 2011, Chaddock-Heyman et al., 2013), which may help improve cognitive functions such as working memory and attention (Phan et al. 2018) . Analysis of longitudinal datasets confirms neuroscience findings in relation to physical activity and improved cognitive function as well as on classroom behaviour, which may enhance academic performance (Steed and Nevill, 2010 and Booth et al, 2014). Steed and Nevill (2010) suggest that further research is needed to establish the optimal intensity and duration of physical activity (and MVPA) to improve cognitive function. Also, conclusions from two recent studies pointed to the positive effects of physical exercise interventions on pre-adolescents' cognitive functions, including working memory and brain development (Kamijo et al, 2011; Chaddock-Heyman et al., 2013).

However, some other studies failed to detect any impact of the physical exercise on pupils' attainment. A three year cluster randomised trial in the USA involving 24 schools tested Physical Activity Across the Curriculum (PAAC), a school based intervention that promoted 90 minutes of MVPA/week in addition to a 60 minute PE lesson[4] found no significant

[2] An MVPA fact sheet (2009) defines *moderate* physical activity as activities equivalent in intensity to brisk walking or bicycling. *Vigorous* physical activity is an activity that produces large increases in breathing or heart rate, such as jogging, aerobic dance or bicycling uphill. http://www.californiaprojectlean.org/docuserfiles/200911_MVPA_FactSheet%5B1%5D.pdf

[3] EEF (2014) Neuroscience and Education: A review of Educational Interventions and Approaches Informed by Neuroscience.

[4] This intensity of physical activity was based on the U.S Department of Health and Human Services guidelines set out in Healthy People 2010.

difference in attainment (a secondary outcome measure) among a sub-sample[5] of participants (Donnelly, 2009). More recently, a trial based in England that tested two interventions to modify MVPA among Year 7 pupils in 60 secondary schools analysed accelerometer data and assessed level of well-being. It concluded that 'modest' classroom initiatives, such as 1:1 peer mentoring sessions between Year 9 and Year 7 pupils to increase physical activity, are unlikely to have a meaningful impact on MVPA or feelings of well-being (Tymms et al, 2016).

There is some evidence pointing to the differential impact of physical activity on academic attainment for boys and girls. For example, Carlson et al. (2008) observed a small but significant benefit of extra physical exercise for academic achievement in mathematics and reading for girls but not for boys. Also, some studies testing the impact of physical interventions on academic achievement found different effect sizes for pupils with low and high prior attainment. Testing a new classroom physical intervention, McClelland et al. (2015) found that the improvements in pupil academic performance were largest for struggling pupils performing in the lowest 20 percent.

The majority of evidence on physical activity and academic attainment or cognitive function has focused on MVPA. However, there is increasing interest in the possible role of short bursts of high intensity or vigorous physical activity (VPA) at influencing the brain and cognition (Moreau, D. et al, 2017; Costigan, S. A. et al, 2016; Kujach S, et al., 2018; Cooper, S. B. et al, 2016).

The Education Endowment Foundation and the Wellcome Trust are working together to build research expertise and knowledge at the interface between neuroscience and education. The evaluation of the Fit to Study intervention, an EEF and Wellcome Trust collaboration, adds to this body of evidence by measuring levels of physical activity and assessing changes in cognitive function and attainment among secondary school pupils.

Intervention

The Fit to Study intervention was designed by academics at Oxford University and Oxford Brookes University in collaboration with stakeholders including the UK Sports Partnership, PE teachers and head teachers. The intervention was piloted in a small number of schools and evidence from the evaluation of the pilot, conducted by NatCen, informed the review and redesign of the intervention. Further information on the rationale for and design of the intervention, as well as details of secondary measures, are available in a protocol paper published by the developers (Wessenar et al 2019).

As highlighted above, neuroscience evidence suggests that physical activity contributes to improved cognitive function and then to improved attainment. Therefore, it could be expected that within one academic year an increase in physical activity will lead to improved attainment.

The Fit to Study intervention aimed to increase activity levels within PE lessons because PE lessons presented the best opportunity within the school day to improve activity.

Previous intervention trials have tended to focus on MVPA sessions of at least 60 minutes per week. However, the Fit to Study developers focused on designing an intervention that could be delivered within the context of the school curriculum. Therefore, in contrast to many previous trials, the Fit to Study intervention focused on short bursts of vigorous physical activity (VPA) that could be incorporated into a regular PE lesson. Short bouts of VPA or high-intensity activity, during which individuals work at around 80% of their maximum heart rate for between 45 seconds and four minutes, have been shown to deliver equivalent fitness benefits to longer, lower-intensity workouts (Buchheit & Laursen, 2013; Costigan et al, 2015). This type of vigorous activity was chosen by the developer as it is brief enough to be incorporated into PE lessons without disrupting curriculum delivery, and there is evidence that it can improve adolescent fitness (Moreau et al, 2017; Costigan et al, 2016; Kujach, S. et al, 2018, Cooper et al, 2016; Logan et al, 2014).

The Fit to Study intervention was developed in two stages:

[5] Sub-sample measures were obtained at baseline and at three years. Detailed descriptions of all sub-sample evaluations can be found: DuBose KD, Mayo MS, Gibson CA, Green JL, Hill JO, Jacobsen DJ, Smith BK, Sullivan DK, Washburn RA, Donnelly JE. Physical activity across the curriculum (PAAC): rationale and design. Contemp Clin Trials. 2008;29:83–93

1. A developmental stage during which the developers explored the most appropriate delivery approaches and design for Fit to Study.
2. A pilot stage involving 8 schools. This was conducted in two phases. 4 schools delivered Fit to Study[6] in the first phase. Emerging findings led to a review of the intervention. The remaining 4 schools delivered the revised Fit to Study intervention during the second pilot phase. Each phase was delivered over one school term.

The intervention was designed as a whole school intervention that could be incorporated in PE lessons. The trial focused on Year 8 pupils (aged 12-13) because it is around this age that physical activity levels begin to drop off, particularly among girls (Inchley et al., 2016), and where the adolescent brain is developing rapidly (Blakemore and Choudhury 2006). Fit to Study is targeted at increasing vigorous physical activity (VPA) during PE lessons. Short bouts of VPA have been shown to deliver equivalent fitness benefits to longer, lower-intensity workouts (Buchheit & Laursen, 2013). This type of activity has been shown to be brief enough to be incorporated into PE lessons without disrupting curriculum delivery (Logan et al., 2014) and may be sufficient to bring about cognitive changes, although evidence is inconclusive (Costigan et al, 2016).

As part of the intervention PE teachers were provided training and guidance so that they could adapt PE lessons to create more opportunities for pupils to achieve the required level of VPA during PE lessons.

Training was planned to be delivered by the developers face-to-face to all Year 8 PE teachers from every intervention school. The training comprised a two-hour session delivered at two hubs where teachers from different schools could attend together. Up to 30 teachers could take part in a session. The initial plan was to deliver the training session before the end of in July 2017 and before the end of the 2016/2017 school year. The possibility of cascading training within a PE department was to be used only if a teacher was unable to attend and this possibility was not offered as an alternative to face-to-face training. Developers provided the face-to-face training where the rationale of Fit to Study was presented and attending teachers were given the opportunity to discuss the ways in which they would deliver Fit to Study within their PE lessons.

Four modes of training were implemented: face-to-face training sessions, live webinar, pre-recorded training videos and cascading.

Teachers who were unable to attend a face-to-face training session were invited to a live webinar covering the content of the face-to-face training in September 2017. Webinars were offered on multiple dates to suit school availability. Where teachers were unavailable to attend a live webinar, they were sent a link to a pre-recorded video of the training hosted online. Uptake of the training by teachers is presented in the implementation and process evaluation section.

Teachers who attended a training session could cascade what they had learned to other PE teachers at their school. On-going support was provided by the developers throughout the school year, including on-going communication and school visits. “Top-up” training was incorporated into the initial school visits made by the developer. These visits also provided an opportunity for any other ad-hoc support required by schools.

Training covered the scientific rationale for the research, what schools would need to do to deliver the intervention, and the requirements for trial participation (for example, filling in lesson logs). More details on the contents of the training is contained in Appendix K.

During the trial Fit to Study consisted of two key components which PE teachers had to implement in each PE class:

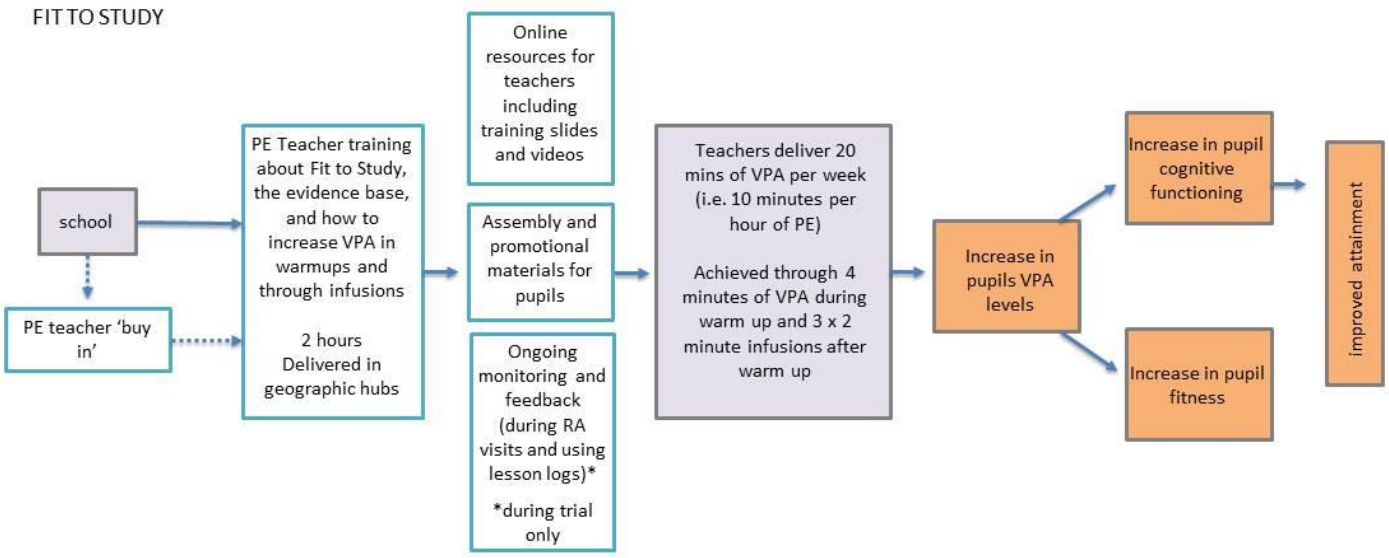
1. Four minutes of VPA during the warm-up at the start of each PE lesson
2. Three ‘fitness infusions’ which are short intense bursts of VPA lasting 2 minutes each.

Therefore, each PE lesson structure included at least 10 minutes of VPA. The assumption was that if pupils would take part in two Fit to Study PE lessons per week, pupils would be engaged in at least 20 minutes of VPA per week which would benefit pupil brain structure function and thereby improve executive function, attention and working memory, which in turn would improve learning – resulting in a positive impact on attainment.

[6] https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/fit-to-study

The intervention was delivered by PE teachers during scheduled PE lessons. The intervention rationale is set out in the logic model below:

Figure 1: Fit to Study logic model



The cognitive domains that have been previously reported to improve with physical activity interventions include executive function and memory (Khan, N. A., and Hillman, C. H., 2014; Mura, G. et al, 2015; Fisher, A. et al, 2011; Monti, J. M., Hillman, C. H., Cohen, N. J., 2012; Hillman, C. H. et al., 2014), functions that would be relevant across a wide range of academic disciplines. In line with this, prior studies have reported effects of physical activity on both maths and literacy (Donnelly et al, 2016; Hillman et al., 2008). However, testing constraints and cost and a desire to minimise school burden meant that a single outcome measure had to be defined for this study. In conjunction with the developers, and agreed upon by EEF, the attainment measure used was a maths test[7].

The National Foundation for Educational Research (NfER) recruited schools for the trial, leading to the total recruitment of 104 schools (more detail can be found in the participant flow diagram in the Impact Evaluation section). The delivery of the Fit to Study intervention in schools was varied, and a number of factors were found to influence its implementation. These are explored in more detail in the implementation and process evaluation section of this report. Thirty-one schools dropped out of the study for a variety of reasons listed in the participant flow diagram in the impact evaluation section.

Evaluation objectives

The efficacy trial aimed to answer the following principal research questions:

- What is the impact of Fit to Study intervention on Maths attainment of participating pupils?
- What is the impact of Fit to Study intervention on Maths attainment of participating pupils (a) with low prior attainment and (b) eligible for free school meals?
- What is the impact of Fit to Study intervention on Maths attainment of participating (a) girls and (b) boys?

Alongside the impact evaluation, a process study to assess implementation fidelity (that is, whether the intervention was delivered as intended) was carried out. Exploring implementation fidelity is important for furthering understanding of any contextual and experiential considerations that affect implementation, and to inform analysis of intended primary and secondary outcomes. The main research questions that the process evaluation sought to answer were:

- How is the programme delivered and what factors influence implementation fidelity?
- What type of PE lesson modifications take place in control schools during the treatment period?

[7] https://www.gl-assessment.co.uk/products/progress-test-in-maths-ptm/

- What more generally are the barriers and necessary conditions for success?
- What is PE teachers' level of engagement with available external support and guidance?

The evaluation protocol[8] for Fit to Study and the statistical analysis plan (SAP)[9] can be found on the Education Endowment Foundation's website.

The table below shows which IPE questions will be answered by which data course including the number of encounters.

[8] https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Neuroscience_-_Fit_to_Study_Trial_Protocol.pdf
[9] https://educationendowmentfoundation.org.uk/public/files/Projects/Fit_to_Study_SAP_2018.03.15_FINAL.pdf

Implementation and process evaluation research questions and data sources

IPE domain	IPE RQ	Data source	Number of encounters
Fidelity, Adaptation	How is the programme delivered and what factors influence implementation fidelity?	Online school survey	All schools (treatment and control group)
		Phase 1 interviews	15 telephone interviews with treatment schools 5 interviews with control group schools
		Phase 2 visits (interviews and observations)	Interviews with PE staff and observations of PE lessons in 7 treatment schools
		Phase 3 interviews	10 telephone interviews with treatment schools
Compensatory activities	What type of PE lesson modifications take place in control schools during the treatment period?	Online school survey	All schools (treatment and control group)
		Phase 1 interviews	5 interviews with control group schools
		Phase 2 visits (interviews and observations)	Interviews with PE staff and observations of PE lessons in 3 control groups schools
Conditions for success	What more generally are the barriers and necessary conditions for success?	Online school survey	All schools (treatment and control group)
		Phase 1 interviews	15 telephone interviews with treatment schools
		Phase 2 visits (interviews and observations)	Interviews with PE staff and observations of PE lessons in 7 treatment and 3 control groups schools
		Phase 3 interviews	10 telephone interviews with treatment schools
Engagement	What is PE teachers' level of engagement with available external support and guidance?	Online school survey	All schools (treatment and control group)
		Phase 1 interviews	15 telephone interviews with treatment schools
		Phase 2 visits (interviews and observations)	Interviews with PE staff and observations of PE lessons in 7 treatment
		Phase 3 interviews	10 telephone interviews with treatment schools

Ethics and trial registration

Ethical approval for the evaluation was obtained from University of Oxford's Medical Sciences Inter-divisional Research Ethics Committee. Approval was granted in January 2017. Schools were invited to sign a memorandum of understanding (MoU) to signal their intention to take part in the trial and associated evaluation activities. Once this agreement to participate in the trial was received, the schools was considered recruited into the trial. See Appendix F for the MoU and information letter for parents.

The trial was registered with ISRCTN. The trial number is 15730512[10]. Oxford University separately registered the Fit to Study trial with ClinicalTrials.gov.

Data protection

Upon the introduction of the General Data Protection Regulation (GDPR) on 25th May 2018, a privacy notice for the evaluation of Fit to Study was introduced. Schools were informed that NatCen would have access to pupil, school and teacher information, recordings of interviews, transcripts, charted data, sample files (including contact details), survey responses and anonymised test results. They were also informed that all responses would be anonymised before being analysed and subsequently archived.

[10] http://www.isrctn.com/ISRCTN15730512

Third parties who would also have access to certain information were listed as academics from the University of Oxford and Oxford-Brookes University, McGowan Transcriptions, Formara Print+ and GL Assessment.

NatCen was the data controller and data processor for this project. This means that NatCen were responsible for deciding the purpose and legal basis for processing data (under Article 6 of GDPR). For this project, the legal basis for processing data was "legitimate interest". NatCen's privacy notice for this project can be found in the appendix and on the NatCen website[11].

Project team

The developers at the University of Oxford and Oxford-Brookes University were responsible for the delivery of Fit to Study in schools.

The evaluation of Fit to Study was led by NatCen's Children and Families Team, who were responsible for the day-to-day project management of the study. The team worked closely with impact evaluation experts from NatCen's Evaluation Team, who provided expertise on the design of the trial and testing. Contact with schools, pupil testing and data entry processes were coordinated through NatCen's Operations Department. The draft analysis plan for the study was written by Nico Jabin, Research Director Evaluation, who has since moved to another organisation. The final version of the plan was agreed with EEF by Vainius Bartasevicius, Fatima Husain, and Lydia Marshall.

Table 2: NatCen evaluation project team

Children & Families Team		
Dr Fatima Husain	Director (interim)	Overall study lead
Dr Lydia Marshall	Research Director	Project lead (until June 2018)
Sandy Chidley	Senior Researcher	Day to day project manager
Emma Forsyth	Research Assistant	Supporting testing and qualitative fieldwork
Evaluation Team		
Vainius Bartasevicius	Senior Researcher	Impact analysis

[11] http://natcen.ac.uk/taking-part/studies-in-field/fit-to-study/privacy-notice/

Methods

Trial design

The trial of Fit to Study was an efficacy trial, exploring the intervention's potential to improve outcomes under as ideal circumstances as possible and looking at a restricted cohort (Thorpe et al. 2009). The trial was designed as a two-arm, two-level cluster-randomised trial, in which pupils (level 1) were clustered within schools (level 2). The intervention was targeted at all Year 8 pupils in participating schools with all Year 8 PE teachers receiving training. However, given that the training could also be cascaded to other teachers in the PE department randomisation was performed at school level. Accounting for the hierarchical structure of the data was necessary to avoid underestimating the standard errors of the treatment effects[12]. Half of the participating schools were randomly assigned to implement the Fit to Study programme while the remaining half were allocated to control and carried on with business as usual. Within randomised schools, a sub-sample of Year 8 pupils were randomly selected for the testing for the primary outcome measure considered in this evaluation report. All Year 8 pupils in participating schools were targeted for secondary outcome measures collected by the developers. Information about the secondary outcome data collection is available in Wassenar et al, 2019, the protocol published by the developer (Wassenaar et al, 2019).

Table 3: Trial design characteristics

Trial type and number of arms		Efficacy, 2-arm trial
Unit of randomisation		School
Stratification variable(s) (if applicable)		Gender-status of schools (co-educational or single-sex)
Primary outcome	variable	Maths attainment
	measure (instrument, scale)	Raw scores achieved in the Progress Test in Mathematics (PTM), Level 13 (in the scale of 0 to 70)
Secondary outcome(s)	variable(s)	-
	measure(s) (instrument, scale)	-

We made one important change to the original trial design described in the protocol. In the protocol we originally proposed to stratify randomisation by the gender-status of schools (co-educational or single-sex) and by geography. Stratification by gender-status was underpinned by the assumption that the delivery of PE lessons may be different in co-educational and single-sex schools. It was also believed that there might be some differences in the implementation of PE teacher training across different geographical regions. Due to changes to the way teachers were trained (following the introduction of the online training element), stratification by geography was no longer required. We also intended to retain stratification by urban/rural status, as this may reflect different existing exercise regimes and thus influence treatment effectiveness. However, stratification by both gender-status and an urban/rural status of schools proved to be impossible because all rural schools selected for the trial were co-educational. Therefore, stratification was done by gender-status of schools only. This change has been described and explained in the SAP.

Participant selection

To take part in the trial, schools had to meet the following criteria:

[12] Education Endowment Foundation, 2018. *Statistical analysis guidance for EEF evaluations* , p. 3. Available at https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf.

- Type: state schools including free schools were eligible to take part. Selective/grammar schools were not eligible.

- Location: For the developers to have ease of access, schools had to be located in one of following geographic regions: Greater London; Thames Valley, Southampton and Portsmouth, Bristol and Bath, Birmingham and Coventry, Cheltenham/Gloucester, Luton, Bedford and Milton Keynes.

School participation was voluntary and financial incentives were used. Both control and treatment schools received £500 for taking part in the trial. Recruitment of schools was led by NFER with support provided by Oxford Brookes University. NatCen contributed to relevant recruitment materials and ensured that information about the trial was provided. These materials included evaluation information documents, the memorandum of understanding (MoU), and parent information and opt-out documents. These can be found in Appendix F.

In each school, all pupils in all Year 8 form groups received the intervention. However, to reduce the burden of testing on schools and in consideration of the costs associated with testing, only a set proportion of form groups in each school were randomly selected for participation in testing for the primary outcome measure. In the protocol we suggested that three Year 8 form groups per school would be selected for pupil testing for the primary outcome. However, this was revised as selecting three form groups independent of the school size would have given more weight to smaller schools. As an alternative, 49% of Year 8 forms were randomly selected for primary outcome testing in each school. Selecting a proportion of forms relative to the school size ensured a more equal weighting to each pupil, rather than attributing more weight to pupils in smaller and less weight to pupils in larger schools.

Outcome measures

The primary outcome of interest for the trial was maths attainment, as measured by the raw scores achieved in the Progress Test in Mathematics (PTM), Level 13 (GL Assessment 2015). PTM is a standardised group test that assesses pupils' mathematical skills and concepts. The test is aligned to the new national curriculum and is an accurate measure of progress in maths. PTM Level 13 is suitable for pupils completing Year 8 and takes 75 minutes to complete. PTM assesses two dimensions of learning:

1. Mathematical content knowledge (Curriculum Content Category)

2. Understanding and applying mathematical processes through reasoning and problem solving (Process Category).

The test was administered by NatCen invigilators in schools post-treatment towards the end of the 2017/18 summer term. Invigilators were blind to the school's treatment allocation and were instructed not to enquire about the school's allocation status. However, we could not control for schools disclosing their status to invigilators. The automated marking was completed by GL Assessment.

No secondary outcomes were assessed during the trial. Apart from attainment, the Fit to Study logic model also identifies the increase in pupils' fitness as a potential outcome. Fitness data, along with a number of other secondary measures, was collected by the developers and falls out of scope of this evaluation (refer to the protocol by Wassenaar et al, 2019).

Key Stage 2 (KS2) scores for maths were used as a pre-test measure in this trial. Data on the raw KS2 scores was obtained from the National Pupil Database (NPD).

Sample size for primary outcome assessment

At protocol stage, only preliminary sample size calculations were provided. The aim was to recruit around 100 schools for the trial and randomly select three Year 8 form groups (around 90 pupils) for assessment in each school. The estimated pupil sample size was 9,000[13].

[13] For full power calculations at protocol, randomisation and analysis stages, please see the impact evaluation section of this report.

In total 106 schools were recruited to the trial, two of which dropped out before randomisation. All Year 8 pupils received the treatment in treatment schools. For the trial only 49% of Year 8 form groups in each treatment and control school were selected for post-intervention assessment. As a result, 8,707 pupils out of 18,299 Year 8 pupils in the 104 randomised schools were randomly selected for testing. However, a further 18 schools dropped out during the intervention phase, making a total of 31 schools that dropped out of the trial before post-intervention testing was completed, reducing the sample size to 6,020 pupils[14]. Pupil absences from post-intervention test further decreased the analysis sample. Post-test data was only obtained for 5,098 pupils from 73 schools. Due to the absence of baseline scores (Key Stage 2 data obtained from the National Pupil Database) for 253 pupils, our final number of analysed pupils was 4,845, with 1,448 of them eligible for FSM. In line with EEF statistical analysis guidance, FSM eligibility was determined using the EverFSM indicator (EVERFSM_6_P)[15]. For the full information on sample sizes as well as MDES achieved at different stages of the trial, please see Figure 2 and Table 7.

Randomisation

School randomisation

Schools were randomly assigned to treatment and control using a stratified block randomisation. Stratification was done by single-sex / co-educational status of schools. Both treatment and the control group had 10 single-sex schools each. Due to changes in the training of PE teachers by the developer, geographical stratification as described in the protocol was not required. Randomisation was performed by Nico Jabin from the evaluation team at NatCen Social Research (the syntax is provided in Appendix C).

Pupil selection for testing

In addition to assigning schools, 49% of Year 8 form groups in each school were randomly selected to participate in primary outcome testing. Randomly selecting form groups as opposed to including all Year 8 pupils into the analysis made the process of post-intervention testing less burdensome for schools and helped to reduce the costs associated with it. Selecting a proportion of forms ensured that pupils from small and large schools are given the same weight in the analysis. Choosing 49% of form groups in each participating school allowed us to approximate as much as possible the number of pupils that we originally planned to test (9,000).

Only those pupils that we intended to test (those in the randomly selected form groups) are considered to be part of the trial for the purposes of this independent evaluation.

An unequal probability sample without replacement was drawn, meaning that each class could only be selected once. Sampling probabilities were proportional to class size. Classes with a larger number of pupils had a higher likelihood to be selected for the trial. With the aim of testing as many Year 8 pupils as possible, we excluded form groups with fewer than 10 pupils if there were classes with more than 10 pupils in the same school. Selection of form groups for pupil testing was carried out by Vainius Bartasevicius (the syntax is provided in Appendix C).

Statistical analysis

Primary intention-to-treat (ITT) analysis

The primary analysis examined the hypothesis that Fit to Study improves attainment in mathematics, as measured by a post-intervention Progress Test in Mathematics (PTM), Level 13 (GL Assessment 2015). The analysis was conducted on an intention-to-treat basis and included pupils in all Year 8 form groups that were randomly selected to be part of the trial within the randomised schools, subject to agreement and successful collection of outcome and baseline data.

Evidence of effectiveness and reported effect sizes were obtained from a baseline-adjusted analysis, in which the dependent variable was the result of the outcome test, and effects were estimated through a multilevel linear model containing a dummy variable indicator capturing treatment/control group membership, the stratification variable (gender-status of school), and pupil level baseline test scores. Baseline test scores was the only pupil-level covariate used to

[14] The details on school attrition as well as the reasons for it are outlined in the impact evaluation section.
[15] The indicator shows whether the pupil has ever been eligible for FSM in the last 6 years up to the pupil's current year (not including nursery).

improve the precision of impact estimates in the main model. The model analysed was a two-level model in which pupils were clustered in schools, with schools modelled as random effects using a random intercept model (see equation (1)).

$$Test_{ij} = \beta_0 + \beta_1 baseline_{ij} + \beta_2 intervention_j + \beta_3 stratifiers_j + u_j + e_{ij} \quad (1)$$

where i presents the individual, j presents the school, and u_j is the school random effect. The intervention effect was estimated by β_2.

The analysis was run in Stata. Analysts were not blind to treatment status of schools.

Imbalance at baseline for analysed groups

Baseline characteristics were summarised by treatment and control group status across schools and pupils.

At school level, the comparison covered:

- School type (academy, free school, etc.)
- Number of Year 8 pupils
- Number of Year 8 forms
- Urban/rural status
- Gender status
- Percentage of absence sessions in previous academic year (16/17)
- Percentage of pupils eligible for FSM

At the pupil level, the following baseline comparisons were presented:

- Eligibility for free school meals
- Gender
- Key Stage 2 maths scores (obtained from the National Pupil Database)
- Number of absences from school in previous academic year (16/17) (obtained from the National Pupil Database)

Imbalances on baseline covariates between the treatment and control groups in the sample as analysed were assessed for the covariates listed above using the appropriate statistical test (two-independent-sample t-test for continuous variables and Fisher's exact test for categorical variables), with a p-value of 0.05 or smaller considered as indication of covariate imbalance. Differences in baseline outcomes were reported as Hedges' g effect sizes.

Sensitivity analyses

Three additional sensitivity analyses were carried out to explore the robustness of the main finding:

1. An unadjusted analysis that did not include any baseline covariates;
2. An adjusted analysis which included the unbalanced covariates in addition to those in the main model;
3. An adjusted analysis that included any variables we believed were likely to be predictive of outcomes.

The second analysis was originally described in the SAP as part of the analyses if imbalances were present.

Missing data

Our primary analysis assumed that any missing outcome data was missing completely at random (MCAR) and used complete case analysis. We also assessed the robustness of our results to alternative assumptions about the mechanisms leading to missing data. We explored the number, pattern and likely reasons for missing outcome data. We ran a drop-out model to assess whether any existing covariates predict the observed loss-to-follow up pattern. We estimated the propensity score using multivariate logistic regression. In this case, the propensity score can be thought of as the conditional probability that a pupil has been tested given the set of existing covariates capturing pupil characteristics.

Having found that a number of covariates were associated with the missingness of outcome data, we imputed the missing post-test scores using multiple imputation, under the assumption that data was missing at random (MAR). We imputed a number of new datasets using a two-level linear model: pupil (level 1), and school (level 2). The model included all variables in the main analysis; those variables that were predictive of missingness; and variables associated with the outcome. The estimated regression equations were then used to generate predicted values for the cases with missing data and estimate treatment effects and standard errors under the alternative assumption. Rubin's rules were used to combine the estimates from multiple datasets into an overall MI estimate (Rubin, 1987). More details on multiple imputation procedure are provided in the impact evaluation section.

In the analysis plan, we also said we would explore the robustness of our results to extreme suppositions under the assumption that data are missing not at random (MNAR). However, after performing the multiple imputation under the MAR assumption, we decided that additional sensitivity analysis was no longer needed. This decision was informed by two key considerations. First, we found that missing outcome data was conditional on a high number of covariates in our dataset. This finding provided support for the MAR assumption and made multiple imputation under the MAR assumption a robust method for assessing the reliability of the primary findings. Second, analysis with imputed outcome data generated very similar results to the ones that were observed in the primary analysis. This led us to conclude with high confidence that the findings of the primary analysis are unlikely to be biased. As a result, further analysis under the MNAR assumption was not necessary.

Treatment Effects in the presence of Non-compliance

A complier average treatment effect (CACE) was estimated to show the effect of Fit to Study on pupils in schools that complied with the assignment to their trial status. In the analysis plan, we indicated that teacher logs would be used as the main source of information for assessing compliance in schools. The developers designed a paper booklet which was distributed to PE teachers in treatment schools. Teachers were asked to complete the log for each PE lesson delivered, noting down the number of minutes of VPA, the length of the warm-up, and the number of fitness infusions implemented. However, only 17 out of 32 treatment schools sent the teacher logs to the developers. Furthermore, 7 of the teacher logs received were not sufficiently complete. As a result, we decided, in conversation with EEF and the Developer, to use the post-intervention school survey for compliance estimation. In the survey, schools were asked to provide an estimate of the % of Y8 PE lessons delivered as intended. Since survey data was retrospective and did not capture the delivery of Fit to Study in each PE lesson, the compliance analysis conducted was less accurate. However, it covered more schools than the analysis based on teacher logs would have done (25 out of 32).

Compliance was defined as the proportion of PE lessons that were delivered as intended. CACE estimates were reported for a range of compliance cut-offs, from 90% compliance to zero compliance. The real CACE estimation was assumed to be generated somewhere between minimal and optimal compliance thresholds (Gerber and Green, 2012).

Given that teachers and pupils in control schools did not have access to the Fit to Study programme, the CACE could be estimated under the assumption of one-sided non-compliance. We labelled as non-compliant any pupils that were selected for testing in treatment schools which did not deliver as many PE lessons with full Fit to Study components as the cut-off value. A complier average treatment effect (CACE) was estimated by dividing the ITT estimate by the share of compliers.

$$CACE = \frac{ITT_y}{\Pr(Compliers)} \quad (1)$$

Subgroup analyses

Three subgroup analyses were carried out, examining whether there was evidence for a differential impact of the intervention on pupils by:

- FSM eligibility
- Gender
- Prior attainment

According to EEF statistical analysis guidance, subgroup analysis by FSM status should always be carried out as FSM pupils are EEF's key target group (EEF statistical analysis guidance, 2018). Also, the level of activity and engagement during PE lessons may vary by gender. Studies report a clear trend of decreasing levels of activity as girls get older, and a widening disparity between girls' and boys' physical activity behaviours (Bailey et al., 2004). Different levels of physical activity among boys and girls may lead to differential impacts of PE interventions on academic attainment. For example, Carlson et al. (2008) identify the relationship between physical activity and learning for girls but not for boys. Moreover, some studies testing the impact of physical interventions on academic achievement found different effect sizes for pupils with low and high prior attainment (McClelland et al., 2015).

Estimation of subgroup effects on the primary outcome involved the re-estimation of the adjusted model described above, with the addition of a further covariate for the particular subgroup concerned. This additional covariate was interacted with the treatment/control group indicator. Additionally, separate models were estimated and reported for each subgroup.

Effect size calculation

Calculation of effect sizes, their variances and 95% confidence intervals was carried out in Stata. Effect sizes were estimated with *esizei* command (for full syntax, see Appendix E). The effect sizes were calculated following Hedges (2007), using the total pooled within-groups standard deviation, and assuming unequal cluster sizes (Hedges, 2007). The effect size was calculated as follows:

$$d = \frac{(\bar{Y}_T - \bar{Y}_C)_{adj}}{\sigma_T}\sqrt{1 - \left(\frac{(N - n_U^T m^T - n_U^C m^C) + n_U^T + n_U^C - 2}{N-2}\right)} \qquad (3)$$

Where $(\bar{Y}_T - \bar{Y}_C)_{adj}$ is the difference between the treatment and control group means taken from the model (1) above, controlling for baseline and stratification variables only. The denominator (σ_T) is calculated as the unconditional variance of model (1), that is, it has no variables entered other than the treatment effect indicator. This calculation assumes meta-analytic inference is aimed at the population of individuals. It is calculated, as follows:

$$\sigma_T = \sqrt{\frac{\sum_{i=1}^{m^T}\sum_{j=1}^{{n_i}^T}\left(Y_{ij}^T - \bar{Y}_{++}^T\right) + \sum_{i=1}^{m^C}\sum_{j=1}^{{n_i}^C}\left(Y_{ij}^C - \bar{Y}_{++}^C\right)}{N-2}} \qquad (4)$$

The second term on the rhs of (3) is calculated using the following:

$$n_U^T = \frac{(N^T)^2 - \sum_{i=1}^{m^T}(n_i^T)^2}{N^T(m^T - 1)}$$

$$n_U^C = \frac{(N^C)^2 - \sum_{i=1}^{m^C}(n_i^C)^2}{N^C(m^C - 1)}$$

$$N = N^T + N^C$$

$$v(d_t) = \left(\frac{N^T + N^C}{N^T N^C}\right)(1 + (\tilde{n} - 1)\rho) + \frac{[(N-2)(1-\rho)^2 + A\rho^2 + 2B\rho(1-\rho)d_T^2]}{2(N-2)[(N-2) - \rho(N-2-B)]} \qquad (5)$$

$$\tilde{n} = \frac{N^C \sum_{i=1}^{m^T} (n_i^T)^2}{N^T N} + \frac{N^T \sum_{i=1}^{m^C} (n_i^C)^2}{N^C N}$$

$$A^T = \frac{(N^T)^2 \sum_{i=1}^{m^T} (n_i^T)^2 + \left(\sum_{i=1}^{m^T} (n_i^T)^2\right)^2 - 2N^T \sum_{i=1}^{n^T} (n_i^T)^3}{(N^T)^2}$$

$$A^C = \frac{(N^C)^2 \sum_{i=1}^{m^C} (n_i^C)^2 + \left(\sum_{i=1}^{m^C} (n_i^C)^2\right)^2 - 2N^C \sum_{i=1}^{n^C} (n_i^C)^3}{(N^C)^2}$$

$$B = n_U^T (m^T - 1) + n_U^C (m^C - 1)$$

$$\rho = \frac{\sigma_B^2}{\sigma_T^2}$$

$$\sigma_B = \sqrt{\frac{\sum_{i=1}^{m^T} \sum_{j=1}^{{n_i}^T} (Y_{i+}^T - \bar{Y}_{*+}^T) + \sum_{i=1}^{m^C} \sum_{j=1}^{{n_i}^C} (Y_{i+}^C - \bar{Y}_{*+}^C)}{m^T + m^C - 2}} \quad (6)$$

We converted to Hedge's g, following the usual formulae:

$$J(df) = 1 - \frac{3}{4df - 1}$$

$$g = j(df)d \quad (7)$$

$$v_g = [J(df)]^2 v_d^T \quad (8)$$

$$SE_g = \sqrt{v_g} \quad (9)$$

Confidence intervals were calculated using the 95% threshold, i.e. ±1.96SE_g.

Implementation and process evaluation

The implementation and process evaluation was designed to explore how Fit to Study was delivered, including the barriers and facilitators to delivery and teacher engagement with support. It aimed to assess implementation fidelity in treatment schools and explore 'business as usual' in control schools. Exploring implementation fidelity is important for furthering understanding of any contextual and experiential considerations that affect implementation, and to inform analysis of intended primary and secondary outcomes along with any unintended outcomes. The key dimensions of implementation that were assessed by the process evaluation were decided at a workshop with the developer and agreed with EEF. They included fidelity, dosage, responsiveness, adaptation, sustainability and scalability.

The main research questions for the implementation and process evaluation from the protocol were expanded to include:

1. How is the programme delivered and what factors influence implementation fidelity?
2. What type of PE lesson modifications take place in treatment schools during the treatment period?
3. What more generally are the barriers and necessary conditions for success?
4. What is PE teachers' level of engagement with available external support and guidance?
5. What type of PE lesson modifications take place in control schools during the treatment period?

The process evaluation was designed as a mixed methods study. The quantitative element involved a pre- and post-intervention school survey. The qualitative element was longitudinal, delivered in three phases. Fieldwork was carried out between September 2017 and June 2018 and included the following components set out in

Table 4. Fieldwork was carried out by members of the NatCen Fit to Study team, as set out in the 'Project team' section.

Table 4: Elements of the process evaluation

Elements of the process evaluation	Components	Aim
Online school surveys	Year 8 PE teachers in all schools (treatment and control) were invited via email to complete a pre- and post-intervention survey. The first survey was conducted in September 2017 and achieved a response rate of 51%. The post-intervention survey was conducted prior to pupil testing and achieved a higher response rate of 78%.	This baseline survey sought to confirm pre-randomisation pupil lists and gather contextual information on how PE was being taught in schools. The post-intervention survey explored teacher perceptions of the intervention in treatment schools and collected cost data. Control schools were asked to provide information on PE delivery.
Phase 1 interviews	20 telephone interviews were conducted with Year 8 PE teachers – 15 in treatment schools and 5 in control schools. These took place during the Autumn term (November – December 2017). Schools were purposively selected to achieve a diverse sample using the following criteria: - School type (mixed, all female) - Size - Geographical location - Type of training received (online or face-to-face) - Sport network status	To explore early implementation of the intervention. Treatment school interviews focused on PE provision in the school, the sign-up process, training experience and early delivery. Control schools were asked about their reasons for signing up and the current focus of their Year 8 PE lessons to describe 'business as usual' PE practice.
Phase 2 school visits	Researchers visited both treatment and control schools to conduct a lesson observation and a face-to-face interview with a Year 8 PE teacher. Lesson observations were conducted using an observation pro forma to capture the key features of the class and lesson, the key elements of the session with regards to the intervention activities and VPA and pupil engagement. Schools were selected from the Phase 1 sample using the same sampling criteria as listed above. In addition to this, schools were also selected based on perceived fidelity (high, medium and low), based on the Phase 1 interviews.	To explore ongoing delivery of Fit to Study in treatment schools. Observations recorded how a Fit to Study lesson was conducted, pupil responsiveness, and any adaptations. The face-to-face interviews were designed to capture reflections on the observed lesson and explore dosage, fidelity, adaptation and responsiveness. In control schools, lessons were observed to record key elements of the lesson and if any Fit to Study 'style' modifications had been made.
Phase 3 interviews	Telephone interviews were conducted with Year 8 teachers in 9 treatment schools during the summer term (June/July 2018). Participants were selected from both the Phase 1 and Phase 2 samples based on emerging evidence.	The final phase aimed to understand how delivery of Fit to Study progressed over time, adaptations made and whether Fit to Study can be sustained and scaled up to include other year groups.

In addition to the above methods we also analysed the lesson logs to measure compliance and an assessment of fidelity and dosage. The developers collected information on the delivery of Fit to Study during Year 8 PE classes in treatment schools throughout the year. this took the form of a paper-based document which teachers were asked to complete after each PE lesson. Teachers returned the completed logs to the developers who then shared log data with NatCen. The developers had initially planned to use an app but the functionality of the app and the technical glitches with the tablet device used meant that this approach to collecting PE lesson data was abandoned before the start of the trial (but after publication of the protocol).

Table 5 below shows the number of planned and achieved interviews and observations during the process study.

Achieved data collection encounters

Corresponding to each element of the primary data collection set out in Table 1 the table below shows the number of interviews and observations we planned to conduct, and the numbers achieved.

Table 5: Total number of qualitative data collection encounters

Type	Theme	Planned qualitative data collection encounters	Achieved qualitative data collection encounters
Phase 1 interviews	Early implementation	20	20
Phase 2 observations	Ongoing delivery	10	9 (one school had stopped delivering Fit to Study so an observation was not appropriate)
Phase 2 interviews	Ongoing delivery	10	10 (including one with a school that had dropped out)
Phase 3 interviews	Fidelity and sustainability	8-10	9 (including six that had taken part in Phase 1 and Phase 2, two that had only taken part in Phase 1 and one that was new to the sample)
Total		48-50	48

Conduct of interviews

The content of each interview was based on a topic guide to ensure systematic coverage of key themes that addressed the process evaluation research questions. This included exploring early implementation, ongoing delivery, fidelity and sustainability at different timepoints throughout the academic year. An example topic guide can be found in the appendix. It was intended to be flexible and interactive, allowing issues of relevance to be covered through detailed follow-up questioning. Separate topic guides were produced for participants from control and treatment schools.

To minimise the research burden on participants, interviews for Phase 1 and Phase 3 were all conducted over the telephone. As part of the Phase 2 site visits, participants were offered the choice of a face-to-face interview on the day of the school visit or a telephone interview at a later date. The interviews were digitally recorded and professionally transcribed. Framework, a systematic approach to qualitative data management and analysis, was used to chart (collate and summarise) and analyse transcribed data. All participants were assured that interview discussions would remain confidential and would be treated in accordance with the Data Protection Act. Phase 3 interviews were conducted in accordance the General Data Protection Regulation (GDPR) after it came into force on 25th May 2018.

SPSS was used to analyse survey data and syntax files were used to ensure a record was kept of how analyses were conducted for quality assurance purposes. Triangulation of data from across the three phases of research was conducted by theme and by school using the Framework approach and cross-referenced with data collected in the post-intervention survey.

Costs

To estimate the per pupil cost to schools of delivering Fit to Study, two sources of data were used.

Information on direct and indirect costs was collected from the Fit to Study development team based on their activities throughout the trial. This included start-up costs associated with training schools to deliver Fit to Study (either face-to-

face or online) and annual running costs associated with providing support for schools throughout the academic year. The approach undertaken was based on EEF guidance[16].

Estimates of time spent by school staff on all Fit to Study related activities (planning and delivery) were collected from intervention schools via the post-intervention online survey. The post-intervention survey achieved a response rate of 78%. Estimates around time spent on the intervention were based on a sub-set of responses from treatment schools who reported running the Fit to Study intervention in at least one Year 8 PE lesson (n=21), and so findings based on the post-intervention survey should be interpreted carefully.

Timeline

A two-stage pilot phase ran from January 2015 to September 2016. At the end of the pilot the intervention design for the trial was finalised by the developers. Planning for the full intervention trial began in September 2016. The intervention was delivered during the 2017/18 school year (September 2017 - June 2018). Randomisation took place in July 2017 and the pre-intervention survey was conducted at the start of the academic year when Fit to Study delivery commenced. The implementation and process evaluation fieldwork was carried out between October 2017 and July 2018. Research activities concluded with post-intervention pupil testing and school survey in June-July 2018. Table 6 sets out the key evaluation milestones. Analysis and reporting were delayed due to changes in the process for accessing data from the NPD data following the introduction of GDPR.

Table 6: Timeline

Date	Activity
January – April 2017	Recruitment of schools
July 2017	Randomisation of schools and Year 8 form groups
September 2017	Fit to Study delivery commenced and pre-intervention survey
October 2017	Process evaluation commenced
June 2018	Fit to study delivery ends
June – July 2018	Post-intervention testing and school survey
September 2018 – April 2019	Analysis and reporting

[16]https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_guidance_to_evaluators_on_cost_evaluation_1.pdf

Impact evaluation

Participant flow including losses and exclusions

NFER was appointed by EEF to recruit schools to the trial. In total 1,348 schools were approached. Of these 1,236 declined to participate or did not respond, 112 agreed to take part in the trial. In total, 104 schools were randomised as two schools dropped out before the randomisation and six schools were excluded. As described in the Methods section, in each randomised school we randomly selected 49% of all Year 8 form groups for testing. In total, 8,707 pupils were selected for testing. This is depicted in Figure 2 and attrition explained in more detail below.

Figure 2: Participant flow diagram

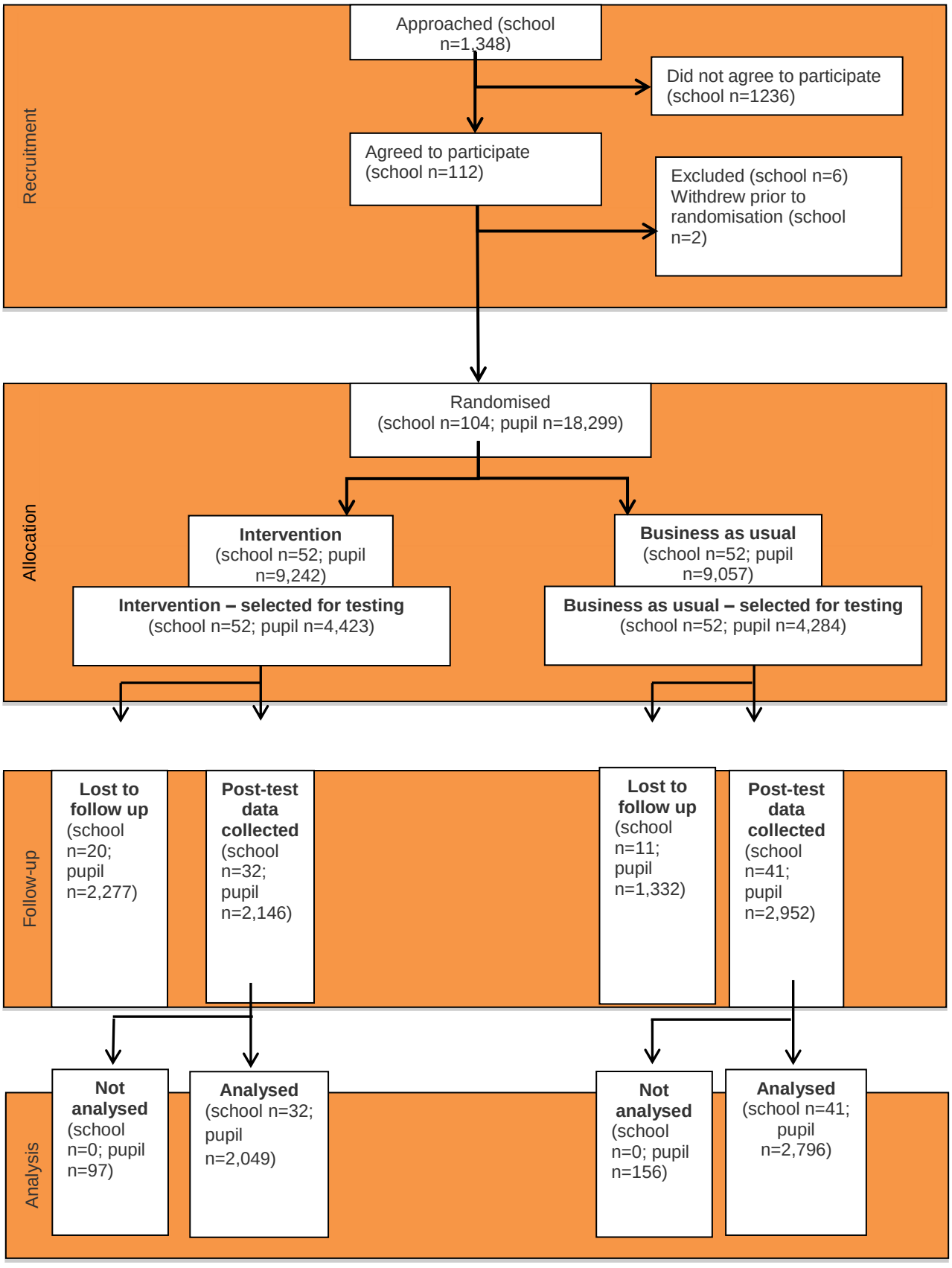


Table 7 sets out the changes in the minimum detectable effect size (MDES) between the protocol, randomisation and analysis stages. At the protocol stage, the MDES was estimated to be around 0.24 standard deviations. This estimation was based on the assumption that 90 schools would be included in the analysis and the intra-cluster correlation (ICC) was 0.16, and baseline covariates explained 50% of variance at pupil level. At randomisation, the MDES decreased to 0.22. This was mainly due to the higher number of randomised schools (104 compared to 90 schools presumed at the protocol stage). While pupils from only 73 schools were included in the final analysis, the MDES remained the same at the analysis stage. This was mainly due to the estimated ICC being lower than originally anticipated (0.11 compared to 0.16). Additionally, the correlation between pre-test and post-test maths scores (0.77) was higher than the one assumed in the randomisation stage (0.63). The MDES for pupils eligible for FSM was only estimated at the analysis stage and was equal to 0.25 standard deviations (please see the table below).

Table 7: Minimum detectable effect size at different stages

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.24	-	0.22	-	0.22	0.25
Pre-test/ post-test correlations	level 1 (pupil)	0.71	-	0.63	-	0.77	0.73
	level 2 (school)	0	-	0.32	-	0.23	0.12
Intra-cluster correlation (ICC)	level 2 (school)	0.16	-	0.16	-	0.11	0.11[17]
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
Average cluster size		90	-	78	-	49	12
Number of schools	intervention	45	-	52	-	32	32
	control	45	-	52	-	41	41
	total	90	-	104	-	73	73
Number of pupils	intervention	4,050	-	4,056	-	2,049	622
	control	4,050	-	4,056	-	2,796	826
	total	8,100	-	8,112	-	4,845	1,448

Attrition

In total, 8,707 pupils from 104 schools were randomly selected for testing. While the number of Year 8 pupils in recruited schools was considerably higher, only those pupils who were randomly selected for testing are formally considered as trial participants. 16 schools with 1,413 pupils dropped out of the trial before scheduling for post-intervention testing started, and another four schools with 268 pupils dropped out during the course of the scheduling process. Five schools with 486 pupils left the trial after scheduling, but before the start of post-intervention testing. Finally, six schools with 520 pupils dropped out during the testing. As a result, testing was conducted in 73 schools with 6,020 pupils.

Schools dropped out before and during testing for a variety of reasons including the time required for facilitating testing, which was considered to be burdensome, clashes with the scheduling of other school activities, such as exams, and the logistical difficulties of bringing together a disparate groups of pupils from across a year group for testing, where

[17] Both overall and FSM ICCs were calculated for the multi-level model containing no covariates (only the intervention allocation variable). Using the primary analysis model (with prior attainment) has almost no effect on the overall ICC (it would still be 0.11) and only a marginal effect on the FSM ICC (it would be 0.10).

form groups had changed from Year 7 to Year 8 (pupil and form lists had been collated pre-randomisation when pupils were in Year 7). Other unforeseen circumstances included flooding in a school.

Due to scheduling and time constraints, some schools did not agree to scheduling a mop-up testing session for pupils absent on the main testing day. Taking into account pupil absences, only 5,290 test results were received. After doing further checking, we found that 158 of these test scores were marked as zero because these pupils did not attend the test. We also found that we had received test scores from 34 pupils who were not selected for testing. These pupils were eliminated from the analysis. This left us with 5,098 pupils with post-test data available.

Our primary analysis model uses baseline scores (Key Stage 2) as a covariate. Due to the absence of baseline scores for 253 tested pupils, our final number of analysed pupils was 4,845. Therefore, the overall rate of pupil-level attrition is 44%. Pupil-level attrition was higher in the treatment group (54%) than in the control group (35%).

Pupil and school characteristics

In expectation, a random allocation procedure should eliminate any meaningful pre-treatment differences between the intervention and control groups. While the possibility of systematic bias is excluded, any given randomisation may still lead to imbalances between the two groups due to random error. In the following tables we check for these imbalances using pupil and school-level characteristics. Four types of imbalances have been assessed: 1) imbalances across randomised pupils; 2) imbalances across analysed pupils; 3) imbalances across randomised schools; 4) imbalances across analysed schools. Randomised sample is composed of pupils who were randomly selected for testing (8,707 pupils from 104 schools). Analysed sample is made up of pupils who were included in the main analysis (4,845 from 73 schools).

Key Stage 2 (KS2) scores for maths were used as a pre-test measure in this trial. Data on the raw KS2 scores was obtained from the National Pupil Database (NPD). Baseline data was missing for 463 randomised pupils (including 253 pupils with available post-test data). As a result, baseline pre-test data for 8,244 pupils was obtained and analysed. Pupils' pre-test scores varied from 2 to 110. The mean pre-test score was 71.9 in the randomised sample; standard deviation – 23.7. More than half of the randomised pupils had a pre-test score of 75 or higher. Histograms detailing the distribution of pre-test scores in the total randomised sample as well as in the treatment and control groups are provided in appendix H. In addition to KS2 scores, a range of other pupil and school-level characteristics were used to check for imbalance between the two trial groups (please see the tables below).

The analysis of randomised sample revealed that, on average, pupils allocated to treatment status had lower pre-test scores (Hedges' g effect size= -0.06), higher number of absence sessions in the preceding academic year (Hedges' g effect size=0.06) and were more likely to be eligible for FSM than control pupils. These differences were statistically significant. There were no statistically significant differences between the treatment and control groups according to gender distribution. It was also found that, on average, treatment pupils belonged to smaller schools (both in terms of the number of classes and pupils) than control pupils. Moreover, the average treatment pupil was enrolled in schools with a higher percentage of absence sessions and a higher share of pupils eligible for FSM than the average control pupil. The full breakdown of the randomised pupil sample by treatment status is provided in

Table 8.

Table 8: Baseline comparison – randomised sample

School-level (categorical)	Intervention group		Control group	
	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Gender status of school Co-educational Single-sex	 3,718/4,423(0) 705/4,423(0)	 3,718(84.1%) 705(15.9%)	 3,568/4,284(0) 716/4,284(0)	 3,568(83.3%) 716(16.7%)
Urban or rural Urban Rural	 3,995/4,423(0) 428/4,423(0)	 3,995(90.3%) 428(9.7%)	 3,898/4,284(0) 386/4,284(0)	 3,898(91%) 386(9%)
School type Academy– Sponsor-led Academy Converter Academy Convertor – Mainstream Community School Foundation School Free School – Mainstream Voluntary aided School Voluntary controlled School	 716/4,423(0) 197/4,423(0) 1,501/4,423(0) 945/4,423(0) 389/4,423(0) 168/4,423(0) 452/4,423(0) 55/4,423(0)	 716(16.2%) 197(4.5%) 1,501(33.9%) 945(21.4%) 389(8.8%) 168(3.8%) 452(10.2%) 55(1.2%)	 721/4,284(0) 82/4,284(0) 1,549/4,284(0) 737/4,284(0) 622/4,284(0) 62/4,284(0) 424/4,284(0) 87/4,284(0)	 721(16.8%) 82(1.9%) 1,549(36.2%) 737(17.2%) 622(14.5%) 62(1.5%) 424(9.9%) 87(2%)
School-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)
Number of Y8 pupils	4,423 (0)	199.9 (59.7)	4,284(0)	208.6 (77.7)
Number of Y8 classes	4,423 (0)	7.2 (2.1)	4,284(0)	7.5 (3)
Percentage of absence sessions in previous academic year (16/17)	4,423(0)	5.5 (1)	4,105(179)	5.3 (0.9)
Percentage of pupils eligible for FSM	4,423(0)	15.5 (8.6)	4,047(237)	14.9 (9.9)
Average KS2 score	-	-	-	-
Pupil-level (categorical)	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Eligible for FSM Eligible Not Eligible	 1,533/4,409(14) 2,876/4,409(14)	 1,533(34.8%) 2,876(65.2%)	 1,374/4,263(21) 2,889/4,263(21)	 1,374(32.2%) 2,889(67.8%)
Gender Boys Girls	 2,007/4,423(0) 2,416/4,423(0)	 2,007(45.4%) 2,416(54.6%)	 2,013/4,284(0) 2,271/4,284(0)	 2,013(47%) 2,271(53%)

Pupil-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Effect Size
Pre-test score (Key Stage 2)	4,216(207)	71.1(23.9)	4,028(256)	72.6(23.5)	-0.06
Number of absence sessions in previous academic year (16/17)	4,422(1)	16.1(21.3)	4,262(22)	15(17.9)	0.06

When assessing the imbalances on pupil-level characteristics in the analysed sample, no statistically significant differences between treatment and control pupils were found. While differences across the two groups did not reach the required level of statistical significance ($p<0.05$), the table provided below shows that treatment pupils had lower average pre-test score (72.8) than control pupils (74). This represents a Hedges' g effect size= -0.05 which is slightly smaller than the difference observed for the sample as randomised. Furthermore, the proportion of boys in the treatment group was lower than in the control group. Statistically significant differences were observed for four school-level variables – number of pupils, number of classes, gender status of school and urban/rural status. As in the case of the randomised sample, analysed treatment pupils were more likely to be enrolled in smaller schools (both in terms of the number of form groups and pupils). However, treatment pupils in the analysed sample were more likely to study in single-sex and rural schools. These imbalances were not observed in the randomised sample. The full breakdown of the analysed pupil sample by treatment status is provided in Table 9.

Table 9: Baseline comparison – analysed sample

School-level (categorical)	Intervention group		Control group	
	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Gender status of school				
Co-educational	1,589/2,049(0)	1,589 (77.5%)	2,280/2,796(0)	2,280 (81.6%)
Single-sex	460/2,049(0)	460 (22.5%)	516/2,796(0)	516 (18.4%)
Urban or rural				
Urban	1,770/2,049 (0)	1,770 (86.4%)	2,561/2,796 (0)	2,561 (91.6%)
Rural	279/2,049 (0)	279 (13.6%)	235/2,796 (0)	235 (8.4%)
School type				
Academy– Sponsor-led	163/2,049 (0)	163 (8%)	413/2,796 (0)	413/ (14.8%)
Academy Converter	100/2,049 (0)	100 (4.9%)	72/2,796 (0)	72/ (2.6%)
Academy Convertor – Mainstream	705/2,049 (0)	707 (34.4%)	1,001/2,796 (0)	1,001/ (35.8%)
Community School	527/2,049 (0)	527 (25.7%)	515/2,796 (0)	515/ (18.4%)
Foundation School	303/2,049 (0)	303 (14.8%)	385/2,796 (0)	385/ (13.8%)
Free School – Mainstream	91/2,049 (0)	91 (4.4%)	33/2,796 (0)	33/ (1.2%)
Voluntary aided School	115/2,049 (0)	115 (5.6%)	308/2,796 (0)	308/ (11%)
Voluntary controlled School	45/2,049 (0)	45 (2.2%)	69/2,796 (0)	69/ (2.4%)
School-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)
Number of Y8 pupils	2,049 (0)	190.2 (54.9)	2,796 (0)	216 (80.5)
Number of Y8 classes	2,049 (0)	6.6 (1.8)	2,796 (0)	7.9 (3.2)
Percentage of absence sessions in previous academic year (16/17)	2,049 (0)	5.2 (1)	2,653 (143)	5.2 (0.84)
Percentage of pupils eligible for FSM	2,049 (0)	14.1 (9.3)	2,597 (199)	14 (9.6)
Average KS2 score	-	-	-	-
Pupil-level (categorical)	n/N (missing)	Count (%)	n/N (missing)	Count (%)

Eligible for FSM Eligible Not Eligible	 622/2,049(0) 1,427/2,049(0)	 622 (30.4%) 1,427 (69.6%)	 826/2,792 (4) 1,966/2,792 (4)	 826 (29.6%) 1,966 (70.4%)	
Gender Boys Girls	 832/2,049 (0) 1,217/2,049 (0)	 832 (40.6%) 1,217 (59.4%)	 1,202/2,796 (0) 1,594 /2,796 (0)	 1,202 (43%) 1,594 (57%)	

Pupil-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Effect Size
Pre-test score (Key Stage 2)	2,049 (0)	72.8 (23.4)	2,796 (0)	74 (23.1)	-0.05
Number of absence sessions in previous academic year (16/17)	2,049 (0)	13.1(15.1)	2,791 (5)	13.1(14.4)	0.002

No statistically significant differences between the randomised treatment and control schools were found. Overall, a high degree of balance was observed between the schools that were randomised to treatment and control, meaning that schools in the two treatment arms shared similar characteristics at baseline. Data on the balance between treatment and control schools in the randomised sample is provided in Table 10.

Table 10: Baseline comparison – imbalances across randomised schools

School-level (categorical)	Intervention group		Control group	
	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Gender status of school Co-educational Single-sex	 42/52(0) 10/52(0)	 42 (80.8%) 10 (19.2%)	 42/52(0) 10/52(0)	 42 (80.8%) 10 (19.2%)
Urban or rural Urban Rural	 47/52 (0) 5/52 (0)	 47 (90.4%) 5 (9.6%)	 47/52 (0) 5/52 (0)	 47 (90.4%) 5 (9.6%)
School type Academies and free schools	 31/52 (0)	 31 (60%)	 30/52 (0)	 30 (58%)
Community and foundation schools	15/52 (0)	15 (29%)	14/52 (0)	14 (27%)
Voluntary aided and voluntary controlled schools	6/52 (0)	6 (11%)	8/52 (0)	8 (15%)
School-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)
Number of Y8 pupils	52 (0)	177.8 (61.2)	52 (0)	174.3 (74.9)
Number of Y8 classes	52 (0)	6.8 (2.1)	52 (0)	6.8 (3)
Percentage of absence sessions in previous academic year (16/17)	52 (0)	5.5 (1)	51 (1)	5.4 (0.91)
Percentage of pupils eligible for FSM	52 (0)	16.8 (9.7)	50 (2)	15.6 (10.4)
Average KS2 score	-	-	-	-

No statistically significant differences between the analysed treatment and control schools were found. While not statistically significant, some differences were observed in relation to the size of schools in the two treatment arms. On average, analysed treatment schools were smaller than control schools both in terms of the number of pupils and classes. There was a high degree of balance according to all other characteristics included in the comparison. Data on the balance between analysed schools is provided in

Table 11.

Table 11: Baseline comparison – imbalances across analysed schools

School-level (categorical)	Intervention group		Control group	
	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Gender status of school Co-educational Single-sex	 25/32(0) 7/32(0)	 25 (80.5%) 7 (19.5%)	 33/41(0) 8/41(0)	 33 (78.1%) 8 (21.9%)
Urban or rural Urban Rural	 28/32 (0) 4/32 (0)	 28 (87.5%) 4 (12.5%)	 37/41 (0) 4/41 (0)	 37 (90.2%) 4 (9.8%)
School type Academies and free schools Community and foundation schools Voluntary aided and voluntary controlled schools	 17/32 (0) 12/32 (0) 3/32 (0)	 17 (53%) 12 (38%) 3 (9%)	 23/41 (0) 11/41 (0) 7/41 (0)	 23 (56%) 11 (27%) 7 (17%)
School-level (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)
Number of Y8 pupils	32 (0)	167 (57.7)	41 (0)	179.6 (79.5)
Number of Y8 classes	32 (0)	6.3 (1.8)	41 (0)	7.1 (3.3)
Percentage of absence sessions in previous academic year (16/17)	32 (0)	5.4 (1.1)	40 (1)	5.3 (0.87)
Percentage of pupils eligible for FSM	32 (0)	16 (11)	39 (2)	15.5 (10.4)
Average KS2 score	-	-	-	-

Overall, the observed imbalances are small and can be attributed to random error. To explore whether the imbalances had any effect on the findings, we ran sensitivity analysis which included unbalanced covariates in addition to those in the main model. The results of this sensitivity analyses are presented immediately after the primary analysis.

Outcomes and analysis

The primary outcome was maths attainment, as measured by the raw scores achieved in the Progress Test in Mathematics (PTM), Level 13 (GL Assessment 2015). No other outcome measures were used. Only 5,098 pupils were tested after the intervention (58.6% of the randomised sample), which was mainly due the high number of schools dropping out of the trial (please see the section on attrition). Pupils' post-test scores varied from 0 to 69. The mean post-test score was 30.4; standard deviation was 15.1. Histograms detailing the distribution of post-test scores in the randomised sample as well as in the treatment and control groups are provided in appendix I.

The primary analysis model is a two-level model in which pupils are clustered in schools. Schools were modelled as random effects using a random intercept model. Apart from the dummy variable capturing treatment/control group membership, two other variables were used as covariates – the stratification variable (gender-status of school) and

pupil-level baseline test scores. Baseline test scores were centred on the mean[18]. Due to the absence of baseline scores for 253 tested pupils, our final number of analysed pupils was 4,845.

The estimated Hedges' g effect size is negative at -0.008 (-0.06, 0.05) and is considerably smaller than the MDES of 0.24. This means that the statistical evidence does not meet the threshold set by the evaluator to conclude the true impact was non-zero. Therefore, it cannot be claimed that the Fit to Study intervention had an impact on the maths attainment of participating pupils. Year 8 pupils in Fit to Study schools did not make additional progress in maths compared to Year 8 pupils in 'business as usual' control schools when assessed at the end of the intervention. The pooled unconditional standard deviation of the outcome was used in the calculation of effect sizes (46.3). The intra-cluster correlation (ICC) was estimated at 0.11, indicating that 11% of the total variance in outcome scores is between schools. The main results of the primary analysis are presented in Table 12 and Table 13.

Table 12: Primary analysis

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Post-test in maths	2,049 (2,374)	30.2 (29.6, 30.8)	2,796 (1,488)	31.1 (30.5, 31.6)	4,845 (2,049; 2,796)	-0.008 (-0.06, 0.05)	0.661

Table 13: Effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance	Population variance (if available)
			n (missing)	Variance of outcome	n (missing)	Variance of outcome		
Post-test in maths	-0.9	-0.34	2,049 (2,374)	40.9	2,796 (1,488)	46.3	44.1	-

Sensitivity analysis

In addition to the primary analysis, three types of sensitivity analyses were listed in the analysis plan: 1) unadjusted analysis that does not include any baseline covariates; 2) analysis which includes the unbalanced covariates in addition to those in the main model; 3) adjusted analysis that includes all the variables that are likely to be predictive of the outcome. The details of which covariates were included in each model are presented in the table below. It is important to note that all continuous variables were centred on the mean.

[18] The current EEF Statistical Analysis Guidance (2018) suggests using raw baseline scores unless there are clear reasons to transform the data (p. 3). The SAP of this evaluation was agreed prior to the finalisation of the new Guidance.

Table 14: Differences in covariates used for each model

Covariates included	Unadjusted model	Main model	Model with unbalanced variables	Model with all potential predictors
Pre-test score (Key Stage 2)	No	Yes	Yes	Yes
Number of absence sessions in previous academic year (16/17)	No	No	Yes	Yes
Eligibility for FSM	No	No	Yes	Yes
Gender	No	No	No	Yes
Gender status of school	No	Yes	Yes	Yes
Urban or rural	No	No	No	Yes
Type of school	No	No	No	Yes
Number of Y8 pupils	No	No	Yes	Yes
Number of Y8 classes	No	No	Yes	Yes
Percentage of absence sessions in previous academic year (16/17)	No	No	Yes	Yes
Percentage of pupils eligible for FSM	No	No	Yes	Yes

Sensitivity analyses indicated that using no covariates or adding additional ones had only a modest impact on the effect sizes. For all sensitivity analyses, as with the main model, the statistical evidence did not meet the threshold set by the evaluator to conclude the true impact was non-zero. In all three sensitivity analyses specified, Hedges' g effect sizes remained considerably smaller than the MDES of 0.24. The strongest negative effect size was estimated in the unadjusted model (-0.02) which is expected given the baseline imbalance observed in pre-tests. However, the confidence intervals indicated that the null hypothesis could not be rejected (-0.08, 0.04). When the model was adjusted for unbalanced predictors or for all potential predictors, the effect sizes were close to zero and equal to -0.004 and -0.007 respectively. Therefore, the findings of the primary analysis were not sensitive to the addition of covariates. The results of the sensitivity analyses are provided in Table 15 and Table 16.

Table 15: Sensitivity analyses

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Post-test in maths (unadjusted model)	2,146 (2,277)	29.8 (29.2, 30.5)	2,952 (1,332)	30.7 (30.2, 31.3)	5,098 (2,146; 2,952)	-0.02 (-0.08, 0.04)	0.465
Post-test in maths (model with unbalanced covariates)	2,049 (2,374)	30.2 (29.6, 30.8)	2,589 (1,695)	30.8 (30.2, 31.4)	4,638 (2,049; 2,589)	-0.004 (-0.06, 0.05)	0.780
Post-test in maths (model	2,049 (2,374)	30.2 (29.6, 30.8)	2,589 (1,695)	30.8 (30.2, 31.4)	4,638 (2,049; 2,589)	-0.007 (-0.06, 0.05)	0.679

with all potential predictors)							

Table 16: Sensitivity analyses: effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance	Population variance (if available)
			n (missing)	Variance of outcome	n (missing)	Variance of outcome		
Post-test in maths (unadjusted model)	-0.9	-0.9	2,146 (2,277)	42.6	2,952 (1,332)	48	45.8	-
Post-test in maths (model with unbalanced covariates)	-0.6	-0.19	2,049 (2,374)	40.9	2,589 (1,695)	44.8	43.2	-
Post-test in maths (model with all potential predictors)	-0.6	-0.28	2,049 (2,374)	40.9	2,589 (1,695)	44.8	43.2	-

Subgroup analysis

Subgroup analyses were carried out to explore whether there was any evidence for a differential impact of the intervention on pupils by FSM eligibility, gender and prior attainment. Following the analysis plan and based on EEF guidance, the interaction terms of the treatment status and the subgroup indicators were used. We explored whether the coefficients resulting from these interactions reached statistical significance at the 95 per cent level and estimated separate models for each subgroup. These models were specified in the same way as the main model.

To test whether the treatment effects differed according to FSM status, we ran the multi-level model which included all the variables from the main model, FSM eligibility and the interaction term of the treatment status and FSM eligibility. The coefficient resulting from this interaction was negative and not statistically significant (b=-0.19, p=0.749). We also ran two separate models for the two groups of interest – pupils eligible for FSM and those who were not. The results of these models are provided in Table 17 and Table 18. No important differences in effect size by FSM eligibility were found. For both groups of pupils, the effect sizes are very small and of negative sign, much lower than the MDES of 0.24.

Table 17: Subgroup analysis by FSM status

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Post-test in maths for those pupils who are eligible for FSM	622 (911)	25.8 (24.7, 26.9)	826 (548)	25.4 (24.5, 26.4)	1,448 (622, 826)	-0.01 (-0.12, 0.09)	0.709
Post-test in maths for those who are not eligible for FSM	1,427 (1,449)	32.1 (31.3, 32.9)	1,966 (923)	33.4 (32.8, 34.1)	3,393 (1,427, 1,966)	-.006 (-0.07, 0.06)	0.789

Table 18: Subgroup analysis by FSM status: effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance	Population variance (if available)
			n (missing)	Variance of outcome	n (missing)	Variance of outcome		
Post-test in maths for those pupils who are eligible for FSM	0.4	-0.34	622 (911)	27.2	826 (548)	24.6	25.8	-
Post-test in maths for those who are not eligible for FSM	-1.3	-0.21	1,427 (1,449)	32.3	1,966 (923)	39.5	36.6	-

Similarly, a multi-level model featuring the interaction term of the treatment status and the pre-test score (KS2) was run. The pre-test score was factored in as a dichotomous variable identifying pupils with higher and lower than median pre-test scores. The coefficient of the interaction term was not statistically significant (b=0.05, p=0.925). Two separate models were run for pupils with higher and lower than median prior attainment. In the group of pupils with higher than median prior attainment a positive treatment effect was observed but its magnitude was negligible and equally consistent with a very small positive or negative effect (Hedges' g=0.0005, CI= (-0.08,0.08)). While among the pupils with lower than median prior attainment the effect size was negative (Hedges' g= -0.02, CI=(-0.1, 0.07)), the confidence intervals are consistent both with a very small positive effect and a very small negative effect. Therefore, no differential effect by prior attainment was found. The results of the sub-group analysis by prior attainment are provided in Table 19 and

Table 20.

Table 19: Subgroup analysis by prior attainment

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Post-test in maths for those pupils who have	1,049 (1,188)	39.9 (39.1, 40.6)	1,484 (822)	40.2 (39.6, 40.9)	2,533 (1,049, 1,484)	0.0005 (-0.08, 0.08)	0.986

higher than median prior attainment							
Post-test in maths for those pupils who have lower than median prior attainment	1,000	20.1 (19.5, 20.7)	1,312	20.7 (20.1, 21.2)	2,312 (1,000, 1,312)	-0.02 (-0.1, 0.07)	0.667

Table 20: Subgroup analysis by prior attainment: effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance	Population variance (if available)
			n (missing)	Variance of outcome	n (missing)	Variance of outcome		
Post-test in maths for those pupils who have higher than median prior attainment	-0.3	0.02	1,049 (1,188)	24	1,484 (822)	31.8	28.8	-
Post-test in maths for those pupils who have lower than median prior attainment	-0.6	-0.32	1,000 (1,186)	17.1	1,312 (666)	21.2	19.5	-

As a third type of subgroup analysis, a multi-level model featuring the interaction term of the treatment status and gender was run. The coefficient resulting from this interaction was negative and not statistically significant (-0.035, p=0.951). When multi-level models were run separately for boys and girls, the observed effect size was positive for boys (Hedges' g=0.00055, CI=(-0.09; 0.09)) and negative for girls (Hedges' g=-0.014, CI=(-0.088; 0.06)). However, in both cases the point estimates are negligible and confidence intervals are consistent with very small negative and positive effects, and this do not allow to conclude that the true impact was non-zero. Therefore, there is no statistical evidence that treatment effects are different for boys and girls. The results of the subgroup analysis by gender are provided in Table 21 and Table 22.

Table 21: Subgroup analysis by gender

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value

Post-test in maths for girls	1,217 (1,199)	30.6 (29.7; 31.4)	1,594 (677)	31.2 (30.5; 31.9)	2,811	-0.014 (-0.088; 0.06)	0.533
Post-test in maths for boys	832 (1,175)	29.7 (28.7; 30.7)	1,202 (811)	30.9 (30; 31.7)	2,034	0.00055 (-0.09; 0.09)	0.986

Table 22: Subgroup analysis by gender: effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance	Population variance (if available)
			n (missing)	Variance of outcome	n (missing)	Variance of outcome		
Post-test in maths for girls	-0.6	-0.5	1,217 (1,199)	35.3	1,594 (677)	36.6	36	-
Post-test in maths for boys	-1.2	0.018	832 (1,175)	28.2	1,202 (811)	36.6	33.4	-

Compliance analysis

A complier average treatment effect (CACE) was estimated to show the effect of Fit to Study on pupils in schools that comply with the assignment to their trial status. Given that teachers and pupils in control schools did not have access to the Fit to Study intervention, the CACE was estimated under the assumption of one-sided non-compliance. The CACE was arrived at by dividing the ITT estimate by the share of compliers.

In the analysis plan, we indicated that teacher logs would be used as the main source of information for assessing compliance in schools. The main advantage of teacher logs was the expected precision on VPA during each PE lesson as reported by teachers. Provided that the teacher logs were sufficiently complete, compliance could have been defined with reference to the proportion of VPA sessions actually delivered, out of the number of sessions that should have been delivered. However, only 17 out of the 32 treatment schools that did not drop out of the trial submitted teacher logs to the developer. Out of the 17 teacher logs that were provided, 7 were not sufficiently complete as they contained information on less than half of the PE lessons that should have taken place in these schools during the 2017/2018 academic year. Due to this high level of missing data, teacher logs from only 10 out of 32 treatment schools could have been used for compliance analysis. As compliance estimates would have been unreliable, teacher logs were not used for analysis.

The second potential source of information for compliance estimation was the post-intervention school survey. In the survey, schools were asked to provide the estimate of the percentage of Y8 PE lessons delivered as intended. Since not all treatment schools provided this information in the survey, we contacted these ‘non-response’ schools by email to try and capture this information. While data from the post-intervention survey is retrospective and not as accurate as the teacher logs, it allowed us to estimate compliance for a larger number of schools. In total, 25 out of the 32 treatment schools that did not drop out of the trial provided intervention fidelity data. We decided to use this type of data for compliance estimation as it covered a much larger number of schools than the data from teacher logs.

For the post-intervention survey data, compliance was defined as the proportion of PE lessons that included Fit to Study delivery as intended. CACE estimates were reported for five different compliance cut-offs: 90%, 75%, 50%, 25% and 10%. For example, the use of 90% cut-off implies that only those schools which delivered 90% or more PE lessons that included Fit to Study delivery as intended were considered as compliant with the intervention. All pupils in compliant schools were treated as compliant. In all five cases, an assumption was made that the seven schools that did not provide data on intervention fidelity were not compliant with the intervention. Table 23 provides the proportions of compliant pupils based on different compliance cut-offs and

Table 24 provides the CACE estimates based on different compliance cut-offs.

Table 23: Share of compliant pupils based on different compliance cut-offs

Compliance at 90% cut-off	Compliance at 75% cut-off	Compliance at 50% cut-off	Compliance at 25% cut-off	Compliance at 10% cut-off
6%	31%	58%	71%	76%

Table 24: CACE estimates based on different compliance cut-offs

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Post-test in maths (90% cut-off)	2,049 (2,374)	30.2 (29.6, 30.8)	2,796 (1,488)	31.1 (30.5, 31.6)	4,845 (2,049; 2,796)	-0.13 (-0.18, -0.07)	0.661
Post-test in maths (75% cut-off)	2,049 (2,374)	30.2 (29.6, 30.8)	2,796 (1,488)	31.1 (30.5, 31.6)	4,845 (2,049; 2,796)	-0.03 (-0.08, 0.03)	0.661
Post-test in maths (50% cut-off)	2,049 (2,374)	30.2 (29.6, 30.8)	2,796 (1,488)	31.1 (30.5, 31.6)	4,845 (2,049; 2,796)	-0.01 (-0.07, 0.04)	0.661
Post-test in maths (25% cut-off)	2,049 (2,374)	30.2 (29.6, 30.8)	2,796 (1,488)	31.1 (30.5, 31.6)	4,845 (2,049; 2,796)	-0.01 (-0.07, 0.05)	0.661
Post-test in maths (10% cut-off)	2,049 (2,374)	30.2 (29.6, 30.8)	2,796 (1,488)	31.1 (30.5, 31.6)	4,845 (2,049; 2,796)	-0.01 (-0.07, 0.05)	0.661

CACE estimates fluctuate between -0.13 and -0.01 depending on the compliance cut-off used. The real CACE estimation is assumed to be generated somewhere between minimal and optimal compliance thresholds. If the medium compliance threshold (50%) is used, CACE is equal to -0.03 (-0.08, 0.03). However, in all cases the estimated effect sizes are lower than the MDES of 0.24. Therefore, the statistical evidence does not meet the threshold needed to reject the null hypothesis. As a result, it cannot be claimed that Fit to Study had an impact on the maths attainment of pupils that benefitted from the intervention.

Missing data

Out of 8,707 pupils who were selected for post-intervention testing, only 5,098 took the test (loss-to-follow up rate of 41.4%). Our primary analysis assumed that any missing outcome data were missing completely at random (MCAR) and used complete case analysis. However, in the analysis plan we also indicated that we would check the sensitivity of our results to alternative assumptions about the types of missing data.

Firstly, we explored the likely reasons for missing outcome data. We ran a series of t-tests and Fisher's exact tests to check if the absence of outcome data was related to any of the variables in our dataset. When comparing tested pupils with those who were not tested, statistically significant differences with regards to the following variables were found: treatment status, gender, FSM status, pre-test score, number of absence sessions in previous academic year, urban/rural status of school, type of school, gender status of school, percentage of absence sessions in previous academic year (school-level), percentage of pupils eligible for FSM (school-level).

Compared to tested pupils, pupils with missing outcome data were more likely to be treated, to be boys, and to be eligible for FSM. They were also more likely to have had a higher number of absence sessions in the previous academic year and a lower prior attainment score. In addition, pupils with missing outcome data were more likely to study in urban

co-educational schools, schools with a higher percentage of absence sessions in the previous academic year and from schools with a higher percentage of pupils eligible for FSM.

We included all the variables listed above into a multivariate logistic regression model and estimated propensity scores. Important differences in the average propensity scores were found between pupils who were tested and those who were not. In the group of pupils with outcome data, the average probability of outcome data being missing was 0.35 and among the pupils that were not tested, the average propensity score was 0.51. These findings support the missing at random (MAR) assumption (EEF statistical analysis guidance, 2018).

The preliminary analysis described above confirmed that the loss-to-follow up rate can be predicted using existing covariates. We therefore opted to use multiple imputation to impute missing outcome data under the MAR assumption. We imputed 40 new datasets using a two-level linear model. In our imputation model, we included all variables that were part of the primary analysis, any variables predictive of the absence of outcome data, and any variables associated with the outcome. Alongside the nine variables that were found to be related to the absence of outcome data, the number of classes and pupils were also included in the imputation model. Outcome data could only be imputed in those cases when the data on all the variables included in the imputation model was available. In total, outcome data was imputed for 3,361 pupils.

Finally, we ran the main model using each of the 40 imputed datasets. Schools were modelled as random effects using a random intercept model. Apart from the dummy variable capturing treatment/control group membership, two other variables were used as covariates – the stratification variable (gender-status of school) and pupil-level baseline test scores (centred on the mean). Rubin's rules were used to combine the estimates from multiple datasets into an overall MI estimate. The findings of the analysis with imputed data are provided in Table 25 and Table 26 below. The estimated effect size is very small (-0.01), indicating that the analysis with imputed data does not change the conclusions drawn from the primary analysis. Since the analysis with imputed outcome data generated very similar results to the ones that were observed in the primary analysis, further analysis under the MNAR assumption was deemed unnecessary.

Table 25: Analysis with imputed data

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
Post-test in maths	4,213 (210)	28.7	3,788 (496)	29.8	8,001 (4,213; 3,788)	-0.01 (-0.05, 0.03)	0.392

Table 26: Analysis with imputed data: effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance	Population variance (if available)
			n (missing)	Variance of outcome	n (missing)	Variance of outcome		
Post-test in maths	-1.05	-0.53	4,213 (210)	42.2	3,788 (496)	44.5	43.3	-

Cost

Fit to Study was free to schools in the context of this trial, where delivery was funded by EEF. This section estimates how much it would cost schools to deliver Fit to Study outside of this trial context.

Our estimate of the cost of a school delivering Fit to Study includes:

- Start-up costs associated with training schools to deliver Fit to Study (either face-to-face or online)
- Annual running costs associated with providing support for schools throughout the academic year.

Table 27 shows the cost of delivering Fit to Study per pupil over 3 years should a school opt for face-to-face training.

Table 28 shows the cost of delivering Fit to Study per pupil over 3 years should the school attend online training.

Our estimates assume the following:

- The number of schools delivering Fit to Study per year is 52 (exactly as in the trial).
- The average number of Year 8 pupils per school is 178 (exactly as in the trial).
- Each school has one PE teacher attend the training (online or face-to-face) once in the three years, who cascades training to others (see IPE findings below).
- Each training session (online or face-to-face) is attended by 10 teachers (in line with the trial)
- Materials created for the trial are now pre-existing resources and would not factor into the per pupil cost for the implementation of Fit to Study outside of a trial. This includes training materials (videos, PowerPoint slides and script) and the Fit to Study website.
- Face-to-face training would take place at two regional hubs in Oxford and Birmingham.

School staff reported that they did not require any additional materials to deliver Fit to Study. The largest costs to schools therefore arise from Fit to Study staff time – for setting up and delivering training and providing ongoing face-to-face and virtual support to schools.

The mode of training makes a moderate difference to the overall cost of delivering Fit to Study. Delivering face-to-face training brings additional costs through Fit to Study staff and teachers' travel costs and venue and refreshment costs. Delivery of Fit to Study with face-to-face training would result in a per pupil cost of £4.80 over 3 years of delivery. Delivery of Fit to Study with online training would result in a per pupil cost of £4.56 over 3 years of delivery.

Table 27: Cost of delivering Fit to Study with face-to-face training

Cost item	Cost per	Cost in year 1	Total cost over 3 years	Average annual cost	Average annual cost per pupil
Start-up costs (year 1 only): Fit to Study training					
Fit to Study staff time to set up training	Training	£700	£700	£233	£0.13
Fit to Study staff time to deliver training	Training	£1,000	£1,000	£333	£0.19
Fit to Study staff travel costs	Training	£120	£120	£40	£0.02
Venue and refreshments	Training	£500	£500	£167	£0.09
Printing training materials	School	£5	£5	£2	£0.01
Teacher travel costs	School	£20	£20	£7	£0.04
Running costs (annual): Ongoing support and resources					
Fit to Study staff time to visit schools and provide telephone support	School	£700	£2,100	£700	£3.93

Fit to Study staff time to update website	Year	£1,400	£4,200	£1,400	£0.15
Fit to Study staff time to develop and distribute termly e-newsletter	Year	£1,050	£3,150	£1,050	£0.11
Website hosting	Year	£150	£450	£150	£0.02
Website maintenance	Year	£1,000	£3,000	£1,000	£0.11
Total cost per pupil per year over three years			**£4.80**		

Table 28: Cost of delivering Fit to Study with online training

Cost item	Cost per	Cost in year 1	Total cost over 3 years	Average annual cost	Average annual cost per pupil
Start-up costs (year 1 only): Fit to Study training					
Fit to Study staff time to set up training	Training	£700	£700	£233	£0.13
Fit to Study staff time to deliver training	Training	£500	£500	£167	£0.09
Fit to Study staff travel costs	Training	£0	£0	£0	£0.00
Venue and refreshments	Training	£0	£0	£0	£0.00
Printing training materials	School	£5	£5	£2	£0.01
Teacher travel costs	School	£0	£0	£0	£0.00
Running costs (annual): Ongoing support and resources					
Fit to Study staff time to visit schools and provide telephone support	School	£700	£2,100	£700	£3.93
Fit to Study staff time to update website	Year	£1,400	£4,200	£1,400	£0.15
Fit to Study staff time to develop and distribute termly e-newsletter	Year	£1,050	£3,150	£1,050	£0.11
Website hosting	Year	£150	£450	£150	£0.02
Website maintenance	Year	£1,000	£3,000	£1,000	£0.11

Total annual cost per pupil over three years	**£4.56**

Costs over time

Most of the costs of Fit to Study come from ongoing support for schools, which would be the same each year. Nevertheless, the costs of delivering Fit to Study would reduce somewhat over time, since schools would only need to attend for training in the first year. This reduction over time would be most noticeable for schools attending the more expensive face-to-face training. Costs in Years 2 and 3 are identical for schools attending face-to-face and online training.

Table 29: Per-pupil costs of Fit to Study with face-to-face training over 3 years

Year 1	Year 2	Year 3	Average over 3 years
£5.77	£4.32	£4.32	£4.80

and

Table 30 show that the per-pupil cost would decrease from £5.77 for schools attending face-to-face training and £5.08 for schools attending online training in Year 1 to £4.32 in the following two years.

Year 1	Year 2	Year 3	Average over 3 years
£5.02	£4.32	£4.32	£4.56

Table 29: Per-pupil costs of Fit to Study with face-to-face training over 3 years

Year 1	Year 2	Year 3	Average over 3 years
£5.77	£4.32	£4.32	£4.80

Table 30: Per-pupil costs of Fit to Study with online training over 3 years

Year 1	Year 2	Year 3	Average over 3 years
£5.02	£4.32	£4.32	£4.56

School staff time

Training

Information on the number of staff who attended training was collected from schools in the end of the year survey and from developers via registers of attendance.

Both the school survey and the attendance registers showed that, on average, schools who took part in face-to-face training sent two members of staff to attend (survey: mean = 1.9, n=10, registers: mean = 2.3, n=14). The process study found that these were Heads of PE and/or PE teachers who would be delivering the intervention.

Schools were not explicitly asked to report the number of hours that staff spent attending training. However, we assume that attending the training would require a whole day of a staff member's time, since schools may have to travel some way to a regional hub in order to attend. We therefore conservatively estimate **two days** of PE staff time (most likely one day of a Department Head's time and one day of PE teacher time) per school to be required for face-to-face training.

Eleven schools that took part in online training completed the end of year survey. These schools reported that an average of four staff took part in a livestream training session or watched a pre-recorded training video (mean=3.8, n=11). We assume that for schools choosing online training, there would be **minimal additional staff time** needed for attending training since teachers could watch the video in a planning meeting or similar departmental session or in addition to a normal school day.

Delivery

The end of year school survey also asked about PE staff time spent on the implementation of Fit to Study. This is additional time spent on Fit to Study (for instance including planning and preparation), and does not include time spent delivering PE lessons since this is staff time that would be required under business as usual.

Table 31 shows that time spent on delivering Fit to Study was minimal. Across all schools, the average was 14 hours, or around two days. However, the large majority (67%) of school leads reported that PE staff spent no more than 5 hours on the implementation of Fit to Study throughout the academic year, with more than half (57%) reporting that PE staff spent between just one and three hours on delivery. This suggests that most schools would require one or two additional days to plan the activities necessary to deliver Fit to Study, but a small group reported requiring a considerably higher number of hours to undertake these activities. However, given the small sample included here, these results should be interpreted with caution. Of the two schools who reported delivering Fit to Study activities in all PE lessons in the survey, an average of 2 hours was reported to be spent on preparing for the delivery of the Fit to Study intervention.

Table 31: Number of hours spent on delivery over the year by staff in treatment schools

Hours	n.	%
1-3	12	57%
4-5	2	10%
6-10	2	10%
11-50	3	14%
51-100	2	10%
Total	*21*	*100%*

The survey asked whether PE staff spent time on individual planning, joint planning and/or promoting Fit to Study throughout the school. Table 32: Activities teachers spent time on in treatment schools shows that most common, reported in 95% of schools, was joint planning – for instance discussing Fit to Study in PE department meetings.

Table 32: Activities teachers spent time on in treatment schools

Activities	n.	%
Joint planning – e.g. discussing Fit to Study in PE department meetings	20	95%
Individual planning – e.g. re-designing lesson plans to accommodate Fit to Study	12	57%
Promoting Fit to Study throughout the school – e.g. in assemblies or all staff meetings	12	57%
Base	**21**	**100%**

Around a quarter of school leads also reported that non-PE staff spent time on Fit to Study. Other staff listed included members of the Senior Leadership Team, PHSE and form tutors and exam officers. However, where school leads specified tasks that these staff had been involved with, they mentioned research tasks rather than tasks associated with the delivery of the Fit to Study intervention. This finding should thus be interpreted carefully as it may not be applicable to programme delivery outside of a trial.

Implementation and process evaluation

This section synthesises the findings from the process evaluation, bringing together the perspectives of teachers who participated in the evaluation research through interviews, observations and data collected in the post-intervention survey. This includes Year 8 PE teachers from intervention schools who shared their experiences of implementing Fit to Study. It sets out the key issues related to implementation, the extent to which the intervention was delivered as intended, perceived outcomes, suggestions for improvement and control group activity.

The implementation and process evaluation was designed to explore how Fit to Study was delivered, including the barriers and facilitators to delivery and teacher engagement with support. It aimed to assess implementation fidelity in treatment schools and explore 'business as usual' in control schools. The key dimensions of implementation that were assessed by the process evaluation included fidelity, dosage, responsiveness, adaptation, sustainability and scalability.

The main research questions were:

1. How is the programme delivered and what factors influence implementation fidelity?
2. What type of PE lesson modifications take place in treatment schools during the treatment period?
3. What more generally are the barriers and necessary conditions for success?
4. What is PE teachers' level of engagement with available external support and guidance?
5. What type of PE lesson modifications take place in control schools during the treatment period?

Programme delivery and fidelity

This section on programme implementation and fidelity discusses:

- Attractiveness of the intervention
- Training
- Planning for and discussions about the intervention
- Ongoing support and resources

Attractiveness of the intervention

Overall, the Fit to Study intervention was attractive to PE teachers, who felt strongly aligned with its core aims. Many PE teachers and leads were motivated to sign up to and take part in the trial because of their interest in the research results. They were keen to be part of a research project that might provide evidence of their belief that that increasing levels of moderate to physical activity (MVPA) could improve attainment, and therefore demonstrate the importance of PE and sport. As well as demonstrating the importance of PE to a wider audience, teachers were also motivated to take part in the Fit to Study intervention due to a concern about the current status of PE in their schools and a desire to prove its value and place in the curriculum:

> *"One of the main motivating factors, I suppose, was to highlight the importance of PE potentially in wider school provision."* (Year 8 PE teacher,1017)

PE teachers also expressed that Fit to Study lined up with their existing professional interest in improving fitness levels in young people and combating obesity throughout their school.

The financial incentive offered to schools taking part in the trial also played a part in the decision for schools to sign up to the intervention, especially in the context of a perceived reduction in school budgets. This incentive would not be present if the trial were to be rolled out outside of a trial context.

Training

Training was provided by the developers to inform PE teachers of the theory behind the Fit to Study intervention and to go through the practical aspects of delivering the intervention. Attendees were shown a PowerPoint presentation explaining the theory and presented instructional videos with examples of Fit to Study warm-ups and infusions. In some

cases, schools sent either multiple members of their PE staff, including the Head of PE, to the training while others sent one member of staff, either the Head of PE or a senior member of PE staff.

Depending on their availability, school staff chose the most suitable mode of training and there were examples of staff attending face-to-face training, online training, and accessing the pre-recorded video links.

Fourteen schools were trained face-to-face, during July 2017. Twenty-two schools were trained via livestream, and all schools were sent the pre-recorded video link of the livestream training with the September newsletter. “Top-up” training was incorporated into the initial school visits conducted by the developer in the autumn term. This involved a member of the developer team attending the school in person to ensure that they were aware of all elements of Fit to Study to be delivered in Year 8 PE lessons.

Attendees at the face-to-face training, which was conducted in small groups of around 10 teachers, appreciated the format as it was more interactive and helped teachers feel confident about what they needed to do to deliver Fit to Study. Watching demonstrations of the warm-up and infusions and being able to ask questions and discuss Fit to Study with other attended was described as beneficial. However, teachers also felt that it was time consuming as in some cases teachers had to travel long distances to attend, and inconvenient for those who struggled to take time off work. If the intervention were delivered to scale, this training format may not be feasible with a larger number of participating schools.

Due to time constraints, the teachers who were not able to take part in face-to-face training in July 2017, were given the option to take part in on-line training, thereby relieving teachers also of the requirement to travel to attend a training session. Those who participated in the live online training also had the opportunity to ask questions via email and found this function helpful. It was suggested that an online chat facility would have made this process easier as there was a delay in the sending and receiving of emails making the process stilted and disrupted the flow of the training. Although the online training was viewed as convenient and less time consuming, it was more difficult for the developers to ensure schools had fully understood the training. Teachers who were unable to attend the face-to-face or online training were sent a link to a recording of the online training session.

The resources made available to schools were the Fit to Study instructional videos, the Fit to Study PowerPoint presentation, and online links to further information about fitness. These teachers cascaded the training by using the resources provided to other members of PE department in their school. Fit to Study information was shared during departmental meetings either by replicating the training in its entirety, or by presenting information perceived to be most relevant to delivery and omitting the theoretical background, in order to save time. Staff trained in this way mentioned that they were confident of implementing Fit to Study as it was part of their normal practice as PE teachers.

Overall, teachers suggested the balance of theoretical and practical content within the training could be shifted, with less time spent on the theoretical background of the intervention and more time dedicated to concise information which focused on the practicalities of delivering the intervention. The desire for more direction was discussed by both teachers who attended the face-to-face training and those who received live online training.

> *“What I needed and wanted was, 'This is what you need to do, this is how you need to do it.'”* (Year 8 PE teacher, 1017)

In addition, teachers who attended the training thought it would be useful to have clear guidelines about what information to prioritise when cascading to other members of staff. They felt this would have helped to ensure consistency in the information that was cascaded.

> *“They've not really shown or given us guidance in terms of how to present to the staff, so I sort of tried to give the training they gave me to them.” (*Year 8 PE teacher, 1092)

During the trial, due to delays in delivering the training, not all schools had been trained before the start of the intervention in September 2017. This was due to assumptions around being able to train staff during the summer holidays and a late start to setting up the training sessions. Consequently, delivery was delayed in a small number of schools. See Appendix K for further details on the contents of training. All treatment school staff had received some form of training by late October 2017. During the Autumn term 2017, the developer also delivered ‘top-up’ training during visits to schools as a way to ensure that the requirements for Fit to Study had been understood and schools

knew what to do to deliver the intervention as intended. These pre-planned visits were adapted to deliver this "top-up" training.

Should the training be scaled up, similar issues with its delivery would most likely persist. Schools might struggle to release more than one member of staff and without clearer guidance for cascading the training to other members of staff it would be hard to ensure consistency in delivery of the intervention. Whilst training using a pre-recorded video is less time consuming for teachers, it becomes more difficult to ascertain whether schools understand the requirements for delivery as intended without some type of 'top-up' face-to-face support.

Planning for and discussions about the intervention

Schools incorporated Fit to Study into their lesson planning and PE department discussions, though the type of planning and preparations carried out by teachers varied. Data from the post-intervention school survey indicates that the majority of schools (91.3%, n = 21) reported setting aside time at the beginning of term to discuss the intervention and its' delivery. Teachers also described having regular informal updates with other PE staff members. Communications about Fit to Study continued throughout the year and included discussion of the perceived benefits of the intervention, possible adaptations, and pupil responsiveness.

Ongoing planning was encouraged by the delivery team to ensure staff delivered Fit to Study throughout the year and during each PE lesson. Just over half (52.2%, n = 12) said that they carried out planning on an individual basis, for example teachers taking it upon themselves to redesign some of their lesson plans. Teachers explained that planning was required for creating sport-specific infusions, for example for football or swimming lessons.

There were two views on the amount of planning needed to incorporate Fit to Study into a PE lesson. On the one hand, teachers felt that the intervention did not require any planning or change in teaching style.

> *"It really didn't change our teaching much to be honest or our planning mode or anything."* (Year 8 PE teacher, 1077)

The other view was that some initial planning was needed to embed Fit to Study in lesson delivery, after which ongoing 'extra' work was not needed:

> *"It's just become an integral part of our lessons now so in terms of planning I don't think it needs any."* (Year 8 PE teacher, 1059)

Overall, it seems that teachers were able to implement Fit to Study with minimal additional resource spent on planning and preparation.

Ongoing support and resources

During the trial, ongoing support for PE teachers, mainly through online resources was made available by the developer. Communication between the developers and treatment schools included email and telephone support and termly visits which included observations of PE lessons. Teachers were positive about the support they received from the developers, felt it was easy to make contact, and their questions were answered quickly.

Examples of additional resources used to help with delivery included the development of an activity sheet for pupils by PE staff so that they could lead the Fit to Study activities themselves. Teachers also expressed an interest in how other schools were implementing Fit to Study and suggested there should be opportunities to share good practice. The developer had put in place an online forum for sharing good practice and incentivised school participation using the offer of prizes, but schools did not take this up.

Although there were open communication channels for those seeking support, the developer experienced challenges in maintaining contact with some schools that were less engaged with the intervention and were not responsive to any communications. This may have been a factor in the higher than expected attrition at school level.

Schools found it helpful to discuss barriers to implementation directly with the developers. These included issues around fitting in the warm-up and three infusions into each of their PE classes. Where teachers raised these issues, they were provided with reassurance that other schools were facing similar challenges. In some cases, schools were instructed

to carry out just two infusions instead of three. Without fully completed teachers logs, it is difficult to assess the extent to which the modification was carried out.

Dosage

Although Fit to Study was reported to be well-structured, and despite the commitment of treatment schools to deliver the intervention as intended, there were a range of issues that affected delivery as intended.

Fit to Study is both prescriptive and flexible. While the four minutes of VPA during the warm-up is necessarily at the start of the lesson, the three infusions can be delivered at any point during the lesson, as long as each PE lesson includes 10 minutes of VPA.

The post-intervention schools survey found that dosage varied from full delivery as intended to schools dropping out of the intervention altogether. Fit to Study was delivered in all lessons as intended in 9% (n=2) of schools. While some schools ensured that the warm-ups and infusions took place irrespective of the context, regardless of sport, weather and location, teachers felt that occasionally Fit to Study activities had to be omitted in order to meet lesson objectives.

> *"Sometimes it could be just that it wasn't feasible inside that lesson to deliver a good or outstanding lesson and have the infusions in there as well." (*Year 8 PE teacher, 1074*)*

Schools raised the issue of having to carry out the warm-up and the three infusions in each PE lesson with the developer. Teachers found it challenging to incorporate the full Fit to Study warm-up in all lessons as it extended warm-up time and reduced teaching time during the lesson. For this reason, the time allocated to the warm-up was often shortened to include only the parts perceived to be most important to the study by the teacher (i.e. the infusions). The warm-up was also often initiated before all pupils were changed and ready to take part.

Teachers also felt that the number of infusions they had to deliver within a lesson was demanding and struggled to incorporate the three required infusions in each lesson. Teachers reported delivering two infusions at once (that is one longer infusion), ending a session with an infusion, or even skipping one or more infusions altogether. The main reasons for these adaptations were a lack of time and the perceived disruption to the flow of the lesson.

> *"It does kind of break up your lesson a little bit, so if you're in the middle of something you bring them in and do an infusion, you can kind of lose the focus*." (Year 8 PE teacher,1053)

After requesting advice from the developer, some teachers, particularly those with shorter (50 minute) PE lessons, reported being informed that they could omit the third infusion occasionally, as long as they carried it out in the majority of PE lessons. Schools also mentioned that the developer told them that other schools were also struggling with fitting in the third infusion, which might have led to a reduced number of infusions becoming more standard practice.

Other unpredictable events throughout the course of the year, such as poor pupil behaviour, medicals, teacher observations and other incidents such as injuries, meant that delivering all aspects of Fit to Study was not possible. Seasonal variations also led to adaptations in Fit to Study, for example, the cancellation of an outdoor PE lesson due to rain led to Fit to Study activities either being altered, reduced or omitted entirely.

> *"There are days where we just can't get it done or we can't implement it in the way that we wished to*." (Year 8 PE teacher, 1104)

Adhering to the requirements to deliver Fit to Study as intended was better at the start of the academic year, but over the course of the three terms, schools reported struggling to include the intervention during PE lessons. For some schools, this eventually led to Fit to Study disappearing completely from regular Year 8 PE practice. Schools preferred to retain the warm-up over the infusions if they did not want to break the flow of their PE lesson, and if they felt that their usual lesson was already included vigorous physical activity. This was because teachers reported always beginning their PE lessons with a warm-up which could easily be replaced by the Fit to Study warm-up.

Concerns were also raised about Fit to Study hindering pupils' progression in certain PE activities and sports. They felt that some pupils were not achieving the level of skills that they should have been due to the focus on Fit to Study and the time take out of PE lessons for to deliver it. Teachers reported being informed by the developer that it would be sufficient to implement the intervention once a week, or every other lesson, in this case.

Quality

A range of factors affected quality including the size of the class, differences in teacher style and time constraints. Teachers reported that they still attempted to use the principles of Fit to Study in their warm up and lesson plans even if they were no longer precisely following the intervention activities. Differences in delivery were reported in terms of how each teacher delivered the warm-up and infusions, depending on their own teaching style and methods and their perceptions of the value of Fit to Study. More experienced teachers were felt to be more traditional in their teaching style and reluctant to adapt their approach to warm-ups and lessons.

> "*Some teachers would have put more focus on the physical side, others would have led heavily on making it very sports specific."* (Year 8 PE teacher, 1017)

It was also highlighted that ensuring the correct intensity throughout each two-minute infusion in larger classes could be difficult. In the observations, teachers were seen to bring pupils together for the two-minute infusion, but the intensity of the infusion activity would not last the prescribed two minutes. In some cases, the infusion activity consisted only of three or four ten-second bursts of activity.

> *"If you're out on a field you've got 30 something kids, it's very difficult to ensure that the intensity of that is being carried out appropriately for every single pupil doing the actual infusion."* (Year 8 PE teacher, 1056)

Quality was also affected due to time pressures. The warm-up, which was seen as regular PE practice already, was implemented by PE staff. Schools would occasionally cut the warm-up down by two to three minutes due to time restrictions, for example, or by starting before all pupils were ready to take part.

Barriers and conditions for successful delivery

This section comments on the practical delivery arrangement employed by schools as well as school-specific barriers to delivery.

Practical delivery arrangements

Schools found that practical issues, such as the planned lesson activity, focus on a particular sport, available space, the location of the lesson, and class size affected delivery.

- **Lesson activity**: Teachers agreed that the intervention was better suited to particular PE activities. These included team sports such as hockey, football and basketball, where the infusions could be incorporated more easily into sport-related activities. It was reported to be more difficult to implement the infusions in lessons where they would '*stand out'* more, such as gymnastics and badminton, and in those where skill development was a priority, such as athletics. This was because the intervention was perceived to disrupt the flow of the lesson and take pupils away from their skill practice.
- **Space**: Adequate space was perceived to be essential for implementing Fit to Study, as teachers highlighted the difficulties they faced delivering the intervention in lessons where space was limited, such as when focusing on sports like gymnastics or trampolining where large amounts of equipment were in use.
- **Location and class size**: Teachers described the benefits and problems associated with delivering the intervention during indoor and outdoor lessons. Indoor spaces were found to be useful for focusing pupils however, lack of space could be a challenge. Indoor delivery was reported to be easier if teaching a smaller class. Outdoors, there was more space but practical issues such as bad weather made outside delivery more challenging.

School-specific barriers to delivery

School-specific factors affecting fidelity included the length of PE classes at some participating schools, varying approaches to teaching PE, existing lesson plans and PE objectives and their compatibility with the intervention.

Where teachers felt that PE lessons were not long enough to be able both to deliver Fit to Study and achieve lesson objectives, time was described as an obvious barrier. This was particularly the case at schools with fifty-minute lessons, but also for those with sixty-minute lessons. The suggested solution to this was longer lessons, or a greater number of lessons per week, which would make it easier to deliver Fit to Study and achieve skills objectives.

> *"…It's just been hard to implement with the progress and objectives that we're trying to teach within our lessons."* (Year 8 PE teacher,1074)

Teaching styles differed across school and there was a concern that the activity-focused direct teaching style and prescriptive lesson structure of Fit to Study was not wholly compatible with independent learning, where pupils are encouraged to explore and develop according to their own interests and needs, currently espoused by schools. It was felt to clash with some schools' pupil-led approach to PE, such as involving pupils in the leading of certain sections of the class, for example, the warm-up.

Responsiveness

Pupil responsiveness varied depending on the level of adaptation and the delivery approach taken by teaching staff. Schools where teachers felt that buy-in was secured from senior members of school staff perceived a greater level of pupil motivation to take part in PE throughout the year. This was due to staff feeling more motivated by senior members of staff to deliver the programme, which filtered through to pupils. This was also the case in schools that adapted the warm-up to be pupil-led and the infusions to be context and sport-specific. School level commitment was evident in instances where Fit to Study had been most successfully and faithfully delivered.

> "*We've tried to make it part of the school. Different staff have asked what we're doing and we've made that very clear to the staff that work with the Year 8s and we obviously made it a school thing, not just a PE thing and the head teacher's been very supportive through it."* (Year 8 PE teacher, 1042)

Pupil responsiveness was also affected by confidence, with those pupils who were perceived to be more physically fit or capable in terms of the specific sport enjoying the infusions more. This was particularly the case for the infusions that were adapted to be sport-specific. Competitive elements, for example splitting the class up in to groups, was seen to increase pupil engagement in the infusions, but this also exacerbated the divide between the confident and less confident pupils.

Teachers struggled to describe any change in children as a result of taking part in Fit to Study. This affected their commitment to the programme and they mentioned that it would be the difficult to continue to deliver the intervention without being able to see any clear outcome.

> *"To stay committed for a year when you're not actually seeing or hearing any benefits is quite a challenge in itself."* (Year 8 PE teacher, 1101)

The lack of visibility of any perceived outcome for children was reinforced by teachers feeling that they were not able to implement Fit to Study correctly.

> *"Sometimes it has been in a rush and so then you kind of feel, well, was there any point in doing that if I've just rushed through it."* (Year 8 PE teacher, 1104)

Teachers also described the intervention as repetitive and found this to be a barrier to keeping pupils engaged. This was anticipated as a potential barrier by teachers at the start of the school year and was reported to materialise in the Spring and Summer terms. Declining levels of engagement, as the 'novelty' of partaking in the intervention wore off,

was described as a key factor in one school's decision to stop delivering Fit to Study half way through the year. Teachers reported that their pupils would rather be doing PE lesson activities and focusing on skills development than the intervention exercises.

> *"…No matter how much we try and motivate them, their effort and their level of engagement in the skill practice is a lot higher than it is in the intervention."* (Year 8 PE teacher,1092)

Were the intervention to be continued into Year 9 for the same cohort of pupils, sustainability would likely be affected by decreasing teacher and pupil motivation.

Perceived outcomes and level of engagement

At the start of the intervention teachers were hopeful for an increase in pupil fitness levels and academic achievement, but there was some concern that the amount of contact time that teachers had with pupils would not be sufficient to enable an increase in fitness, and as a result improve academic outcomes.

Perceptions of the outcomes of the intervention at the end of the school year ranged from none to increased fitness and improved pupil social skills, such as confidence and leadership, as demonstrated by pupils leading activities during the class. As PE teachers were not always in direct contact with other teachers, they struggled to ascertain if Fit to Study had had any effect on academic attainment.

Overall, teachers concluded that the value to their pupils might have just been limited to physical outcomes, such as increased fitness, but they did feel hopeful about the intervention as a larger scale study that had the potential to boost the importance of PE across their schools. Teachers were optimistic that the study might provide evidence of a positive link between increased physical activity and fitness and academic attainment. In turn, they hoped that this would enable them to bargain for increased PE lesson time.

Teachers felt that Fit to Study helped to increase fitness levels, particularly for children in lower skill groups, in that it kept pupils active for longer. They also commented on the fact that it enabled teachers to incorporate fitness-related activities even in non-physically demanding classes or classes that included greater time spent waiting or in rotation, such as badminton or gymnastics.

The Fit to Study warm-up was a good resource and was suited to being pupil-led. This required pupils to research and then lead the warm up, which was perceived to help to develop leadership and social skills and the confidence of those who might not otherwise fully engage with PE. It was perceived by teachers to be well suited to being pupil-led as it contained short, highly prescribed activities that all pupils could partake in simultaneously for a controlled amount of time under the supervision of the teacher.

> *"It worked well with our Year 8s and it has actually brought them on socially and they have started to become more confident with it. So we thought, well, let's try it with our Year 10 and see if it was the same outcome."* (Year 8 PE teacher, 1104)

However, pupil enthusiasm waned over the course of the year. Teachers reported that although pupils got used to the infusions, towards the end of the year the lower ability pupil seemed to be going through the motions and the higher ability pupils were keen to get on with the PE lesson. In their view this was a reflection of their general attitude to PE lessons.

> *"when you say, 'Come on, we've got infusions', you do generally get a groan … if you're in the middle of a football game and we've stopped it and you're doing infusion, they're like, 'Well, I want to get back to playing football'." (PE teacher 1053)*

Perceptions of the intervention

At the end of the trial, the majority (68%, n = 19/28) of treatment schools that responded to the survey indicated that they would recommend the Fit to Study intervention because it helped to promote the value of physical exercise and raise awareness of the importance of PE in the school curriculum. In addition, school felt that the intervention made it easier to

- integrate high intensity exercise into PE lessons, and
- keep pupils as active as possible.

Interestingly, teachers who were positive about the intervention emphasised that they felt their pupils had enjoyed the infusions and that Fit to Study had become embedded in the delivery of Year 8 PE lessons at their school. Fit to Study was also perceived to be a useful teaching tool for achieving focus within lessons, for example, by bringing pupils in small groups together at set, regular intervals throughout the PE class. Teachers acknowledged that as a concept Fit to Study was valuable but needed to be adapted:

> *"…we just need to tweak it, and for them to work with schools to make it as good as it can be" (YR 8 PE teacher 1074)*

The schools survey data indicates that thirty two percent (n = 9/28) of schools said they would not recommend Fit to Study because of concern about:

- the potential disruption to lesson delivery and lesson 'flow' caused by Fit to Study
- perceived hindrance to pupil progress and development in PE due to taking up too much regular teaching time.

While the premise of the intervention was attractive, teachers were concerned by the requirement to deliver it in all lessons regardless of lesson content and learning objectives. Furthermore, the intervention was perceived to support a method of teaching that PE staff felt to be outdated, less engaging and less pupil-led:

> *"It's almost that Victorian style of teaching where people stand in rows and they swing their arms around, and then they leave. And that's not really what school PE should look like these days."* (Year 8 PE teacher,1053)

Unintended consequences

A number of unexpected consequences arising from intervention delivery were identified. These included:

- **Disrupted flow of lesson –** It was hard for some schools to engage their pupils as they were oriented around traditional sports and activities. Pupils were not thought to like being taken away from the standard curriculum that they had become accustomed to. Teachers also found the nature of delivery to disrupt the flow of PE lessons.

  > *"One of the things we did find is that sometimes to actually stop the flow of the lesson and go to something completely different was quite confusing and disruptive for the students."* (Year 8 PE teacher, 1074)

- **Curriculum affected –** Teachers found that covering the entire curriculum was made more difficult by having to include the Fit to Study intervention activities in every Year 8 PE lesson throughout the academic year. This was felt more acutely in schools where the intervention activities were not well adapted or suited to the PE sport or activity being taught.

  > *"The key challenge was trying to get everything in whilst actually still delivering a curriculum."* (Year 8 PE teacher, 1101)

- **Fit to Study lessons less active than business as usual –** It was suggested that business as usual PE lessons contained an active and dynamic warm-up. Because teachers had to stop planned activity to bring pupils together to deliver Fit to Study infusions; this was felt to increase inactive time which could potentially lead to lessons being less active overall than regular PE lessons.

> *"It's not just the two minutes and the ten minutes, it's all the dead part around it."* (Year 8 PE teacher, 1092)

Adaptations and modifications

Experiences of delivery ranged from delivering exactly as intended, with no adaptations, to adapting the intervention and delivering it in the way that suited the school. At times this would include not delivering any infusions at all.

School staff found that the intervention was not 'one size fits all' and described it as being difficult to deliver during PE lessons focused on specific sports such as, such as gymnastics and athletics, where equipment is involved and space limited, and summer sports that focus on tactical skills, such as cricket. Sports such as trampolining were particularly difficult because they required large equipment and the teacher's attention being focused on one pupil at a time. Combined with the 'dead' time for other pupils while they rotated on the equipment, this was one of the more challenging sports for Fit to Study delivery.

After discussion with the developers, schools felt empowered to alter the warm-up and infusions for two reasons, to make things sport-specific and fit into the flow of their lesson and to reduce monotony for children and maintain motivation. Schools mentioned letting pupils lead the warm-up and reported that the developer gave their permission for these kinds of adaptations, as long as the total quantity of VPA throughout the class remained.

> *"They've* [the developer] *been quite willing to listen to how people adapt and how we can make things work*." (Year 8 PE teacher, 1017)

Schools also mentioned making adaptations without consulting the developer first where they have felt it appropriate.

> *"We've taken it on ourselves to try and think up ideas rather than asking the developers."* (Year 8 PE teacher, 1104)

The main changes to Fit to Study activities revolved around making them more interesting in order to maintain pupil motivation. These adaptations included making the warm-up pupil-led and the infusions sport-specific. Teachers were also keen to ensure that Fit to Study matched the school's priorities.

> *"We actually, to some extent, have gone away from their Fit to Study style infusions because we didn't feel it worked for our students or worked for us as a school. So we've adapted and rewritten our own infusions so the way we do it is that infusion one is done as part of the student led warm-up."* (Year 8 PE teacher, 1059)

Omissions and adaptations were justified by teaching staff in that they felt they had to prioritise their teaching requirements. Conversations with the developers led to schools perceiving the entire intervention to be adaptable, especially in terms of ensuring the PE curriculum is followed.

> *"We're trying to run the curriculum alongside this programme, so it's probably been a bit of a compromise and the best of both worlds."* (Year 8 PE teacher, 1101)

The adaptable nature of the intervention was viewed positively, and this was felt to have benefitted pupils in that the intervention could be adapted to their school's context and requirements. The warm-up in particular was altered in order to make them more high intensity and raise pupils' heart rates.

> *"By tailoring it to our needs and the way we deliver things it has really, really taken off and benefitted the students now."* (Year 8 PE teacher, 1059)

However, it is possible in that adapting the intervention to a certain extent would result in schools not carrying out the required amount of VPA as prescribed by the developer.

In addition to the high level of flexibility with which Fit to Study was delivered within PE lessons, the adaptability of the infusions had wider application. The infusions were reported to have been used in other non-PE classes to help calm and control a class when they were being loud and disruptive.

> *"They're adaptable. You could do them in a classroom. I've done it with my form […] I've got a year-ten form, and I said, 'Right, everyone, sprinting on the spot, ten seconds', and then I did the shoulder rolls. […] It's something that has actually helped me as a teacher, because I can use it for other stuff."* (Year 8 PE teacher, 1042)

Formative findings

The findings from the process evaluation suggest that Fit to Study was appropriate and valued by teaching staff. Strategies to increase Year 8 pupils' level of fitness were felt to tie in well with the overall aims for Year 8 PE. However, some alterations were made to the intervention activities by teachers to facilitate the implementation of Fit to Study throughout the academic year. While teachers understood the value of the intervention as a whole they struggled to deliver it as intended. The challenges to delivering the intervention as intended combined with a lack of visible benefits for pupils meant that maintaining Fit to Study was difficult for teachers.

Aspects of intervention delivery that could be reviewed

1. **Training** – Teachers who attended face-to-face training felt that the training could be shortened. They commented on the overly theoretical elements of the training which could be reduced to focus more on the practical delivery of Fit to Study and how to effectively cascade learning to other staff. Ensuring training was held before the end of the Summer term in the preceding academic year would also enable delivery from the first Year 8 PE lesson of the delivery year. It was also thought that there was still a need to balance online and face-to-face modes of delivery for training, and it was noted that online training could be done at the teacher's own pace at a convenient time.

   > "*I was happy with it being face-to-face but that's because I'm old-fashioned. I'm sure that other people would like it online given the busy nature of teaching and the profession."* (Year 8 PE teacher, 1097)

2. **Support for adaptions** – Providing teachers with further support for making adaptations (for example, different versions of infusions for indoor and outdoor activities, for different sports, for different length PE lessons, and for different weather conditions) would help teachers with implementation and pupil motivation. Adaptations to the warm-up and infusions would also help to keep children motivated,

   > *"To re-energize the staff in doing it, re-energize the students so they're not doing the same activity every time."* (Year 8 PE teacher, 1021)

   This could take the form of termly advice with sport-specific infusion options to help maintain the flow of PE lessons and fit in with wider PE priorities. These adaptations would need to reduce the time taken to implement Fit to Study activities.

   > *"We don't want the same stretches being done for football and for badminton, for example."* (Year 8 PE teacher, 1059)

3. **Fitting in with curriculum –** Fit to Study activities could better support the learning objectives of the class. Adapting the intervention activities to fit in with curriculum demands would allow the opportunity to have some lessons based on skills (without infusions) and some fitness-based (with infusions).

4. **Pupil-led elements** – Resources could be more pupil-oriented to enable pupils to develop and lead Fit to Study activities. This could help with pupil engagement and motivation and develop soft skills, such as confidence and leadership.

> "*If they were to produce something a bit more pupil-friendly so that they can potentially lead themselves through it, that would be more useful."* (Year 8 PE teacher, 1017)

5. **Modernising the intervention activities** – The warm-up and infusions could be updated to more adequately reflect the current PE curriculum and modern teaching styles. One way to do this could be through making the infusions more dynamic and sport or skill-related.

> *"For me, is quite alien now to be standing there with a group in front of me all doing the same activity. That's just a bit sort of old-school."* (Year 8 PE teacher, 1053)

6. **Beginning delivery with Year 7** – The beginning of Year 7 is a logical starting point for an intervention as pupils are more likely to be new to a school. If delivery of Fit to Study began at this point then it would potentially be embedded into PE by Year 8 and could become a whole school intervention.

> *"It makes sense to start as you mean to go on in Year 7 and get them, you know, get them forming those good habits. And actually then you've got five years to actually see significant impact."* (Year 8 PE teacher, 1101)

Control group activity

The trial assumed a 'business as usual' approach whereby control schools were not invited to deliver Fit to Study. Control schools described a range of aims for Year 8 PE, including maximising participation, preparation for GCSE, and encouraging healthy competition and resilience. As would be expected, improving pupil fitness was also described as a priority for PE across all year groups and therefore formed a key part of control schools' 'business as usual'.

Schools gave examples of how fitness was incorporated into PE, including weekly extra-curricular fitness sessions and providing dedicated spaces for fitness. Pupil-led warm-ups did feature in control schools' descriptions of their 'business as usual' for Year 8 PE, however, from the qualitative interviews and lesson observations, there was no evidence of control schools undertaking any compensation activities, such as Fit to Study-style modifications to lessons. However, two schools reported that they were delivering Fit to Study which may have been a reference to taking part in the trial and evaluation activities rather than delivery of the intervention itself.

Control schools reported involvement with other PE interventions and these were focused on increasing girl's participation in exercise, such as the Girls Active programme run by the Youth Support Trust. While only targeted at girls, teachers at control schools felt that this was complementary to Fit to Study in that it had a similar goal to increase activity levels. In control schools who were not currently involved in any PE interventions other than Fit to Study, it was expressed that girls' uptake of, and confidence in, PE was an area requiring more work.

Similar to treatment schools, control schools were focusing on increasing the amount of physical activity and fitness of their Year 8 cohorts. This could have eroded any larger differences between the treatment and control group resulting the lack of any observable impact.

Conclusion

Interpretation

This trial did not find evidence that increasing vigorous physical activity (VPA) by delivering the Fit to Study intervention during PE lessons improves maths attainment of participating pupils. Year 8 pupils in Fit to Study schools did not make additional progress in maths compared to Year 8 pupils in 'business as usual' control schools. Furthermore, no additional progress was made by pupils who are eligible for Free School Meals in Fit to Study schools. There is mixed evidence of the effectiveness of interventions promoting MVPA on cognitive function and attainment, and unfortunately due to implementation issues, the null result found by this evaluation has to be interpreted very cautiously and further rigorous research is required. The findings presented above run counter to the conclusions from recent evidence that pointed to the positive effects of physical exercise interventions on pupils' cognitive functions (Kamijo et al, 2011; Chaddock-Heyman et al., 2013), but are closely aligned with other studies that also found null results on attainment, even with considerably more intense interventions (Donnelly, 2009).

Low compliance with the intervention and high attrition were the main limitations of the impact evaluation. The overall pupil-level attrition was 44% and reached 54% among the treatment pupils. High and uneven attrition had a negative impact on the precision and security of this trial. Low intervention fidelity made the hypotheses testing more difficult. While this trial assumed the presence of a single 'business as usual' condition, it is likely that the actual 'business as usual' PE lessons differed in their intensity across schools and PE teachers. Additionally, the absence of a uniform control condition could have contributed to the null finding.

The implementation study indicates that another reason for the null finding could lie in the way the Fit to Study intervention was designed. Although there were reported instances of Fit to Study helping to increase children's level of physical activity, PE teachers felt that bringing pupils together for the VPA infusions required making regular stops to the lesson which resulted in the increase of 'dead' time. In some cases, this could lead to Fit to Study lessons being less active overall than regular PE lessons. This finding relates to the more general observation on the differences in PE regimes across schools and PE teachers. Furthermore, evidence from the implementation and process evaluation indicates that it is likely that the intervention did not increase activity in Year 8 PE classes in the treatment group more than in the business as usual group. It is to be noted that teachers recognised the importance of improving pupils levels of fitness and were keen to find ways to do so. It may be for this reason that a majority responded that they would recommend Fit to Study despite the implementation issues experienced.

Another problem related to the Fit to Study intervention design was the negative impact that the separate VPA infusions were perceived to have on the lesson 'flow'. Some PE teachers claimed that it was very difficult to incorporate the VPA into certain lessons, viewing 'fitness infusions' as disruptive to lesson delivery. Changes to intervention dosage by delivering fewer infusions, for example, and adaptations made to accommodate factors such as sports skills and seasonal effects, were reflected in low intervention fidelity in treatment schools. A reduction in intervention intensity was also reported and subsequently permitted by the developer, with Fit to Study intervention activities therefore not being delivered in each Year 8 PE lesson. This was confirmed by the low level of compliance reported in the school survey, with less than a tenth of treatment schools (9%, n=2) reporting delivering Fit to Study in all Year 8 PE classes. As a result, it is also not possible to know whether 10 minutes of VPA in each PE lesson is the optimum amount of VPA to achieve cognitive change.

Additionally, various difficulties were faced when organising the training for PE teachers prior to the start of the intervention. The training was intended to be delivered face-to-face in the summer preceding the trial, but low attendance meant that online training had to be arranged for the start of the autumn term for untrained schools. This delayed implementation in some schools. 'Top-up' training then had to be incorporated into school visits to ensure that all schools were able to deliver as intended.

Only two out of 32 treatment schools that conducted the post-intervention testing reported that all PE lessons across all Year 8 form groups in 2017/2018 academic year were delivered in line with the Fit to Study intervention model. 17 treatment schools indicated that more than half of all Year 8 PE lessons were delivered as planned. The implementation and process evaluation also found that many participating schools were more strongly motivated to take part in the

research project than they were attracted to the intervention itself. This could pose a threat to uptake and delivery of Fit to Study if it were rolled out in the future.

It is also important to stress that problems around the implementation of this trial as well as the difficulties in post-intervention testing resulted in high attrition figures. The overall rate of pupil-level attrition was 44%. Furthermore, pupil-level attrition was significantly higher among the pupils in Fit to Study schools (54% compared to 35% among the pupils in control schools). High and uneven attrition had a negative impact on the precision of this trial. However, it did not lead to the increase in Minimum Detectable Effect Size (MDES). This was mainly due to the estimated intra-cluster correlation (ICC) being lower than originally anticipated (0.11 compared to 0.16).

The null results of the trial may indicate weaknesses in the design of the intervention requiring a review of Fit to Study's component parts and studying their fit within PE lessons of different lengths, the PE curriculum, and teaching styles that incorporate elements of inquiry-based learning.

Limitations

Low compliance with the intervention and high attrition were the main limitations of the impact evaluation.

- Only 17 of the 32 Fit to Study schools that conducted the post-intervention testing reported that more than half of all PE lessons across all Year 8 form groups in 2017/2018 academic year were delivered in line with the Fit to Study intervention model.
- Overall pupil-level attrition was 44% and reached 54% among the treatment pupils.
- The absence of a uniform 'Business as Usual' in control schools may mean that in some cases the intensity of PE lessons in BAU schools was not lower than in Fit to Study schools.
- Variations in fidelity makes it unlikely that 10 minutes of VPA was achieved by pupils during each PE lesson.

Given the lack of specificity of Fit to Study and the uncertainty around whether 10 minutes of VPA in each PE class equates to a sufficient amount of MVPA to make a difference to cognitive ability, it is assumed that the null result is generalisable to all Year 8 pupils in state secondary schools in England. However, even if the schools that participated in the trial were broadly representative of schools nationally, the high level of attrition, and possible unmeasured characteristics (such as, the desire to take part in research or to deliver an intervention designed by academics) mean that schools that opted into the trial may still be systematically different, affecting the external validity of the results presented in this report, and increasing the need for further robust research in this area.

Future research and publications

The limitations to this evaluation, including its implementation, mean that further research is required into the likely links between increased physical activity, increased cognitive function and increased attainment. The developers of Fit to Study will report on the secondary measures that they have captured within this trial, including measures of activity, fitness, wellbeing and cognitive function.

Beyond the current trial, a review of the intervention design alongside research to see what type of (M)VPA elements could be successfully embedded within PE lessons, without affecting PE lesson structure and supportive of teaching styles, is needed. This would help create a firmer grounding on which to base any future research into a Fit to Study-style intervention.

Following a design review, more rigorous research would also be needed to confirm the impact of the Fit to Study intervention on fitness outcomes, ensuring that any covariates associated with the improvement in maximum aerobic capacity are controlled for.

References

Bailey, R., Wellard, I. and Dismore, H. (2004). Girls' participation in physical activities and sports: benefits, patterns, influences and ways forward. Technical Report.World Health Organisation. Available at http://www.icsspe.org/system/files/Stead%20and%20Neville%20-%20The%20Impact%20of%20Physical%20Education%20and%20Sport%20on%20Education%20Outcomes.pdf

Blakemore, S., & Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. Journal of Child Psychology and Psychiatry, 47(3-4), 296–312.

Booth, J. N., Leary, S. D., Joinson, C., Ness, A. R., Tomporowski, P. D., Boyle, J.M., and Reilly, J. J. (2014). Associations between objectively measured physical activity and academic attainment in adolescents from a UK cohort. *British Journal of Sports Medicine,* 48(3), 265-270.

Buchheit M, Laursen PB. High-intensity interval training, solutions to the programming puzzle: Part I: Cardiopulmonary emphasis. Sport Med. 2013;43:313–38.

Carlson SA, Fulton JE, Lee SM, et al. (2008). Physical education and academic achievement in elementary school: data from the early childhood longitudinal study. *American Journal of Public Health*, 98:721-727.

Centers for Disease Control and Prevention (2010) *The association between school based physical activity, including physical education, and academic performance*. Atlanta, GA: U.S. Department of Health and Human Services.

Chaddock-Heyman, L., Erickson, K., Voss, M., Knecht, A., Pontifex, ,M. Castelli, D., Hillman, C., and Kramer, A. (2013) The effects of physical activity on functional MRI activation associated with cognitive control in children: a randomized controlled intervention. *Frontiers in Human Neurosci*ence 7(72).. doi: 10.3389/fnhum.2013.00072

Cooper, Simon B., Karah J. Dring, and Mary E. Nevill. "High-intensity intermittent exercise: effect on young people's cardiometabolic health and cognition." *Current sports medicine reports* 15.4 (2016): 245-251.

Costigan SA, Eather N, Plotnikoff RC, Hillman CH, Lubans DR. High-Intensity Interval Training for Cognitive and Mental Health in Adolescents. Med Sci Sports Exerc. 2016;48:1985–93.

Costigan SA, Eather N, Plotnikoff RC, Taaffe DR, Lubans DR. High-intensity interval training for improving health-related fitness in adolescents: a systematic review and meta-analysis. Br J Sports Med. 2015;bjsports-2014-094490.

DCMS and DfE (2015) 2010 to 2015 government policy: sports participation https://www.gov.uk/government/publications/2010-to-2015-government-policy-sports-participation/2010-to-2015-government-policy-sports-participation

Department for Education (2013) Evidence on physical education and sport in schools: key findings. Report Ref: DFE-00092-2013

Donnelly, J., Greene, J., Gibson, C., Smith, B., Washburn, R., Sullivan, D., DuBose, K., Mayo, M., Schmelzle, K., Ryan, J., Jacobsen, D., and Williams, S. (2009)Physical Activity Across the Curriculum (PAAC): a randomized control trial to promote physical activity and diminish overweight and obesity in elementary school children. *Preventive Medicine*, 49(4):336-341.

Donnelly, J. E., Hillman, C. H., Castelli, D., Etnier, J. L., Lee, S., Tomporowski, P., ... & Szabo-Reed, A. N. (2016). Physical activity, fitness, cognitive function, and academic achievement in children: a systematic review. *Medicine and science in sports and exercise*, *48*(6), 1197.

Education Endowment Foundation, 2015. *Policy on Analysis for EEF Evaluations.*

Education Endowment Foundation, 2018. *Statistical analysis guidance for EEF evaluations*. Available at: https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf [Accessed 17 April 2019.

EEF (2014) Neuroscience and Education: A review of Educational Interventions and Approaches Informed by Neuroscience.

Erickson, K., Voss, M., Prakash, R., Basak, C., Szabo, A., Chaddock, L., Kim, J., Heo, S., Alves, H., White, S., Wojcicki, T., Mailey, E., Vieira, V., Martin, S., Pence, B., Woods, J., McAuley, E., and Kramer, A. (2011) Exercise training increases size of hippocampus and improves memory, *PNAS*, 108 (7): 3017-3022. https://doi.org/10.1073/pnas.1015950108

Fisher, Abigail, et al. "Effects of a physical education intervention on cognitive function in young children: randomized controlled pilot study." BMC pediatrics 11.1 (2011): 97.

Gerber, A. S and Green, D.,2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.

GL Assessment, 2015. *Progress Test Series: Measure and Track Progress in English, Maths and Science: Assessment Overview*. London: GL Assessment.

Hedges, L. V., 2007. Effect Sizes in Cluster-Randomized Designs. *Journal of Educatoinal And Behavioural Statistics,* 32(4), p.341-370. Available at **https://www.researchgate.net/publication/250185702_Effect_Sizes_in_Cluster-Randomized_Designs**

Hillman, Charles H., Kirk I. Erickson, and Arthur F. Kramer. "Be smart, exercise your heart: exercise effects on brain and cognition." Nature reviews neuroscience 9.1 (2008): 58.

Hillman, Charles H., et al. "Effects of the FITKids randomized controlled trial on executive control and brain function." Pediatrics 134.4 (2014): e1063-e1071.

House of Commons: The Culture, Media and Sport Committee (2014) Women and Sport. London: The Stationery Office.d

Inchley J, Currie D, Young T, Torsheim T, Augustson L, Mathison F. Health behaviour in school-aged children (HBSC) study: International Report from the 2013/2014 survey. Heal Policy Child Adolesc. 2016

Kamijo, K., Pontifex,M., O'Leary, K., Scudder, M., Wu, C-T., Castelli, D., and Hillman, C.(2011) The effects of an afterschool physical activity program on working memory in preadolescent children Dev Sci. Author manuscript; available in PMC 2012 Sep 1.

Khan, Naiman A., and Charles H. Hillman. "The relation of childhood physical activity and aerobic fitness to brain function and cognition: a review." Pediatric exercise science 26.2 (2014): 138-146.

Kujach S, et al. (2018) A transferable high-intensity intermittent exercise improves executive performance in association with dorsolateral prefrontal activation in young adults. Neuroimage 169:117–125.

Logan GRM, Harris N, Duncan S, Schofield G. A review of adolescent high-intensity interval training. Sport Med. 2014;44:1071–85.

McClelland, E., Pitt, A., and Stein, J. (2014)Embodied Cognition in the Classroom . *Improving Schools*. Vol. 18(1). doi.org/10.1177/1365480214562125

Monti, Jim M., Charles H. Hillman, and Neal J. Cohen. "Aerobic fitness enhances relational memory in preadolescent children: the FITKids randomized control trial." Hippocampus22.9 (2012): 1876-1882.

Moreau, D., Kirk, I. J., & Waldie, K. E. (2017). High-intensity training enhances executive function in children in a randomized, placebo-controlled trial. *Elife*, *6*, e25062.

Mura, G., Vellante, M., Egidio Nardi, A., Machado, S., & Giovanni Carta, M. (2015). Effects of school-based physical activity interventions on cognition and academic achievement: a systematic review. CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders), 14(9), 1194-1208.

Ofsted (2013) Beyond 2012 – outstanding physical education for all. Available at **www.ofsted.gov.uk/resources/120367**.

Ortega, F., Cadenas-, C., Migueles J., Labayen I., Ruiz J., Sui X., Blair S., Martínez-Vizcaino, V., and Lavie C. (2018) Role of Physical Activity and Fitness in the Characterization and Prognosis of the Metabolically Healthy Obesity Phenotype: A Systematic Review and Meta-analysis. *Progress in Cardiovascular Diseases*. doi: 10.1016/j.pcad.2018.07.008

Phan, D-V., Chan, C-L., Pan., R-H., Yang, N-P., Hsu, H-C., Ting, H-W., and Lai, K. R. (2018) A Study of the Effects of Daily Physical Activity on Memory and Attention Capacities in College Students. *Journal of healthcare engineering.*doi.org/10.1155/2018/2942930

Roberts, C., Freed, B., McCarthy, W. (2010) Low aerobic fitness and obesity are associated with lower standardized test scores in children. The Journal of Pediatrics, 156:711–18.

Rubin, D. B., 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Stead, R. and Nevill, M. (2010) The impact of physical education and sport on educational outcomes: a review of literature. /www.icsspe.org/system/files/Stead%20and%20Neville%20-%20The%20Impact%20of%20Physical%20Education%20and%20Sport%20on%20Education%20Outcomes.pdf

Tymms, P., Curtis, S., Routen, A., Thompson, K., Bolden, D., Bock, S., Dunn, C., Cooper, A., Elliott, J., Moore, H., Summerball, C., Tiffin, P., and Kasim, A. (2016) Clustered randomised controlled trial of two education interventions designed to increase physical activity and well-being of secondary school students: the MOVE Project. *British Medical Journal Open* available at http://bmjopen.bmj.com/content/6/1/e009318.full.pdf+html

US Department of Health and Human Sciences (July 2010). The Association Between School-Based Physical Activity, Including Physical Education and Academic Performance.
https://www.cdc.gov/healthyyouth/health_and_academics/pdf/pa-pe_paper.pdf

Warburton, D., Nicol, C., and Bredin, S., (2006) Health benefits of physical activity: the evidence. *CMAJ*, 174(6): 801–809. doi: 10.1503/cmaj.051351

Wassenaar TM, Wheatley CM, Beale N, Salvan P, Meaney A, Possee JB, Atherton KE, Duda JL, Dawes H, Johansen-Berg H. Effects of a programme of vigorous physical activity during secondary school physical education on academic performance, fitness, cognition, mental health and the brain of adolescents (Fit to Study): study protocol for a cluster-randomised trial. Trials. 2019 Apr 2;20(1):189. doi: 10.1186/s13063-019-3279-6.

Appendix A: EEF cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention over three years. More information about the EEF's approach to cost evaluation can be found **here**. Cost ratings are awarded as follows:

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: Progress Test in Mathematics (PTM), Level 13 (GL Assessment 2015).

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%					
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%			Adjustment for threats to internal validity **[0]**		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%	**1**				**1**
0	No comparator	>=0.6	>50%					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Low -Moderate	RCT design. Although there was some imbalance at baseline, and these were higher than 0.05 (0.06 for KS2 and absence sessions - denoting a 'moderate threat) the sensitivity analysis demonstrates this does not appear to have any notable effect on outcomes.
Threat 2: Concurrent Interventions	Low	There doesn't appear evidence or rationale that implementation of FtS is correlated with implementation of another programme. There were perhaps opportunities to more closely monitor/report conditions in comparison conditions.
Threat 3: Experimental effects	Low	School-level randomisation precludes contamination effects. No direct evidence for compensation rivalry or resentful demoralisation.
Threat 4: Implementation fidelity	Low - Moderate	Variations in training modes/ uptake and delivery of infusions are noted within IPE section. Although there is acknowledged limitations in fidelity data (retrospective recall), CACE analysis does not indicate differences as a result of compliance.
Threat 5: Missing Data	Low	Although missing data is notable, analyses do not indicate differences between complete case and imputed findings does not change interpretation of findings.
Threat 6: Measurement of Outcomes	Low	Outcome tests have good validity and reliability for the target population – marking was blinded.
Threat 7: Selective reporting	Low	SAP was pre-published, and changes to approach were noted in the SAP with reasons provided.

- **Initial padlock score:** [1] Padlocks – This was a randomised design with an MDES of 0.24. However, high attrition was the main limitations of the impact evaluation. The overall pupil-level attrition was 44% and reached 54% among the treatment pupils.

- **No adjustment for threats to validity:** Although there are some indications of some threat, these have mostly been addressed through are logically presented statistical analysis (e.g. sensitivity and CACE analysis), though there are opportunities to more carefully consider incremental validity of intervention in relation to usual practice.

- **Final padlock score:** 1 Padlock

Appendix C: Randomisation syntax

School randomisation.do - Printed on 04/18/2019 04:05:52 PM

```
1   /*
2   Fit to Study randomisation of schools
3   Nico Jabin
4   21/06/2017
5   */
6
7   cap log close
8
9   log using "randomisation_log", replace
10
11  use "I:\Workdocs\P12262 EEF_Fit to Study
    Trial\Secure\Randomisation\classsample_urind.dta", clear
12
13  local stratifiers "gender"
14
15  set seed 20170622 //from random.org
16  di c(seed)
17
18
19  *** remove schools that dropped out ***
20  *di sname if school==1085
21  *di sname if school==1023
22  *drop if school==1085 | school==1023
23
24  ********************************************
25  ****** assign schools                ******
26  ********************************************
27
28  /*
29  *** check stratifier groups ***
30  bysort `stratifiers': count //classes
31  preserve
32    collapse (first) `stratifiers' , by(school)
33    bysort `stratifiers': count //schools
34  restore
35  ** => drop rurality as stratifier
36
37  local stratifiers "gender"
38  bysort `stratifiers': count //classes
39  preserve
40    collapse (first) `stratifiers' , by(school)
41    bysort `stratifiers': count //schools
42  restore
43  */
44
45  gen double ranclass=runiform()
46
47  * allocate schools a random number equal to random number of 1st class
48
49  cap drop schid
50  bysort school: gen schid=(_n==1)
51  bysort school: gen ranschool=ranclass[1]
52
53  sort schid `stratifiers' ranschool
54
55  by schid `stratifiers': g treatment = (_n <= (_N * 0.5))
56  by schid `stratifiers': replace treatment = round(runiform()) if (mod(_N, 2)) & (_n==
    _N)
57  sort school schid
58  by school: replace treatment = treatment[_N]
59  replace treatment = . if exclude | !sampled
60
61  label define treat 0"control" 1"treatment"
62  label values treatment treat
63
64  preserve
65    drop if !sampled
66    collapse (first) treatment (sum)pupnum, by(school)
67    ta treatment, summarize(pupnum)
68  restore
69
70
```

Random selection of classes.do - Printed on 04/18/2019 04:06:52 PM

```
clear
clear matrix
clear mata
set more off
*version 14.1

**************************************
* Evaluation of Fit to Study
* Random selection of classes for testing
* Vainius Bartasevicius - February 2018
* The National Centre for Social Research (NatCen)
**************************************

log using "xxx.log", replace

cd "xxx"

*** Importing file ***

use "Fittostudy_classsample", clear


set seed 20180226

di c(seed)

g exclude = 0
g sampled = 1

**************************************
*** Exclude classes with one pupil *****
**************************************

replace exclude = 1 if pupnum==1

**********************************************
*** Exclude classes with few pupils if there are classes with
*** more than 10 pupils in the same school
**********************************************

sort schoolid pupnum
by schoolid: g maxclass = pupnum[_N]
replace exclude = 1 if (maxclass > 10) & (pupnum < 10)
drop maxclass
cap drop sampled

******************************************************
***** Select 49 percent of classes in each school *****
******************************************************

gsample 49 if !exclude [w=pupnum], percent strata(schoolid) wor g(sampled)
drop exclude

**********************************************************
**** Run summary statistics ************************
**********************************************************

summarize pupnum if sampled==1

di r(sum)

******************
* End of do-file *
******************

log close
```

Appendix D: Effect size formulae

Cohen's d effect size was calculated as follows:

$$d = \frac{(\bar{Y}_T - \bar{Y}_C)_{adj}}{\sigma_T}\sqrt{1 - \left(\frac{(N - n_U^T m^T - n_U^C m^C) + n_U^T + n_U^C - 2}{N - 2}\right)}$$

Where $(\bar{Y}_T - \bar{Y}_C)_{adj}$ is the difference between the treatment and control group means, controlling for baseline and stratification variables only. The denominator (σ_T) is calculated as the unconditional variance, that is, it has no variables entered other than the treatment effect indicator. This calculation assumes meta-analytic inference is aimed at the population of individuals. It is calculated, as follows:

$$\sigma_T = \sqrt{\frac{\sum_{i=1}^{m^T}\sum_{j=1}^{{n_i}^T}\left(Y_{ij}^T - \bar{Y}_{++}^T\right) + \sum_{i=1}^{m^C}\sum_{j=1}^{{n_i}^C}\left(Y_{ij}^C - \bar{Y}_{++}^C\right)}{N - 2}}$$

We converted to Hedge's g, following the usual formulae:

$$J(df) = 1 - \frac{3}{4df - 1}$$

$$g = j(df)d$$

$$v_g = [J(df)]^2 v_d^T$$

$$SE_g = \sqrt{v_g}$$

Confidence intervals were calculated using the 95% threshold, i.e. ±1.96SEg.

Appendix E: Effect size syntax

FtS primary analysis.do - Printed on 04/18/2019 02:29:38 PM

```
clear
clear matrix
clear mata
set more off
*version 14.1

log using "FtS primary analysis", replace

****************************************
* Evaluation of Fit to Study
* Primary Analysis: ITT Effect Size
* Vainius Bartasevicius - January 2019
* The National Centre for Social Research (NatCen)
****************************************

cd "P:\Working\Data merging"

use "Fittostudy_merged", clear

*****************************
*** Check the sample sizes and remove cases with missing values
*****************************

* Check sample size

ta treatment if Selectedfortesting==1, m

* Remove cases that have missing values needed for the analysis model

foreach var in treatment totalrawscore genderstatus baseline {

    keep if !missing(`var') & Selectedfortesting==1 & URN!=136592 & URN!=136883

}

* Check complete case sample size

ta treatment, m

*****
* 1. 2-Level Hierachical Linear Model
*****

mixed totalrawscore treatment cbaseline i.genderstatus || URN:, mle variance

*****
* 2. Treated and Control group sample sizes and raw means
*****

ta treatment, matcell(x)
local n1=x[2,1]  //Treated
local n2=x[1,1]  // Control
local N=x[1,1]+x[2,1] //Total
display _column(20) `n1' _column(40) `n2' _column(60) `N'

ttest totalrawscore, by(treatment)
esize twosample totalrawscore, by(treatment)


*****
* 3. Adjusted means
*****

mixed totalrawscore treatment cbaseline i.genderstatus || URN:, mle variance
matrix define k=r(table)
local Yadj_treated=k[1,5]  + k[1,1]
local Yadj_control=k[1,5]
display  _column(20)  `Yadj_treated' _column(40) `Yadj_control'

*****
* 4. Pooled unconditional variance
*****
```

FtS primary analysis.do - Printed on 04/18/2019 02:29:38 PM

```
 73
 74  mixed totalrawscore || URN: if treatment==1, mle variance
 75  matrix define z=r(table)
 76  local SEt=z[2,1]
 77
 78  mixed totalrawscore || URN: if treatment==0, mle variance
 79  matrix define y=r(table)
 80  local SEc=y[2,1]
 81
 82  display _column(20) `SEt' _column(40) `SEc'
 83
 84  local s_t=`SEt'*sqrt(`n1')
 85  local s_c=`SEc'*sqrt(`n2')
 86
 87  display _column(20) `s_t' _column(40) `s_c'
 88
 89  local pooled_variance = sqrt((((`n1'-1)*`s_t'^2)+(`n2'-1)*`s_c'^2)/(`n1'+`n2'-2))
 90  display _column(20) `pooled_variance'
 91
 92  ****
 93  * 5. Effect size calculations
 94
 95  esizei `n1' `Yadj_treated' `s_t' `n2' `Yadj_control' `s_c'
 96
 97  local cohen_d=r(d)
 98  local cohen_d_lb=r(lb_d)
 99  local cohen_d_ub=r(ub_d)
100  local hedge_g=r(g)
101  local hedge_g_lb=r(lb_g)
102  local hedge_g_ub=r(ub_g)
103
104  display _column(20) `cohen_d_lb' _column(40) `cohen_d' _column(60) `cohen_d_ub'
105
106  display _column(20) `hedge_g_lb' _column(40) `hedge_g' _column(60) `hedge_g_ub'
107
108  *****
109  * 6. Standard error of Cohen's D
110  *****
111
112  local cohen_d_SE = sqrt(((`n1'+`n2')/(`n1'*`n2')) + ((`cohen_d'^2)/(2*(`n1'+`n2'))))
113
114  display _column(20) `cohen_d_SE'
115
116
117  *****
118  * 7. Standard error of Hedge's G
119  *****
120
121  local hedge_g_SE = sqrt(((`n1'+`n2')/(`n1'*`n2')) + ((`hedge_g'^2)/(2*(`n1'+`n2'))))
122
123  display _column(20) `hedge_g_SE'
124
125  ******************
126  * End of do-file *
127  ******************
128
129  log close
130
```

Appendix F: Consent materials

MEMORANDUM OF UNDERSTANDING (MoU)

CONSENT FORM

Please sign both copies of the Consent Form, retaining one and returning the second copy to NFER in the pre-paid envelope provided. Alternatively, please scan and email the form to fittostudy@nfer.ac.uk

School Name:	

Please read the following statements and initial the boxes if you agree with the statements: *Please initial each box*

I confirm that I have read and understand the information sheet for the above study, I have passed a copy of the information sheet to my PE Dept., and I have had the opportunity to ask questions. ☐

This is a project that aims to test whether the nature of PE lessons affects pupils' fitness, wellbeing, cognitive function and academic attainment. I understand that my school will be randomly assigned to deliver our normal lessons or the new PE programme. ☐

I understand that my school's participation is voluntary and that I am free to withdraw at any time, without giving any reason. ☐

I agree to facilitate the activities involved in the 'Fit to Study' research project as described in the information sheet, including administering the fitness tests, online questionnaires and cognitive assessments at the beginning and end of the research period, and arranging for my Year 8 PE staff to attend the training if my school is randomly selected to deliver the new PE programme. ☐

I agree to provide information for, and facilitate the activities required for, NatCen Social Research's Independent Evaluation, including completing surveys, observations, interviews with relevant staff and pupils, and supporting the administration of pupil attainment tests. ☐

I agree to share background data (full name, gender, date of birth, Free School Meal eligibility, form group and UPN) about about my current Year 7 pupils (Year 8 2017/18 academic year) with NFER, the 'Fit to Study' research team, and the evaluation team at NatCen Social Research. ☐

I understand that the 'Fit to Study' research team will store information collected from individual members of my school either in a secure digital form, in a locked filing cabinet, or under identifier codes rather than names. Designated individuals from the research team and NatCen Social Research may view documents containing participants' names when monitoring or auditing the study. I give permission for these individuals to view this personal data. Maintenance of confidentiality of information is subject to normal legal requirements. ☐

I understand that the Fischer Family Trust will collate and anonymise pupil background and test data for upload to the UK Data Archive, to be made available in an anonymised form with restricted access for research purposes. ☐

I know whom I can contact if I have any concerns or complaints about the study. ☐

I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee. ☐

I agree that my school will take part in the above study. ☐

Name _____ Date _____ Signature _____ /P.T.O

MEMORANDUM OF UNDERSTANDING (MoU)

REPLY FORM

Please complete the following information and return to NFER using the pre-paid envelope provided or scan and email to fittostudy@nfer.ac.uk *Please complete Part 1, then Part 2 or 3 as appropriate.*

PART 1

	Are your details correct?	Please amend any incorrect details.
School Name		
Head Teacher		
Telephone		
Fax		
Email		

PART 2

My school **will** take part in this evaluation and agrees to the conditions stated in this Memorandum of Understanding (MoU)

Head Teacher/ SMT signature:.....

Name of nominated 'Fit for Study' contact in the school (ideally the Head of PE or a Member of the PE Dept.):

Mr/Mrs/Miss/Ms/Dr: .....

Job Title: .....

Contact Phone Numbers: Work.....Mobile:.....

Contact Email Address: .....

Please indicate the number of Year 7 students currently on roll and the number of classes

Number of Year 7 pupils	Number of Year 7 classes

PART 3

My school is unable to take part in this evaluation. ☐ *Please tick*

Thank you for completing this form. Please return the form to NFER using the pre-paid envelope provided, or scan and email to fittostudy@NFER.ac.uk

Education Endowment Foundation

Funded by the Education and Neuroscience Programme. Grant Ref: 2001

OXFORD BROOKES UNIVERSITY

nfer Evidence for Excellence in Education

NatCen Social Research that works for society

Appendix G: Example of a topic guide and observation pro-forma

NatCen
Social Research that works for society

Fit to Study: Evaluation

Early Implementation

Topic guide for staff in treatment schools – November 2017

Aim of the face to face interview:
The aim of the interview is to explore early implementation of the Fit to Study intervention in schools. The interviews will explore staff experiences of training and early implementation, and views on on-going delivery.

Face to face interviews will be carried out with participating staff in fifteen treatment schools in November 2017.

The topic guide:
This guide sets out a number of topic areas and broad themes to cover during the interviews. The interview is conversational in style and will develop and expand on issues brought up by the participant. The guide does not contain follow-up probes and questions like 'why', 'when', and 'how', etc. as participants' contributions will be explored in this way, as far as is feasible, during the interview. Researchers will use prompts and probes in order to understand how and why views and experiences have arisen.

The interview will last for **approximately 30-40 minutes.**

1. **Introduction**
   - Introduce yourself and NatCen Social Research
   - Introduce the study:
     - Evaluation of the Fit to Study intervention
     - Commissioned by the Education Endowment Foundation
     - Overall project aims:
       - To conduct a process study as part of a randomised control trial (RCT)
       - The process study focuses on early implementation and delivery
     - The purpose of these interviews is to gain an overview of Year 8 PE provision and to understand schools' experiences of the Fit to Study training and how they are going about incorporating the intervention into their Year 8 PE provision.
   - Interview is one of a number of interviews with teachers participating in the intervention.
   - Digital recording – check OK, and reassure re: confidentiality – data will not be shared between participants
   - Data kept securely in accordance with Data Protection Act
   - How we'll report findings – anonymity of teachers and schools
   - Reminder of interview length – 30-40 minutes. Check OK.
   - Any questions/concerns?

1

TURN ON RECORDER

2. **Background and context**
Aim: To gather background information on their role within the school

- Overview of role
  - Length of time in the school
  - Role and level of seniority
  - Teaching responsibilities (PE, other subjects, extracurricular)
- School context
  - Type of school (comprehensive; selective; academy)
  - Size (of the school as a whole, and Year 8 specifically)
  - Gender breakdown for Yr 8
  - Levels of disadvantage (e.g. FSM take-up)
  - Other contextual information

3. **Overview of PE in the school**
Aims: To gain an understanding of existing Year 8 PE provision, and any overlaps or differences with the aims of Fit to Study.

- Year 8 PE lessons
  - Priorities
    [ONLY ASK BELOW IF HAVE NOT COMPLETED SURVEY]
  - Frequency
    - Number of times a week
  - Length of time
  - How Year 8 PE is divided
    - Ability, gender, etc.

- Extracurricular sports provision, levels of participation / pupil engagement (if girls or mixed school – are there differences in uptake and type of sports engaged in)
- School encouragement to take part in sports
  - Do pupils engage in sports on weekends?

4. **Taking part in Fit to Study**
Aims: To explore decision-making regarding taking part in the intervention, motivations for participating and expectations for the intervention.

- How they first heard about the intervention
- Who signed the school up
  - Why
  - How the decision to take part discussed in school
    - Among PE staff
    - With other teachers
    - With pupils (if at all)

- Planning burden
  - Type of preparation and planning required – who involved?
  - Any modifications to PE lessons needed

- Initial impressions

- Understanding of Fit to Study when school signed up
- Information provided by developers
- What they expect to achieve by delivering the intervention

5. **Training**
Aims: To understand what support and training teachers need to deliver the Fit to Study intervention. To explore how well-equipped teachers felt to deliver Fit to Study.

- Whether attended Fit to Study training

FOR STAFF WHO ATTENDED FIT TO STUDY TRAINING

- Training
  - When was the training
    - Date/time of year (i.e. before summer holidays or in September/October)
  - Mode – online or face-to-face
    - If online: live session or pre-recorded session?
  - Where did training take place
  - Duration
  - How many PE staff many attended

- Cascading training **(if not all staff attended training)**
  - How discussed in school or staff meetings
  - Whether they are cascading training to other staff
  - How they are cascading training
  - Whether and what guidance and support they were given to cascade training to others
  - Any aspects they felt were difficult to cascade to colleagues

- Understanding
  - Aims of intervention
  - What mild to vigorous physical activity (MVPA) is
  - Understanding of how Fit to Study relates to MVPA

- Takeaways
  - What they understood that they were expected to do
  - How confident they felt about delivering the intervention
  - Perceived benefits to school/to pupils
  - Key challenges anticipated

- Views on training
  - What worked / What should change
  - Views on:
    - Format of training (i.e. online vs. face-to-face)
    - Content (balance between theory vs. practice)
    - Materials provided for training sessions
    - Resources provided to support delivery

FOR STAFF WHO DID NOT ATTEND FIT TO STUDY TRAINING

- If training was cascaded to them

- If yes:
  - Who cascaded training
  - How was training cascaded
    - Individual or department-level training
    - What (if any) resources were shared

- Understanding
  - Aims of intervention
  - What mild to vigorous physical activity (MVPA) is
  - Understanding of how Fit to Study relates to MVPA

- Takeaways
  - What they understood that they were expected to do
  - How confident they felt about delivering the intervention
  - Perceived benefits to school/to pupils
  - Key challenges anticipated

- Views on cascaded training
  - What worked / What should change
  - Views on:
    - Content (balance between theory vs. practice)
    - Materials provided during cascaded training
    - Resources provided to support delivery

6. **Ongoing support (for all participants)**

- Whether received further support
  - Requested this further support
  - Usefulness
- Whether they have they asked for more support
  - If yes:
    - Why
    - Who is providing it
    - How they access it

- Any support / resources they would like that are not currently available

7. **Early Delivery**

- When started delivery
  - If not started:
    - Why
    - When do they plan to begin
  - Whether all Year 8 PE lessons included
  - Any communication with pupils about intervention

- Whether Fit to Study is being delivered as intended (based on guidance/training)
  - Any changes being made to PE lessons
    - What and why [ON TOP OF / REPLACING EXISTING PROVISION?]
  - How closely they feel they are following:

        - The 10 minute warm up
        - The fitness infusions
    - What is working well / not so well

- What makes delivering Fit to Study lessons easy or more difficult
    - Whether Fit to Study is better suited to
        - Indoor or outdoor lessons
        - Particular sport or activity

8. **Reflections on the Fit to Study intervention**
   Aim: To explore views about potential benefits of the Fit to Study intervention for schools and pupils and identify potential facilitators and barriers to achieving these benefits.

    - Responsiveness to Fit to Study lessons
        - For:
            - PE teachers
            - Pupils
                - Do pupils notice any difference
                - Taking part in warm-up and infusions – differences among pupils (gender/level of fitness/LDD/health issues

    - Involvement with other PE interventions (external programmes and networks e.g. *This Girl Can*, sports networks) (*if applicable*)
        - What they are hoping to achieve with various interventions
        - How different/complementary to Fit to Study
        - How Fit to Study interacts with other interventions

    - Overall reflections and recommendations
        - Strengths / weaknesses of Fit to Study intervention
        - How different to usual PE lessons
        - Any changes planned to delivery throughout the year
            - If so, what and why
        - What value do they see to the intervention (compared to what they expected at the start)

9. **Ask if anything else to add**

TURN OFF RECORDER

Thank and close.

NatCen
Social Research that works for society

Fit to Study – Phase 2 - Lesson Observation Pro Forma (control)

School:	Researcher:
Teacher (name and position):	Date:
Start of lesson time:	Finish of lesson time:
Lesson plan provided by teacher: Yes / No (request lesson plan)	
A. Key features of class (number of children, level if set by ability, gender, any other relevant characteristics of pupils):	

B. Key features of lesson (sport focus, whether indoor or outdoor, weather conditions):
C. Describe how the lesson is introduced to pupils (i.e. focus on raising heart rate, having fun, building skills) and how they talk to pupils about raising MVPA/heartrate:

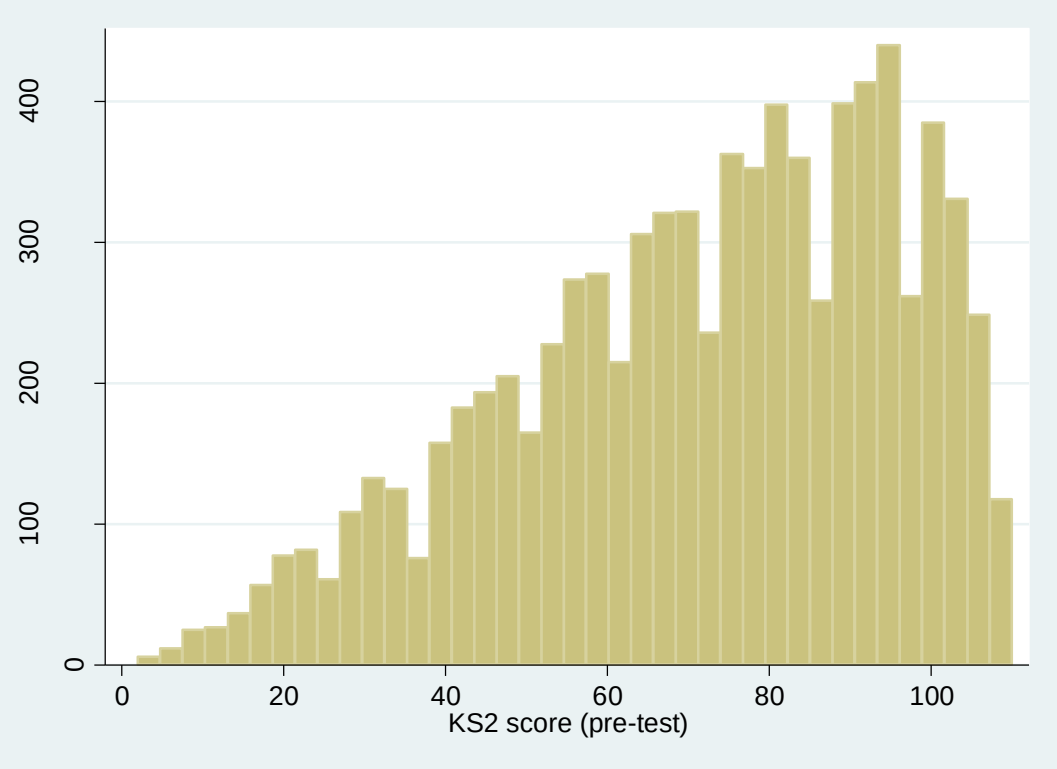
D. Describe the key elements of the lesson (e.g. registration, warm-up, small group activities), give timings (using stop watch). Record any other ways teacher seeks to ensure that pupils are 'as active as possible for the duration of the lesson'.	**Level of physical activity** (note where pupils are static, engaged in low level activity or in MPA and VPA; note any FtS type infusions).

E. Record any observations regarding barriers to / variation in pupils' engagement in MVPA (aerobic activity). Note whether (and which i.e. Boys? Girls? ...) pupils are struggling / less able to keep up (stopping in the middle of an activity; breathless; complaining). Record observations of teacher's reaction to pupils who are struggling.

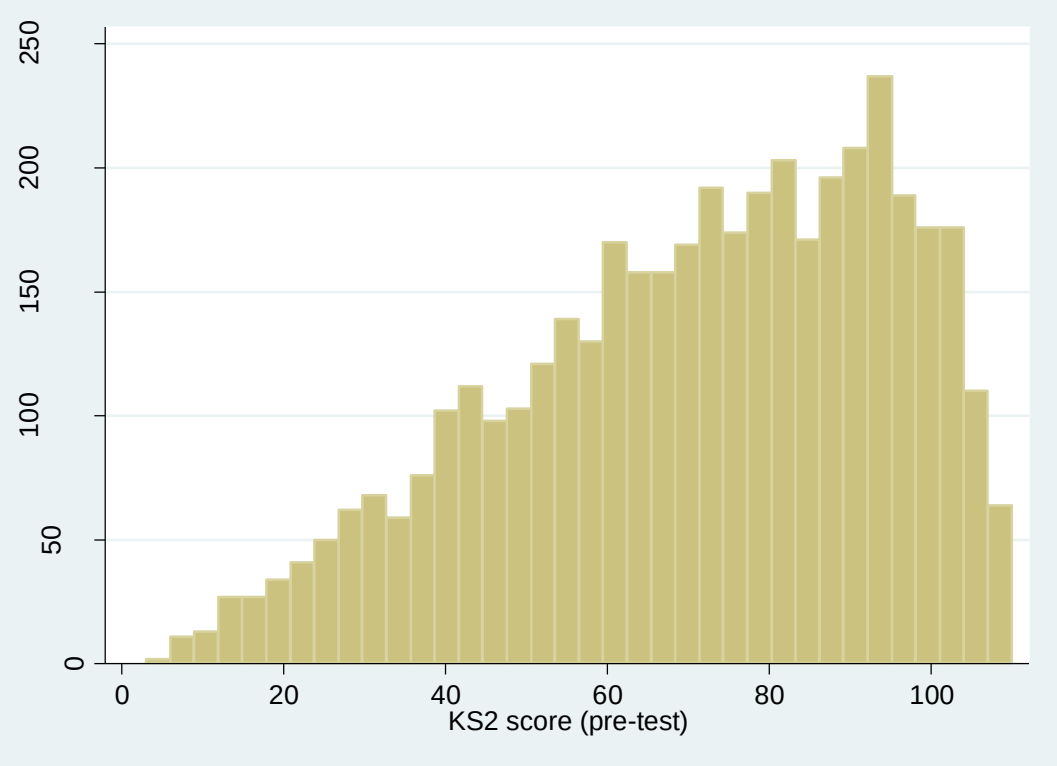
F. Other comments (e.g. any other observations to prompt on in interviews):

Appendix H: Histograms of pre-test scores

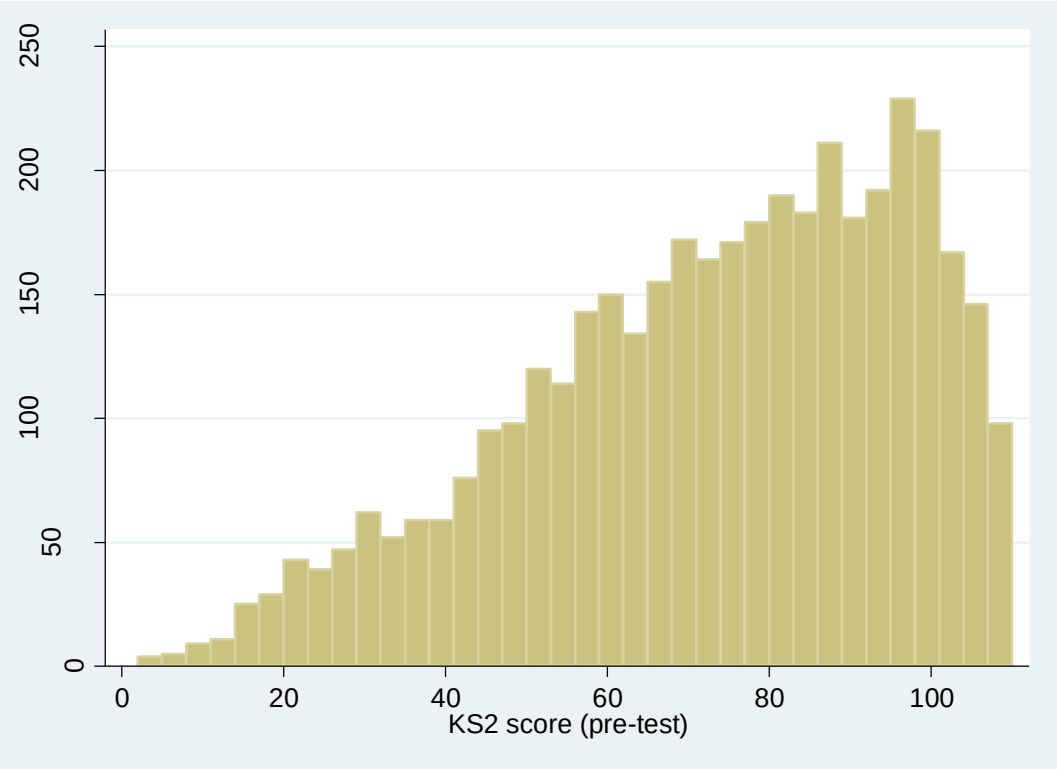
Distribution of pre-test scores in the total randomized sample:



Distribution of pre-test scores in the treatment group:

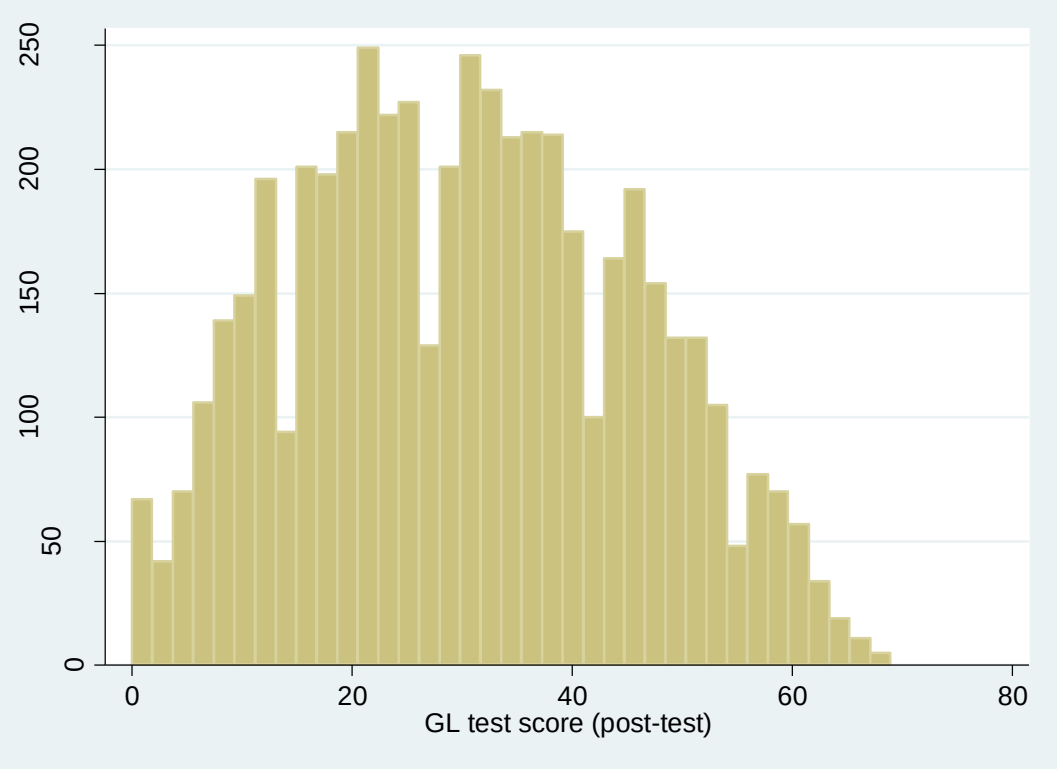


Distribution of pre-test scores in the control group:

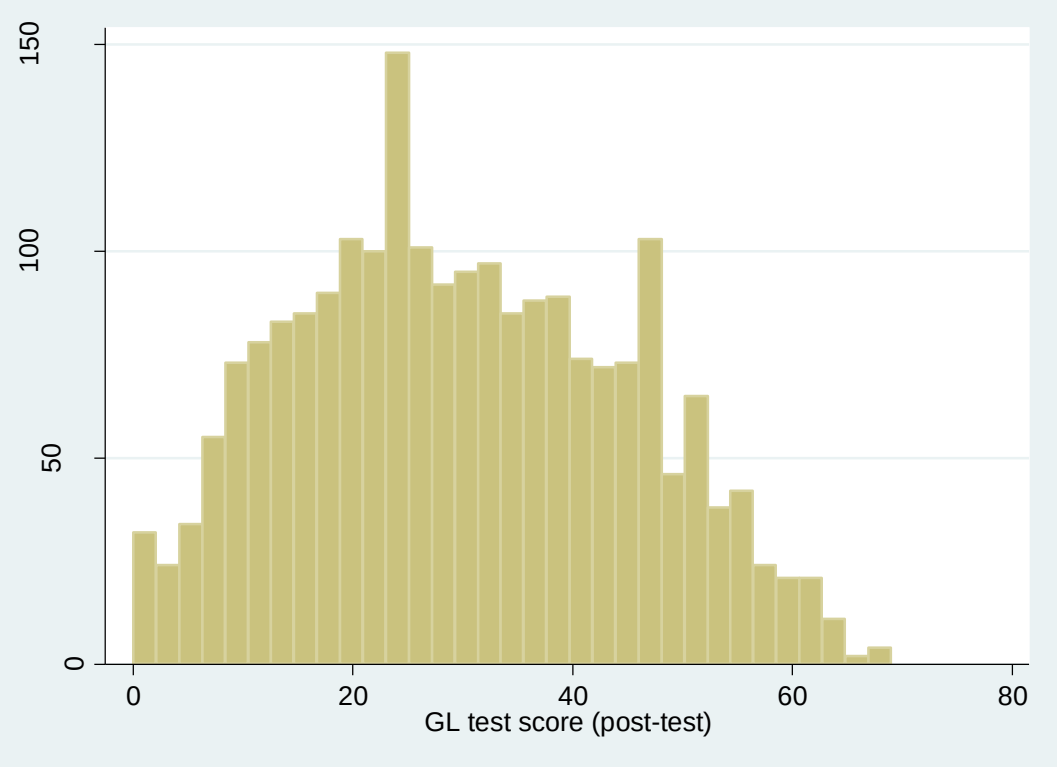

0
50
100
150
200
250
0
20
40
60
80
100
KS2 score (pre-test)

Appendix I: Histograms of post-test scores

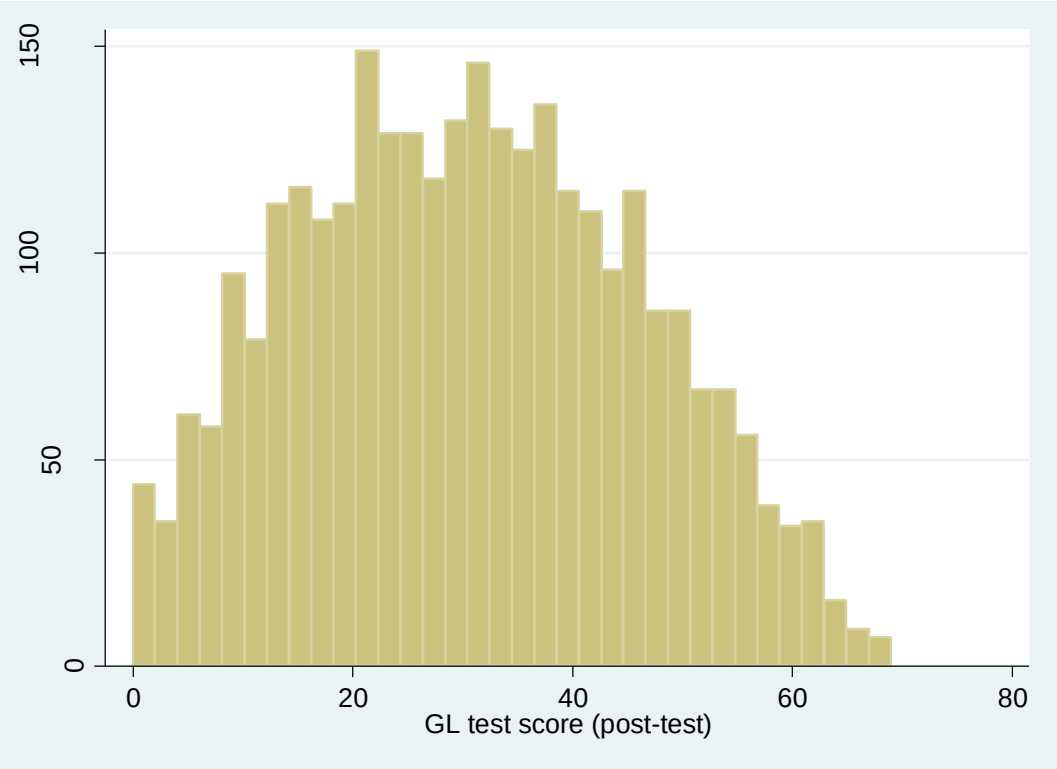
Distribution of post-test scores in the total randomized sample:



Distribution of post-test scores in the treatment group:



Distribution of post-test scores in the control group:


150
100
50
0
0
20
40
60
80
GL test score (post-test)

Appendix J: Privacy notice

NatCen
Social Research that works for society

Fit to Study – Privacy Notice

In line with the new EU General Data Protection Regulation (GDPR), there are certain things that we need to let you, a research participant, know about how your information will be processed. In this privacy notice, we explain the legal basis for data processing, who will have access to your personal data, how your data will be used, stored and deleted and who you can contact with a query or a complaint.

What is the legal basis for processing my data?
The National Centre for Social Research (NatCen) is the data controller and data processor for this project. This means that we are responsible for deciding the purpose and legal basis for processing data. For this project, the legal basis for processing data is "legitimate interest".

You can contact NatCen with any questions about the research at FitToStudy@natcen.ac.uk.

Who will have access to my personal data?
NatCen are carrying out this research and will have access to pupil, school and teacher information, recordings of interviews, transcripts, charted data, sample files (including contact details), survey responses and anonymised test results. All responses are anonymised before being analysed and archived.

Academics from the University of Oxford (http://www.ox.ac.uk/) and Oxford-Brookes University (https://www.brookes.ac.uk/), who are conducting some separate research, will also have access to pupil, school and teacher information.

McGowan Transcriptions (https://www.mcgowantranscriptions.co.uk/) is the transcription service we use to transcribe our interview and focus group data. They will have had access to recordings and transcriptions from all interviews and focus groups. McGowan Transcriptions is on our approved supplier list and compliant with all of our information security policies.

Formara Print+ (http://www.formara.co.uk/) is the printing company we use to print our materials. They will have had access to pupil names, dates of birth and UPNs for printing pupils lists and test paper labels. Formara Print+ is on our approved supplier list and compliant with all of our information security policies.

GL Assessment (https://www.gl-assessment.co.uk/) is the provider of the test papers we use on testing projects. They will have access to pupil names, dates of birth and UPNs for marking test papers. GL Assessment is on our approved supplier list and compliant with all of our information security policies.

How will my data be treated?
The results collected are used for research purposes only. Data gathered from interviews and surveys with teachers will be used to inform the process study element of the project, while pupil data and test results will be used for the impact study element.

At the end of the research, all pupil data will be anonymised before being archived. All personal information, and any other data held on the project, will be securely deleted once the project is complete in July 2019.

What are my rights?

- The right to be informed – You have the right to know about the purpose of this research. This is covered by this Privacy Notice, which has been sent out to all schools participating in this research, and our study page.
- The right of access – You have the right to see the data held about you on this project. Written or verbal requests to access the data held about you on this project will be replied to within one month.
- The right of rectification – You have the right to have inaccurate personal data rectified, or completed if it is incomplete. Written or verbal requests will be replied to within one month.
- The right to erasure – You have the right to have personal data deleted. Written or verbal requests will be replied to within one month. The right is not absolute and only applies in certain circumstances, for example, at the end of this project we will have no further need for your personal data and it will be securely deleted.
- The right to restrict processing – This is an alternative to the right to erasure. The right is not absolute and only applies in certain circumstances, for example, when a request is made to verify the accuracy of any personal information held about you the right to restrict further processing may be applied.
- The right to object – You have the right to object to the processing of your personal data. This right is not absolute and only applies in certain circumstances.
- Rights in relation to automated decision making and profiling – This project involved no automated individual decision-making or profiling.

More information on your rights under the General Data Protection Regulation (GDPR) can be found on the ICO's website.

Who can I contact with a query or a complaint?

If you have any questions about how your data will be used, please contact the NatCen's Data Protection Officer, Emmanuel Ogungbe, at Emmanuel.Ogungbe@natcen.ac.uk.

Under GDPR, you have the right to lodge a complaint with the Information Commissioner's Office. Please go to www.ico.gov.uk for more information.

Appendix K: Training content

Training for Fit to Study comprised 5 different modes, not all of which were delivered to each school:

1. Face-to-face training
2. Online livestream training
3. A video recording of the online livestream training
4. Cascaded training
5. 'Top-up' training

Face-to-face training comprised the scientific rationale for the research, what schools would need to do to deliver the intervention, and the requirements for trial participation. This mode of training also involved practical elements to demonstrate possible infusions for various sport-specific PE lessons. Teachers were able to ask questions and clarify their understanding at any point throughout the session.

Online training contained the same elements as the face-to-face training, bar the practical elements. It did, however, include opportunities for teachers to send in questions via email to clarify their understanding and ask questions.

Video recorded training covered the same elements as the online training but did not allow for any interaction between participant and trainer, as in the online livestream training. This video was disseminated to those unable to attend the online livestream training.

Cascaded training involved teachers who had either attended the face-to-face training or livestream training passing on any knowledge of the intervention to other teachers in their PE department as required. Teachers reported a preference for covering only the perceived salient points for delivery, an for example, mentioned omitting the theoretical background to the intervention.

'Top-up' training was delivered by visiting developer staff as part of their initial schools visits in the Autumn term of 2017 in order to ensure that all schools were aware of the intervention activities and how to deliver them and answer any questions related to implementation of the intervention activities. For the most part this training included addressing any concerns regarding the training in a responsive manner, as opposed to delivering any segments of the face-to-face training programme.

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

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

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn