Education Endowment Foundation

Writing About Values

Evaluation report and executive summary

November 2018

Independent evaluators:

Dr Beng Huat See, Dr Rebecca Morris, Professor Stephen Gorard and Dr Nadia Siddiqui

Durham University

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

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

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

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children. This programme was co-funded by the Department for Education as part of an EEF funding round on Character Education.

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

Danielle Mason
Head of Research

Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP
p: 020 7802 1679
e: danielle.mason@eefoundation.org.uk
w: **www.educationendowmentfoundation.org.uk**

About the evaluator

The project was independently evaluated by a team from Durham University:

Dr Beng Huat See, Dr Rebecca Morris, Professor Stephen Gorard and Dr Nadia Siddiqui.

The lead evaluator was Dr Beng Huat See.

Contact details:
Dr Beng Huat See
School of Education
Durham University
Leazes Road
Durham
DH1 1TA

Email: b.h.see@durham.ac.uk

Contents

Executive summary

The project

Writing about Values aimed to improve the academic performance of disadvantaged Year 10 and 11 pupils by reminding them of their important values. During English lessons, pupils wrote reflective essays about core values, such as relationships with friends and family, sport or music. These writing exercises aimed to remind pupils of positive aspects of their lives, and were administered by English language teachers at the beginning of the academic year, before mock GCSEs, and just before the actual GCSE exams began. Research suggests that an awareness of negative stereotypes about the academic performance of disadvantaged pupils can cause harmful feelings to these pupils and have a negative impact on academic outcomes. The project is based on the theory of self-affirmation, which suggests that engaging in value affirmation writing activities can give individuals a positive sense of value and negate these harmful feelings, fostering academic learning and improving performance.

In this randomised controlled trial, Year 10 and 11 pupils in 29 secondary schools in the South East of England were randomly allocated to either an intervention or a control group. Teachers and pupils were not told which pupils were in each group, or about the theory behind the intervention, because there is evidence that knowledge of the purpose of this type of intervention can reduce its effectiveness. This was achieved by administering the writing exercises in plain individually-named envelopes, and giving the project the generic title of 'Writing about Values'. While all pupils participated in the trial, the target participants were pupils from disadvantaged backgrounds, defined as those eligible for free school meals in the past 6 years.

A team from the University of Sussex implemented the intervention, drawing on research evidence from studies in the U.S. An implementation and process evaluation assessed whether the writing exercises were administered with fidelity. The trial began in September 2016 with the final exercise taking place in May 2017. This is the first report, discussing the impact of the intervention on GCSE results for 5,619 Year 11 (age 15–16) pupils who took their GCSEs in May/June 2017. A 2019 report will present the GCSE results for the initial Year 10 cohort. The programme was co-funded by the Department for Education as part of an EEF funding round on Character Education.

Key conclusions	
1.	Among disadvantaged pupils, those who received the self-affirmation intervention made slightly more progress between the end of primary school and GCSEs than the comparison pupils, but the size of the impact was very small and further analysis suggests the impact was close to zero. This result has a high security rating.
2.	Pupils who completed more writing exercises made slightly more progress. This may mean that the intervention can lead to better outcomes if implemented more thoroughly, but might also be because the kind of pupils who completed more exercises would make more progress anyway.
3.	Neither pupils nor teachers were aware of the nature of the intervention or why the tasks were undertaken because there is some wider evidence that knowledge about the purpose of the intervention can reduce its effectiveness. Schools considering this approach should bear in mind the difficulty of replicating these conditions.

EEF security rating

These findings have high security. The trial was an efficacy trial, which tested whether the approach worked under developer-led conditions. The trial was a well-designed, two-armed randomised controlled trial. It was well powered and relatively few pupils (155 out of 1506 disadvantaged pupils) who started the trial were not included in the final analysis. There were some important differences in prior attainment between the pupils allocated to intervention and those allocated to control. This can be

taken into account in the analysis, but choices about how to do so affect whether the size of the impact estimate is just about one month's progress (as reported in summary table) or zero.

Additional findings

The intervention had a small positive impact on the progress that disadvantaged pupils made from KS2 to GCSE, approximately equivalent to one month of additional progress, although the impact reduced to zero when the analysis controlled for pupils' prior attainment. There is no evidence that the intervention benefitted all children.

There was a small positive correlation between the number of writing exercises completed and the outcomes. This might be because the kind of pupils who completed more exercises would make more progress anyway. An analysis comparing the progress of those who completed the first writing task to similar pupils in the control groups showed little difference between groups.

The exercises were delivered as intended in most schools, with teachers successfully integrating them into the start of English lessons. Some teachers appreciated the opportunity for pupils to undertake some 'free writing', without any assessment constraints.

Cost

The cost of running the one-year intervention is minimal, estimated at £1.89 per pupil (assuming 300 pupils per school) per year when averaged over three years.

Table 1: Summary of impact on primary outcome

Group	Effect size	Estimated months' progress	EEF security rating	No. of pupils	EEF cost rating
GCSE Attainment 8 EverFSM6	0.05	1	🔒🔒🔒🔒🔓	1,351	£~~£~~~~£~~~~£~~

Introduction

Intervention

The Writing About Values project aimed to improve the academic performance of disadvantaged GCSE pupils. It is based on research about 'self affirmation', though this term was not used to describe the project publicly as the theory suggests that knowledge of the intended outcomes could compromise its efficacy. The theory states that disadvantaged pupils are aware of stereotypical beliefs about how pupils like themselves perform academically, and that this awareness can itself detract from their performance. Undertaking self-affirming activities, such as writing about values that are important to them, can help to protect pupils' self-worth and free up cognitive resources to engage more effectively with their learning.

Based on studies from the U.S., this project tested the impact of writing short essays during English lessons, with individual pupils randomised between intervention groups (writing essays about values that are important to them, such as friendships and honesty) and control groups (writing essays about values important to other people).

Detailed description of the intervention

1. **Brief name:** Writing About Values

2. **Why: Rationale, theory and/or goal of essential elements of the intervention**

   The intervention was developed by the University of Sussex and is based on the self-affirmation theory that individuals are motivated to maintain a sense of self-worth, and when that self-worth is threatened, they typically engage in defensive behaviours to protect their self-worth. This is particularly so with some disadvantaged groups. The theory is that getting these individuals to engage in self-affirming activities, such as writing about values that are important to them, can help give them a sense of value and thus alleviate negative feelings associated with their perceptions of themselves (stereotype threat). Previous evidence (for example, Steele and Aronson, 1995; Sherman and Cohen, 2006) suggests that this can free their cognitive resources to allow them to engage in learning. An important element of this intervention is that neither the teachers nor the pupils should know what the intervention is about as there is some evidence that knowledge about the aim of the intervention can interfere with its efficacy. Therefore all efforts are made to avoid revealing this to them. The aim of the intervention is to reduce the impact of stereotype threat on academic performance by raising self-integrity via a writing activity.

3. **Who: Recipients of the intervention**

   The recipients of the intervention are secondary school pupils in Year 10 (age 14–15) and Year 11 (age 15–16) from socio-economically disadvantaged backgrounds. EverFSM6 (children eligible for free school meals in the last six years) was used as a measure of socio-economic disadvantage because this was the definition underlying the Pupil Premium.

   (Note that this report presents the results for the Year 11 only. Year 10 results will be reported in 2019 when the GCSE results are available).

4. **What: Physical or informational materials used in the intervention**

   The intervention involved three writing activities which were presented in booklets that were placed in named envelopes and distributed to pupils individually. Teachers gave a short structured introduction to the task and further instructions were on the booklets. The first writing task involved the treatment group writing about values that are important to them, and the

control group writing about values that are not important to them but that might be important to other people. A list of values was provided for the pupils from which they chose two or three to write about. Examples of such values include: enjoying sports, being honest, and relationships with friends. For the second writing activity, the treatment pupils wrote about things/people that matter to them, while the control pupils wrote about things they did that morning. In the third writing exercise, treatment pupils selected from a list of values those that are important to them and were asked to write about what they will do to show that these are important to them or how much they enjoy doing them. These values could be relationships with friends, having a sense of humour, being with family, and following government and politics. Apart from the short writing exercise, pupils continued as per normal in their regular English lessons. See Appendix C1 and C2 for examples of the writing exercises.

5. What: Procedures, activities and/or processes used in the intervention

The intervention involves students reflecting on important personal values in a series of short structured writing exercises. These are open tasks and allow pupils to respond however they choose. Pupils are told that their work will not be marked or read by their teachers, and that they do not need to focus on grammar or spelling; content and ideas are more important. The idea is to get them to write freely about their thoughts, beliefs and views.

The writing exercises (see Appendices A1 and A2 for examples) were delivered during regular English language lessons. The writing exercise booklets were placed in named envelopes and distributed to pupils individually. Teachers gave a short structured introduction, explaining that the task is a writing exercise focusing on pupils' own thoughts and ideas and that there are no right or wrong answers to the task. They explained that it is about the process of doing the activity rather than the teacher providing feedback so the exercise will not be read or marked. All instructions were on the booklets, so there was minimal input from the teachers on the completion of the task. All completed booklets were then placed in the envelopes, handed to the teachers and collected by the delivery team. All teachers involved were trained to use the materials and administer the writing exercises, but not told what the intended outcome of the intervention was in order to avoid contamination. This is important because previous research has shown that knowledge about the purpose of the exercise reduces its efficacy (Sherman et al., 2009). Teachers and pupils were also blind to treatment groups as all pupils had a writing task to do—different only in terms of content. It was also important that the writing exercise was presented by the teachers as part of the regular English activity so that pupils see it as part of a creative writing task and not associate it with an external research study.

Teachers were given scripts to introduce the task and also pre-prepared responses to questions that pupils are likely to ask. These scripts were in the form of FAQs (see Appendix D). The researchers who observed the delivery of the intervention were also blind to treatment conditions as all pupils were involved in a writing exercise that differed only in terms of content. The two exercises are very similar (see Appendix C1 and C2).

6. Who: Intervention providers/implementers

The intervention is developed by a team of social and developmental psychologists at the University of Sussex who adapted the workbooks, training materials and teacher instruction sheets from those previously used in studies in the U.S. (such as Sherman et al., 2013). They also conducted all the teacher briefing sessions. The writing tasks were delivered by English language teachers in the participating schools. All workbooks were placed in named envelopes and collected from pupils by the teachers and picked up by developers.

7. How: Mode of delivery

The writing exercises were delivered by English language teachers as a whole-class activity and as part of their regular English lessons. Efforts were made to ensure that these exercises were delivered as naturally as possible to avoid pupils linking them to a research project.

8. Where: Location of the intervention

The intervention was delivered in the pupils' usual English language classroom.

9. When and how much: Duration and dosage of the intervention

The writing exercises were administered three times in a year, once at the beginning of the academic year, once before mock GCSEs and once just before the actual GCSEs exams commenced. This is in line with current research suggesting that these writing activities are most effective if the first exercise is administered at the start of the academic year and subsequent exercises just before a stressful event, such as before final exams. Each writing activity was expected to be approximately 10 to 20 minutes.

Although not part of the intervention, a questionnaire survey was administered to measure the impact on pupils' self-efficacy. To ensure that pupils do not associate the survey with the intervention and also that the survey does not interfere with the intervention, the post-survey was delivered after the final writing task.

10. How well (planned): Strategies to maximise effective implementation

To maximise effective implementation a number of strategies were used to safeguard the covert nature of the intervention. For example, observation visits were kept to a minimum to avoid pupils linking it to a research project. It was also planned that the same schools were not visited for both observation of the delivery of the exercises and for the administration of the survey questionnaire so that pupils did not link the survey to the writing tasks. Where evaluators visited the schools to observe the delivery of the exercises, it was explained to them that the visitors were there to observe how English was being taught in other schools. Feedback from pupils was obtained only from the Year 11 and only after their GCSE exams. Interviews with teachers were conducted only after the third writing exercise and only in very general terms about the writing activity itself, and not about the specifics or theory of the intervention. Teachers were also given strict instructions to use prescribed answers to pupils' queries about the purpose of the exercises (see A3).

Furthermore, unlike previous EEF projects, the protocol and the statistical analysis plan for this trial were not published on the EEF website and no reference was made to the actual name of the intervention (Self-Affirmation). Instead the intervention was referred throughout as Writing about Values.

11. How well (actual): Evidence of implementation variability

In practice, the developers and the evaluators were fairly successful in keeping the nature of the intervention from both pupils and teachers. For example, briefings to schools were presented in a very general way and although some reference was made to the evidence-based nature of the intervention, teachers were not made aware of the full background of the intervention. The focus of the briefing was on the delivery of the exercises, and how teachers should ensure that pupils were not aware that they were taking part in a research project. However, while writing exercises were delivered by English language teachers, the survey was administered by the form tutors. In one school, some form tutors were aware that the survey and the writing tasks were part of the University of Sussex research project, although they were instructed beforehand not to make this connection known to pupils. Form tutors were contacted

immediately by the developers to instruct them not to raise pupils' awareness about this connection.

Some teachers were able to incorporate the writing task into their normal exam preparation. Interviews with teachers after the third writing exercise also revealed that they were not aware of the specific nature of the intervention.

However, there was some ambivalence about the use of the sealed envelopes. A couple of pupils in one school were suspicious about the envelopes, commenting that they looked conspiratorial, like a 'kind of social experiment'.

Before the main trial, a pilot was conducted to test the writing materials, training manuals, and the delivery procedure (Appendix E).

Background evidence

There is already considerable policy and practice activity being undertaken on the assumption that an individual's aspirations, attitudes and behaviour (motivation, self-concept, self-belief and locus of control) can be influenced to improve educational outcomes. Most of the research conducted in this area is based on correlational studies using path analysis as a proxy for causality (for example, Marsh and Martin, 2011). The evidence of a causal effect remains unclear. The question is whether pupils with high motivation or aspiration perform well, whether high performance leads to higher motivation, aspiration and self-belief, or whether both are a consequence of something else. The ongoing debate about the sequence of events cannot be resolved without some more closely controlled and independent trials. This was one of the main recommendations in the report to the Joseph Rowntree Foundation based on a review of 166,000 studies (Gorard, See and Davies, 2012). The self-affirmation project attempts to contribute to this developing evidence base.

The theory is that the writing activities give pupils a sense of value, alleviating negative feelings associated with their perceptions of themselves. Initial effects might be that they feel less threatened, more confident and this can affect peers' and teachers' expectations to do better. Previous evidence suggests that self-affirmation interventions have positive and long-term results on improving academic achievement (for example, Good et al., 2003; Cohen et al., 2009).

The advantage of this approach is that no stigma is attached to individual pupils and the cost of delivery is minimal apart from the initial training of teachers and the costs of printing the exercise booklets and the teacher manuals.

The aim of the project was to evaluate the impact of the self-affirmation intervention on reducing the effect of negative stereotype threat on academic performance. The intervention is based on the hypothesis that students from some stigmatised groups are aware that they are the target of a negative stereotype regarding their academic performance (Steele, 1997). This can (a) lead to anxiety about confirming this negative stereotype during school assessments, which can undermine performance, or (b) elicit a defence mechanism, known as 'disidentification', in order to protect their self-concept from being devalued by the negative stereotype. Disidentification results in academic achievement being discounted or devalued (Crocker and Major, 1989; Major et al., 1998), and can reduce learning and motivation.

The self-affirmation strategy has been employed to alleviate the effects of stereotype threat on low performing students, especially those from ethnic minority backgrounds, (Oyserman et al., 2006; Miyake et al., 2010) by getting them to write positive statements about themselves (Sherman and Cohen, 2006; Steele, 1988). It has been found that this can help ameliorate the detrimental effects of stereotype threat on academic performance. If this approach is found to be effective in raising

attainment for disadvantaged children it can prove to be very attractive as it is almost cost-free, simple to implement, and would appear to generate few, if any, contra-indications.

Most of the studies conducted so far are based in the U.S. and the results are mixed but promising showing that it is particularly effective in raising the attainment of ethnic minority groups (Cohen et al., 2009; Cohen et al., 2014; Sherman et al., 2013). Cohen et al. (2009), for example, found that although there were no overall gains in grade point averages across four core academic subjects in both treatment and control groups, the African-American pupils in the treatment group improved their Grade Point Average (GPA) score by 0.24 points, and the low-achieving African-American students by 0.41. The intervention also appeared to reduce the likelihood of low-achieving African-American students being assigned to a remedial programme or retained in grade. A longitudinal experiment (Sherman et al., 2013) also showed that Latino American students using the 'Writing About Values' approach earned higher grades and were more likely to continue to higher education than those in the control. White students were not similarly affected. A recent and larger study (Borman, Grigg and Hanselman, 2015) also showed a small positive impact on minority pupils' standardised maths test scores. A positive effect was also reported for female college students from more advantaged backgrounds (Mikaye et al., 2010). The women in the self-affirmation group achieved significantly higher grades than women in the control group. The gender gap between men and women was also reduced with the women moving up from average to above average range.

On the other hand there are studies which suggest no effects on either academic or other outcomes. Simmons' (2011) study on 47 African-American high school students, for example, found that students taught the strategy did not achieve higher GPAs nor were psychologically more engaged. A potentially important difference between these studies is in the timing of the intervention. Unlike Cohen et al.'s (2009, 2006) and Miyake et al.'s studies (2011), Simmons administered the intervention some time after the beginning of the term, whereas Cohen and his colleagues administered the intervention very close to the start of the term before students had the opportunity to experience negative stereotype influence. Cohen et al. and Miyake et al. also administered the intervention immediately before or after a threatening event and in the regular classroom, while Simmons administered the intervention in a different setting from their regular lesson (for example, a cafeteria or another classroom). Also students were offered monetary incentives to complete the post-measure and this may have reduced the stereotype threat for the participants. Other studies suggest that the writing exercise alone is not enough. A supportive classroom environment is needed for the intervention to have any impact (Dee, 2015).

Most of these studies were conducted in the U.S. and focus on African-American or Latino-American students or college women. No independent randomised controlled trial of such an intervention had been conducted in the U.K. on the academic outcomes of the general student population. This efficacy trial is the first large-scale randomised controlled trial ever conducted in the U.K. to test the causal effect of the self-affirmation theory on academic attainment.

Evaluation objectives

The objective of the trial was to evaluate the impact of the self-affirmation intervention on the academic performance of Year 10 (age 14–16) and Year 11 (age 15–16) pupils in England. This report presents the attainment outcomes for the Year 11 cohort for which the KS4 results were available. The second report, which will be published in 2019, will report the attainment and self-efficacy outcomes for the Year 10 cohort. Since the self-efficacy survey had to be taken after the final exams (to avoid compromising the effects of the intervention), the Year 11 did not take the post-survey after their GCSE when most would have left school. Only the Year 10 cohort completed both the pre- and post-survey. For this reason, only the Year 10 self-efficacy outcomes will be analysed and this analysis will be published in the 2019 evaluation report.

The research questions are:

1. What impact does the self-affirmation intervention have on the academic attainment of disadvantaged pupils, defined as those who have been eligible for FSM at some point in the last six years (EverFSM6), using the individual pupils' Attainment 8 measure at GCSE after one year of treatment (for initial Year 11 group)?

2. Is there a sustained impact of the self-affirmation intervention on the Attainment 8 measure at GCSE for EverFSM6 pupils after two years (one year after the end of the intervention)? (This analysis is based on the initial Year 10 group.)

3. What impact does the self-affirmation intervention have on all pupils (EverFSM6 and non-EverFSM6) using the individual pupils' Attainment 8 measure at GCSE after one year of treatment (for initial Year 11)?

4. Is there a sustained impact of the self-affirmation intervention on the Attainment 8 measure at GCSE for all pupils (EverFSM6 and non-EverFSM6) a year after the end of the intervention (for initial Year 10)?

5. Is there a sustained impact of the self-affirmation intervention on pupil's self-reported self-efficacy?

Research questions (1) and (3) are answered in this report. Research questions (2), (4) and (5) will be answered in the 2019 report.

Ethical review

The evaluation, as distinct from the intervention, raised few additional ethical issues. The attainment outcomes are based on the GCSE exams which happen as a matter of course in schools, so did not pose any ethical issues. Additional information about pupils' self-efficacy was collected via a survey questionnaire. All participants in interviews and observations were informed that participation was voluntary and that they could withdraw consent at any stage. The evaluation was conducted in accordance with the British Educational Research Association's professional Code of Practice and approved by Durham University's ethics committee on 25 August 2015. The study also had the approval of the University of Sussex's Sciences and Technology Cross-Schools Research Ethics Committee. These committees ensure that all pupil and assessment data is treated with the strictest confidence; that no individuals or schools would be identified or identifiable; and that all results are reported in aggregated form. The data would be shared by Durham University with the Education Endowment Foundation data archive.

In accordance with the ethical guidelines, opt-out consent was sought from parents. This was used to indicate agreement to participate. Pupils whose parents opted out were offered the control writing exercise, but all data from these pupils were excluded from the analyses. School level agreement was collected via signed Memoranda of Understanding (see Appendices B and C).

Project team

The project involves two teams of people: the project developer, also known as the delivery team and the evaluation team.

Delivery team

The delivery team is made up of staff from the University of Sussex, led by Dr Matthew Easterbrook. Other team members include Professor Peter Harris, Professor Robin Banerjee, Dr Marlon Nieuwenhuis and Dr Kerry Fox. They were responsible for the design of the intervention, recruitment of schools, training of teachers and the design and delivery of the self-efficacy survey. They had first

contact with the schools and were responsible for collecting the parental consents and other data from schools.

Evaluation team

The evaluation was carried out by staff from Durham University, led by Dr Beng Huat See who managed the project, including arranging fieldwork, managing communications with the EEF and the delivery team, and the impact analyses. Dr Rebecca Morris was responsible for the process evaluation report, and Professor Stephen Gorard was responsible for the design of the trial and the randomisation process. Dr Nadia Siddiqui supported the research team in the development of the final report. A team of postgraduate ad hoc researchers assisted with data collection for the process evaluation. They included Laurence Droy, Eszter Newmann, Szilvia Schmitsek and Richard Barrie.

Trial registration

The trial was registered on 30 June 2016. The trial registration number is: ISRCTN79754465

Methods

Trial design

This was a two-year, double-blind randomised controlled efficacy trial involving Year 10 (age 14–15) and Year 11 (age 15–16) pupils from 29 schools in the South East of England in the academic year 2016/2017. It is double-blind in that both pupils and teachers were not aware of the results of the randomisation process. It was possible for both teachers and pupils to be blinded because both treatment and control groups had to do the writing activity, which differed only in terms of the content. This is equivalent to the control group having a placebo. The two writing activities are very similar with only a slight difference, so it is not obvious to teachers or pupils which is the intervention and which is the placebo.

Evaluation of impact for Year 11 was undertaken at the end of the first year following release of their GCSE results, whereas impact evaluation for the Year 10 cohort will be at the end of the second year (2018) to test the sleeper or sustained effect of the intervention. For this reason, two reports will be generated: one for the Year 11 cohort and one for the Year 10 cohort a year later. This report is for the Year 11 cohort only.

Pupils were individually randomised within schools, stratified by year group and FSM status to either receive the intervention (writing about values important to them) or an alternative exercise (writing about things that are not important to them, but that might be important to other people).

Participant selection

Schools were recruited by the team from the South East of England (areas in and around Sussex and Buckinghamshire) using a combination of strategies: contacting schools in neighbouring localities, through local authorities, academy chains and school improvement advisors' network of schools; canvassing at premium leads and heads of school meetings; county council secondary school conferences, and via direct contacts from schools through the EEF websites. It is therefore not possible to indicate the number of schools approached.

The targeted schools were those not in special measures, with a minimum of 10% of pupil population eligible for FSM. Priority was given to schools with a high proportion of FSM children.

Once schools confirmed participation, a Memorandum of Understanding outlining the intervention and the trial was sent to school heads who then signed an agreement to comply with the requirements of the trial, including releasing teachers for training, delivering of the writing exercises and administering the self-efficacy survey (see Appendix F). To minimise dropout after randomisation, all participating schools were offered an incentive payment of £1,000 for completing the trial, delivering the three writing exercises and administering the non-attainment surveys.

Opt-out consent was sought from parents through the school before randomisation. The Project Team delivered information letters and opt-out consent slips to participating schools. The schools distributed the information letters and opt-out slips to parents/carers, and were responsible for collecting these for the delivery team. The procedure for distributing and collecting opt-out forms is clearly explained in the MOU.

Eligible pupils were all Year 10 and Year 11 pupils in the 29 schools. The consent form (see Appendix G), drafted by the delivery team, provides information about the intervention and what participation entails. Parents were informed that if they opted out, their child could still take part in the intervention (the writing exercises) but their data would not be used in the evaluation.

Outcome measures

Academic attainment and self-efficacy outcomes were chosen as the outcomes of interest because the basic tenet of the self-affirmation intervention is that it counters the negative stereotype threat faced by disadvantaged groups. This enhances their self-worth and self-efficacy, which in turn, improves their academic attainment. The self-efficacy measures are therefore to test the linked mechanisms between improvement in self-efficacy and academic attainment. The EEF is interested in the outcomes of disadvantaged pupils, and it was agreed that the definition of 'disadvantaged' would be 'pupils ever eligible for FSM in the last six years' as this is the definition underlying the Pupil Premium eligibility. However, based on prior work conducted by the developers and others (for example, Bowen et al., 2012; Cohen and Sherman, 2014; Stephens et al., 2012), there is some evidence that the intervention may be more relevant to historically low performing groups, such as ethnic minority groups in the U.S., women in STEM subjects, and those from lower income groups. Stephens et al., for example, found that the intervention closed the achievement gap between students from lower socio-economic backgrounds and their more advantaged peers by 50%. Bowen et al.'s study suggests that the intervention can prevent a decline in performance among low-income pupils. In line with this evidence, the intervention was expected to have a positive effect on the traditionally negatively stereotyped group, such as the FSM pupils in U.K. For this reason, an additional analysis was conducted for this group of pupils.

For this report, only the attainment outcomes for the Year 11 cohort are analysed. The second report to be published in 2019 will report on the attainment and the self-efficacy outcomes for the Year 10 cohort when the GCSE results are available.

Primary outcomes

The primary outcomes were:

- Attainment 8 at GCSE for FSM pupils (based on EverFSM6) after one year of treatment (for initial Year 11 pupils); and
- Attainment 8 at GCSE for FSM pupils (based on EverFSM6) after two years (one year after the end of the intervention) for pupils who received the intervention when they were in Year 10 (to be published in the 2019 report).

Secondary outcomes

The secondary outcomes were:

- Attainment 8 at GCSE for all pupils (including both EverFSM6 and non-Ever FSM6) after one year of treatment (for initial Year 11 pupils); and
- Attainment 8 at GCSE for all pupils (both EverFSM6 and non-Ever FSM6) after two years (for pupils who received the intervention when they were in Year 10) (to be published in the 2019 report).

It was decided after the original protocol was written to also evaluate the impact of the intervention on pupils who are currently eligible for free school means because there is some evidence from theory and previous work by the developers that the intervention may be more relevant to these pupils than those who have been on FSM at some point in the last six years. Therefore, the secondary outcomes also include:

- Attainment 8 KS4 scores for FSM pupils (based on current FSM status) after one year of treatment (for initial Year 11 pupils); and
- Attainment 8 KS4 scores for FSM pupils (based on current FSM status) after two years (one year after the end of the intervention) for pupils who received the intervention when they were in Year 10 (to be published in the 2019 report).

Non-attainment outcome

The non-attainment outcome was pupils' perceived self-efficacy, measured using the subscales of the Motivated Strategies for Learning Questionnaire (Pintrich et al., 1993). These non-attainment surveys were designed, collected and electronically marked by the delivery team, but analysed by the independent evaluators.

Other data

Besides the attainment and non-attainment outcomes, pupils' background characteristics such as age, date of birth, sex, ethnicity, first language, and SEN were also collected from schools as a routine part of being entered into the randomisation. This data was uploaded for all pupils at the outset from each school's SIMS or similar. These were eventually linked via Unique Pupil Number (UPN) to the individual post-test scores.

Pupils' prior attainment in KS2 maths and reading (point scores) and FSM status were collected from the National Pupil Database (NPD). The scores were used as a pre-test to establish baseline equivalence.

Sample size

The sample size calculation was based on the assumption that there would be 25 schools and two year groups (Years 10 and 11). Assuming an average of five forms in each year group (which is typical in secondary schools), there would be 125 forms (25 X 5) per year group. Working on the assumption of an average of 30 pupils per form, there would be 3,750 pupils (30 X 125) for each year group (or 7,500 in total). Randomising individual pupils to treatment conditions, there would be 1,875 pupils in each arm for each year group (or a total of 3,750 in each arm). Assuming around 25% of EverFSM6-eligible pupils overall, this would mean around 470 EverFSM6 pupils per arm per year group (or 940 per arm of the trial).

Traditional power calculations are based on the approach of significance testing (Gorard et al., 2017) which is misleading. They are therefore not included here. Instead, we calculate the sample size needed for any 'effect' size to be considered secure by considering *a priori* the number of 'counterfactual' cases needed to disturb a finding (Gorard and Gorard, 2016). This number needed to disturb (NNTD) is calculated as the 'effect' size multiplied by the number of cases in the smallest group in the comparison (that is, the number of cases included in either the control or treatment group, whichever is smaller). This approach allows for estimating ES and sample size using the following formulae:

NNTD = ES*n

Therefore, n = NNTD/ES and

ES = NNTD/n

This is a useful measure of the scale of the findings to chance (and their variability as represented by the standard deviation used to compute the 'effect' size), taking into account the scale of the study. It can then be extended to compare this sensitivity directly to other more substantial sources of error such as the number of missing values/cases. The number of cases actually missing a value can be subtracted from the NNTD to give an estimate of how large the 'effect' size would be even in the extreme situation that all missing cases had the 'counterfactual' score hypothesised in the NNTD calculation. Here the 'counterfactual' score is one standard deviation away from the mean of the group with the largest number of cases. The standard deviation would be added if the mean of the smaller group (in scale) were smaller than the mean of the larger group, and subtracted if the mean of the smaller group was the largest. (Gorard et al., 2017).

Based on Gorard et al. (2016), NNTD of 50 can be considered a very strong and secure finding. Using this as a working assumption, the number of cases needed in each group (assuming equal size) to detect an 'effect' size of 0.2 (which is typical for an education intervention) will be 250 (or 50/0.2). This is assuming no attrition.

The NNTD calculation concerns the security of a difference, and so is relevant to internal validity only. Issues such as clustering are therefore irrelevant. In addition, as pupils were individually randomized within schools and analysis would be of all pupils in the two groups and not by schools, clustering effects, if there were any, should be evenly spread between the two groups across all schools.

This report concerns the initial Year 11, so all calculations will be based on this cohort only. Assuming 25% of pupils are EverFSM6 (n = 470 per arm), we would expect to be able to detect an 'effect' size of 0.11, or 50/470 (ES = NNTD/n). In reality 29 schools were recruited and 5,619 of the pupils were in Year 11 (n = 2,810 treatment and n = 2,809 control). Of these, 26.8% (or 1,506) were known to be eligible for EverFSM6 (n = 706 treatment; n = 800 control). This will enable us to confidently detect an 'effect' size of +0.07. This is different to that reported in the Statistical Analysis Plan, which combined the sample for both Year 10 and 11 pupils. (Note: The original plan was to recruit 25 schools, but because recruitment was done in the summer term, it was anticipated that some schools might drop out over the summer. To allow for this the developers over-recruited. It was agreed with the EEF that recruitment would stop when 30 schools had provided the pupil data by the deadline).

However, whatever the sample size, it is important that all allocated cases are retained as any attrition can bias the results. Over-sampling in order to cater for subsequent attrition does not minimize bias.

Randomisation

Randomisation was at the individual level, stratified by year group and FSM status. This meant that year groups were randomised separately, and FSM and non-FSM pupils separately within each year. So, in effect, there were four randomisations: first by year group and then by FSM eligibility. This was to ensure equivalence in treatment groups. Stratifying the groups by FSM status makes it possible to estimate the valid effects within that group. Schools were not informed of the results of the randomisation, but since the treatment and control pupils were to be given different writing tasks, the developers were informed of the result of the randomisation immediately in order for the named exercise booklets to be printed on time.

As soon as pupil data from all participating schools was received, randomization was carried out using a random number generator on Excel and conducted in the presence of members of the evaluation team (Professor Stephen Gorard and Dr Rebecca Morris). Schools provided a list of all the Year 10 and Year 11 pupils with their background data for randomisation. This had to be carried out in the summer before the Autumn term started and before the precise number of pupils could be confirmed because the intervention necessitates that the first writing exercise had to be administered at the start of the school term. And since the two intervention groups would be doing different exercises the developers needed to know the number of pupils in each arm of the intervention so that the right number of exercise booklets and named envelopes could be printed before term started. As randomisation was carried out before the new school year began, it was anticipated that some pupils might leave and others might arrive. It was therefore agreed with the EEF that only schools that completed the baseline survey would be included in the trial and analysis.

Analysis

A copy of the Statistical Analysis Plan is available in Appendix H.

Primary analysis

This analysis reports the findings based on the first cohort of Year 11 pupils for which KS4 results were available. The Year 10 pupils' results will be reported in a follow-up report in 2019 when their KS4 results become available. The results of the self-efficacy survey will also be published in the 2019 report.

Primary intention-to-treat (ITT) analysis

The primary analysis is intention-to-treat. Those pupils randomised to receive the intervention were analysed regardless of whether they received the intervention or not. This means that a school that was included in the randomisation but then decided not to participate in the trial was also included in the analysis. The primary impact evaluation is based on the difference between groups in terms of the gain scores between KS2 results for maths and reading and KS4 GCSE Attainment 8 outcomes. The differences are expressed as effect sizes (Hedge's) and converted to progress in months. Significance tests and confidence intervals are not reported here but the latter can be calculated given that the means, standard deviations, and effect sizes are reported. We have not calculated any p-values as they are not relevant (for further explanation, please refer to Gorard, 2016; Cohen, 1994; Colquoun, 2014, 2016; Trafimow and Rice, 2009; Perezgonzalez, 2015; Pharoah et al., 2017).

Although pupils were individually randomized within schools, there was no issue of clustering as analysis would be of all pupils in the two groups and not by schools. Any clustering effects, if any, would be evenly spread between the two groups across all schools. The mean scores of all the pupils in the control group and treatment group in the schools would be the same as the mean scores of all treatment and control pupils in the whole trial.

Imbalance at baseline

To establish baseline equivalence we used the 'effect' sizes for each measurement at the outset, and we also present the characteristics of schools in each group. To cater for any initial imbalances between groups we also present the gain scores analysis. For the benefit of readers we present the pre, post, and gain score 'effect' sizes, regardless of imbalance.

Missing data

Missing cases or missing data are to be expected in all real-life research. We cannot assume that these are random. For example, some pupils may not have a test score because they were on long-term sick leave, excluded from school, refused to take the exam, or had learning difficulties. Some may have come from overseas or from independent schools and so would not have KS2 results. Dong and Lipsey (2011) demonstrated that any missing values can create bias, even if attrition is balanced between comparator groups. And where such attrition is not random (as is most often the case) it can bias the estimate of the treatment effect, and the bias can still be large even when advanced statistical methods like multiple imputations are used (Foster and Fang, 2004; Puma et al., 2009). Such bias can threaten the validity of any conclusion reached (Shadish, Cook and Campbell, 2001; Campbell and Stanley, 1963; Little and Rubin, 1987). Therefore any attrition has to be taken seriously in randomised controlled trials. Dong and Lipsey suggested using baseline covariates to reduce bias introduced by attrition. However, the condition is that these covariates must be correlated with the outcome variable and the propensity to respond. In reality it is difficult, if not impossible, to determine if these are correlated because we do not know why some people may be missing post-test scores. Others suggest substituting existing data for the data that is missing, but since we have little or no knowledge of the missing cases, doing this will simply increase the potential for bias. We therefore present differences in pre-test scores (KS2 maths and reading) between cases dropping out from both groups (where these are available).

To decide whether the missing cases would have altered the outcome, we estimate what happens if all these missing cases had counterfactual scores. For example, if the outcome is positive, we estimate

what the results would be if all the missing cases had negative scores. To do this we first calculate the number of counterfactual cases needed to disturb the headline finding (see Gorard and Gorard, 2016). The number of counterfactual cases determines whether the number of missing cases is large enough to alter/explain the findings. The number of counterfactual cases is calculated using the effect size multiplied by the number of cases in the smaller group minus the number of missing cases. The bigger this number is the more stable is the result as this means it will take this many counterfactual cases to reduce the effect size to zero (see section on Sample size above).

Fidelity analysis

Two analyses were carried out. The first is a correlational analysis comparing the outcomes of pupils with the number of exercises completed (dosage). The number of exercises is used as a continuous variable in the analysis. This will be zero for all cases in the control group. In addition, the regression analysis (see additional analyses below) also includes the number of exercises completed as a predictor.

To estimate the effects for the subgroup of treatment students who complied with their treatment assignment, the Complier Average Causal Effect (CACE) analysis was performed. Comparison is made of the average outcome of treatment pupils who complied with control children who would have complied if given the treatment (Nicholls, undated; Dunn, 2010). Compliance is defined as completion of the first writing exercise (as defined by the developers) because theoretically the first writing exercise is supposed to be the most impactful (Cohen and Sherman, 2014; Cohen et al., 2012; Garcia and Cohen, 2012) as it is expected to trigger a recursive adaptive response to a threatening environment in a feedback loop. For example, if a student performs/behaves better as a result of the first activity, their self-confidence may improve and their teacher may have higher expectations of them. This could lead to better performance and the process perpetuates itself. The second and third exercises are meant to provide the boost to this process. It is more difficult to trigger a positive response later in the year once expectations set in. Therefore, it is important that pupils complete the first writing exercise.

Given that we know the overall results for both groups and the data for those in the treatment group who complied and who did not comply (cells labelled A to K in Table 2), we can calculate the average outcome for those in the control group who would have complied if given the treatment. We assume that because of randomisation, the proportion of compliers in both arms of the trial is the same (on average), and the average outcome for those in the control group who did not comply (I) will be the same as the outcome of non-compliers in the treatment group (D). We may conclude:

- proportion in treatment group who complied is A/E;
- number in control group who would have complied (G) will be A/E*J
- number of non compliers in control group (H) = J-G
- the average outcome for compliers in the control group (x) is calculated thus:

  x = ((K*J - H*I)/G))

Table 2: Estimation of Complier Average Causal Effect

Participants	Compliers		Non-compliers		All	
	N who completed first writing task (proportion of who complied)	Mean	N who did not complete first writing task (proportion who did not comply)	Mean	Total N	Mean
Treatment	At	B	C	D	E	F
Control	Ac = A/E*G	*x*	Cc=G-Ac	D	G	H

Secondary outcome analyses

Secondary outcome analyses are comparisons of pre, post, and gain score 'effect' sizes for:

- all pupils (that is, both EverFSM6 and non-EverFSM6) after one year; and
- FSM pupils (based on current FSM status) after one year of treatment (for initial Year 11 pupils).

Additional analyses

Two multivariate regression models were created to estimate how much of a difference being in a treatment group made. Both used the post-test scores (Attainment 8 KS4 scores) as the dependent variable, and total prior test scores (KS2 maths and reading) and membership of treatment group as predictors. One was based only on EverFSM6 pupils (equivalent to the headline analysis using gain scores), and the other using all pupils.

An additional regression analysis was performed as a test of fidelity using KS2 maths and reading results and the number of exercises completed or dosage (with zero for control pupils) as the predictors and KS4 Attainment 8 scores as the dependent variable. This additional analysis was specified in the Statistical Analysis Plan (see Appendix H).

Effect size calculation

'Effect' sizes are calculated as the difference between mean post-test (and gain scores) for each variable, expressed as Hedges' *g*. We do not report 'confidence intervals' but this can be easily computed if any reader wishes to do so as the number of cases for each group, and the effect size for each comparison are reported.

For categorical variables in the 'self-efficacy' questionnaire, we calculated the 'effect' sizes based on post-intervention odds ratios—or changes in odds where the groups are clearly unbalanced at the outset. All of these are presented with the number of counterfactual cases needed to disturb the results. The results of the self-efficacy will be reported in the second report.

Process evaluation

The process evaluation was conducted by the independent evaluators, in cooperation with the developers, to gather formative evidence on all phases and aspects of the trial from the selection and retention of schools, through initial training and delivery of the intervention, to evaluating the outcomes. All data was collected by the evaluators, unless otherwise stated. The purpose was to assess fidelity to treatment, identify potential for contamination, as well as to gauge participants' perceptions of the intervention and to see if the activity gave any clues to teachers about the intervention. The latter is particularly relevant to this trial as any knowledge of the purpose of the intervention can affect motivation and thus negate any potential effect (Sherman et al., 2009).

The aims of the process evaluation are to assess:

- the fidelity to training;
- teachers' delivery of the intervention;
- the contents of the writing exercises;
- staff and student's views of the intervention;
- possible indication of contamination or diffusion; and
- barriers and challenges to implementation.

The main method of data collection was classroom observations of the delivery and administration of the self-efficacy surveys and writing exercises. These were conducted by the independent evaluators and were carried out to be as integrated and non-intrusive as possible to minimise disruptions to classroom activities. The classroom observations were to see whether teachers stuck to the scripts, that the right pupils were given the correct writing exercise, and if there was any possibility that pupils could swap exercises with their classmates. It was decided with the EEF and the delivery team that a sample of 12 classes in six schools would be selected at random for the classroom observations. The schedule of visits was agreed with the project team and the schools. However, observation visits were made to only five schools as it was difficult to get schools to agree to our visit due to difficulties in organisational and logistic arrangements within the schools.

Observations of surveys were conducted in five schools to see if, in general, they were carried out consistently across schools, if there were any potential irregularities or questions from pupils, and pupils' reaction to the question items. It was important that the writing exercise was seen as part of the regular English lesson (and not part of research). Different schools were visited for the survey and classroom observations. Feedback about the survey was also collected from pupils. We asked pupils:

- what they thought of the questionnaire;
- whether there were any questions which were ambiguous or which they found difficult to answer; and
- whether there were any questions they had not answered and why.

We also had conversations with teachers to find out if they had observed any changes in pupil behaviour and to gather their views on the writing exercises. The intervention necessitates that participants should have no knowledge of the intervention because being aware of the purpose of the intervention can interfere with its potential impact. Therefore, the conversations with teachers had to be carefully conducted in order to avoid directly mentioning the name of the intervention or its outcomes. Since the trial (but not the intervention) was to continue for another year (to observe the longer term effects on the initial Year 10 cohort), the conversations with the teachers were also partly to find out if teachers and pupils had any knowledge of the intervention. As interviews with pupils were not possible while the trial was still running, feedback from the pupils could only be collected after their GCSE exam. This was to minimise the potential of interference with the intervention. As pupils effectively left school after their GCSE exams, emails were sent out to pupils through the schools inviting them to respond to a short questionnaire in which they were asked:

- if they could recall the writing exercise;
- what aspect of the exercise they could remember;
- what was the most memorable information that they have learned from the activity; and
- if they could see any value in the writing exercise.

Information about dosage was collected from the project delivery team who kept a log of the number of exercises completed by each pupil.

Costs

The approach to estimating costs is based on the EEF cost guidance.[1]

The cost of running the intervention is estimated based on information provided by the delivery team. This includes:

[1] **https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_guidance_to_evaluators_on_cost_evaluation.pdf**

- cost of printing the pupil exercise booklets, teacher manuals for the training, teacher scripts used for administering the exercises, and the named envelopes;
- staff time in sorting out the exercise booklets in envelopes and time spent photocopying and collating the booklets; and
- cost of initial training for teachers.

The cost per pupil is estimated by dividing the total amount by the number of pupils/teachers per school. Training costs will be incurred only in the first year but the costs will be spread over three years.

Timeline

The Pilot started in Jan 2016 in five schools. The main trial began in September 2016 with 29 schools.

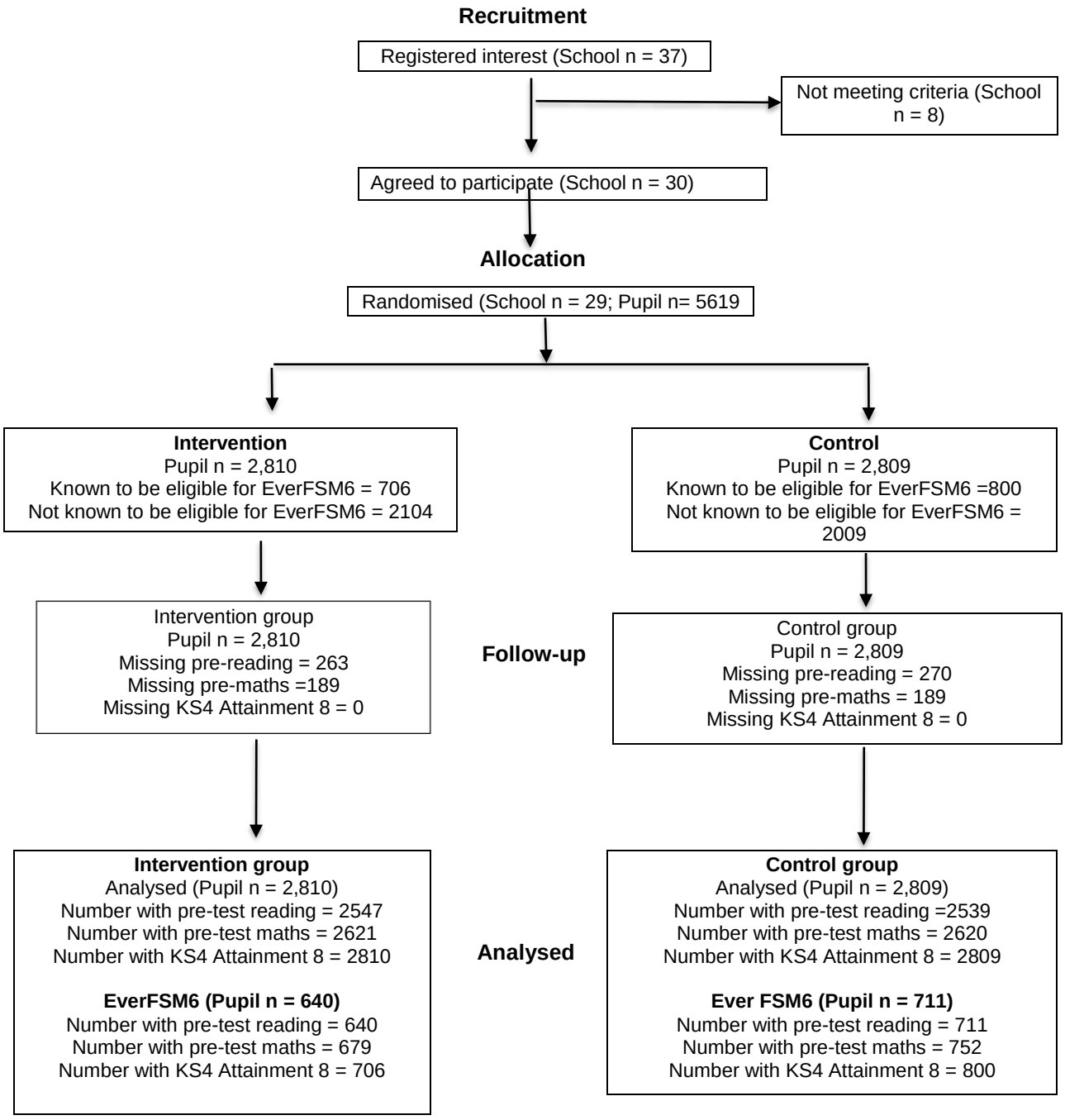
Table 3: Timeline

	Date	Activity
	July 2015	• First meeting with the EEF and delivery team to discuss the intervention, implementation and timescale for the pilot and main trial
	January 2016	• Recruitment of 'friendly' secondary schools by the delivery team to help in the formative process of finalising materials for the pilot
	November 2015 to March 2016	• Recruitment of schools for the pilot and the main trial running concurrently carried out by the delivery team
Pilot Phase	February to March 2016	• Delivery team collect pupil data from schools for pilot • Randomisation of pilot pupils conducted by the evaluators
	April 2016	• Observation of training of teachers in pilot schools (evaluation team) • Collect and evaluate training and teaching materials (evaluation team) • Administer pilot non-attainment survey instrument (delivery team)
	May 2016 (before GCSE)	• Delivery of treatment in pilot schools (delivery team) • Light touch observation of delivery of intervention in pilot schools (evaluators) • Interviews with teachers and pupils to identify potential risks (evaluators) • Collect feedback on the survey instrument from pupils and teachers (evaluators)
	May to July 2016	• Revision of survey instrument by project delivery team
	August 2016	• Protocol updated to include self-efficacy as the non-attainment measure for the main trial (evaluators)
Preparation for Main Trial	December 2015	• Development of recruitment materials by project delivery team
	January to June 2016 (while pilot is going on)	• Observe the recruitment of schools to main trial (evaluators)
	June/July 2016	• Project delivery team arrange for schools to sign MOU, send out parental consent forms to schools and collect pupil background data (UPNs, names, and FSM status) • Evaluation team carried out randomisation in July and results revealed to the delivery team

		immediately to allow for the writing exercise booklets to the printed over the school holiday
Main Trial begins	September 2016	• Conduct non-attainment survey (undertaken by the delivery team in the presence of evaluation team) • Observe training of teachers by evaluators • Delivery of first writing task by the delivery team • Ongoing light touch process evaluation of delivery of intervention by evaluators
	December 2016/January 2017	• Delivery of second writing exercise before mock GCSEs by the project delivery team
	May 2017	• Delivery of third writing exercise before GCSEs by the delivery team
	June to July 2017	• Conduct follow-up non-attainment survey for only Year 10 pupils after school exams (by delivery team) • Interview Year 11 pupils and teachers by evaluators
	July to August 2017	• Collect and enter survey responses by project delivery team • Analyse non-attainment outcome conducted by the project delivery team
	October 2017	• Apply for KS4 attainment results from NPD • Report writing
	April 2018	• Draft report for Year 11 cohort submitted on 17 April

Impact evaluation

Figure 1: Flow chart of participants through the study[2]



[2] Despite the two-level randomisation (on FSM and year group) there is still imbalance between groups. This is because randomisation was carried out before pupils and schools were confirmed to facilitate the printing of the writing booklets. When the trial started one school was excluded as it did not complete the baseline survey (a condition for inclusion in the trial). Additionally, 99 pupils have neither KS2 nor KS4 results, and of these, 69 did not take the baseline survey. These were therefore excluded from the trial. As a result, the initial balance between the two groups is slightly skewed.

Attrition

There was a total of 5,619 Year 11 pupils in this trial. Post-test (GCSE Attainment 8) scores were available for all pupils with pre-test data. The pupils without a pre-test measure were excluded from the analysis. A number of pupils did not have KS2 results for either maths or reading or both because they did not take the KS2 test (for example, they could be from independent schools or from overseas) or were absent. No school dropped out of the trial.

Table 4: Minimum detectable effect size at different stages

Stage	N [schools/pupils] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Power	Alpha	Minimum detectable effect size[3] (MDES)
Randomisation (All)	29 schools 5,619 pupils (2,861; 2,860)	0.67	0.00	80%	0.05	0.06
Randomisation (EverFSM6)	29 schools 1,506 pupils (706; 800)	0.60	0.00	80%	0.05	0.12
Analysis (EverFSM6)	29 schools 1,351 pupils (640; 711)	0.60	0.00	80%	0.05	0.12

Pre-test is the standardised combined KS2 reading and maths scores.
Post-test is KS4 Attainment 8 score.
The columns for power, alpha, and MDES are required by the EEF, although all are meaningless in this context.
Intra-cluster correlation is 0. There is no clustering because analysis is on the individual level and not by schools.

Pupil and school characteristics

Table 5 shows that the characteristics of the pupils in the trial schools are broadly similar to those in secondary schools in England, although they are slightly more likely to have a higher proportion of EverFSM6 and SEN pupils. The trial schools tend to have a lower proportion of pupils achieving five A*–C GCSEs compared to the national average. They are also more likely to have a higher proportion of White British pupils and lower proportion of EAL pupils. The trial schools are, on average, bigger schools compared to the average secondary schools in England. This reflects the success of the recruitment strategy—specifically to target bigger schools with higher than average proportions of disadvantaged pupils.

[3] MDES calculated using PowerUp tool: Dong, N. and Maynard, R. (2013) 'PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies'. http://web.missouri.edu/~dongn/PowerUp.htm

Table 5: Comparison of trial schools and all state-funded secondary schools in England

Variable	All secondary schools (N= 3,401)		Trial schools (N=29)	
School-level categorical variables	n	%	n	%
Academy converter	1,408	41	6	21
Academy sponsor	586	17	8	27
Community	594	17	9	29
Foundation	272	8	3	10
Voluntary aided	282	8	2	7
Voluntary controlled	45	1	1	3
Others (free schools, City Technological Colleges)	214	6	-	-
Ofsted Rating	**All state-funded secondary N= 3,148 inspected as at 31 Mar 2016)**		**Trial schools n/N (missing)**	
Outstanding	700/3,148	22%	4/29 (0)	13%
Good	1,686/3,148	54%	16/29 (0)	57%
Requires improvement	634/3,148	20%	7/29 (0)	23%
Inadequate	128/3,148	4%	1/29 (0)	3%
No information	-	-	1/29 (0)	3%
School-level (continuous)	**All secondary schools**	**Mean**	**n/N (missing)**	**Mean**
Size of schools	3148	939	29/29 (0)	1162 (6 schools have fewer than 800 pupils)
Pupil-level (categorical)	**All secondary schools**	**Percentage**	**Trial schools n/N (missing)**	**Mean (%)**
Proportion achieving 5 A*–C (including English and Maths)	3,148	57.1	29/29 (0)	51.1
Proportion of pupils eligible for FSM	3,148	13.2	29/29(0)	14.3
Proportion of pupils eligible for FSM EverFSM6	3,148	29.3	29/29(0)	32.0
Proportion of pupils with SEN	3,148	12.7**	29/29(0)	14.1

Proportion of pupils with SEN statement or EHC plan	3,148	1.7**	29/29(0)	1.7
Proportion of pupils with first language not English	3,148	15.7	29/29(0)	12.1
Proportion of pupils who are White British ethnicity	3,148	70.9	29/29(0)	76.0

Data from Department for Education 2015/2016 Performance Tables: http://www.education.gov.uk/schools/performance/download_data.html and 2016 Annual Schools Census https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2016. Ofsted ratings for intervention schools are taken from the latest inspection reports. National data for Ofsted ratings is based on inspections completed between 1 September 2015 and 31 March 2016, taken from https://www.gov.uk/government/statistics/maintained-schools-and-academies-inspections-and-outcomes-as-at-31-march-2016.

Table 6 shows that the two achieved groups are balanced at pre-test. There is no difference between groups for reading (ES = 0), with the intervention group being only slightly ahead for maths (ES = +0.03). The EverFSM6 pupils in the intervention, however, are slightly behind at pre-test for both reading (ES = -0.06) and maths (ES = -0.05).

Table 6: Comparison of pupils' baseline characteristics

Characteristics of pupils at randomisation	Intervention		Control	
	n/N (missing)	Percentage	n/N (missing	Percentage
Proportion of boys	1430/2810 (0)	50.9	1393/2809 (0)	49.6
Proportion of pupils eligible for EverFSM 6	706/2810 (0)	25.1	800/2809 (0)	28.5
Proportion of current FSM pupils	305/2795 (15)	10.9	373/2788 (21)	13.4
Proportion of pupils with SEN	421/2810 (0)	15.0	422/2809 (0)	15.0
Proportion of pupils whose first language is not English	314/2808 (2)	11.2	338/2800 (9)	12.0

	Raw means				
	Intervention		Control		
	n/N (missing)	Mean (sd)	n/N (missing	Mean (sd)	Effect size
KS2 Maths (point scores)	2621/2810 (189)	27.8 (5.14)	2620/2809 (189)	27.7 (5.23)	+0.02
KS2 Reading (point scores)	2547/2810 (263)	30.4 (9.68)	2539/2809 (270)	30.4 (9.57)	0.00
KS2 Reading and Maths combined	2547/2810 (263)	58.6 (12.99)	2539/2809 (270)	58.5 (12.93)	+0.01

EverFSM6 pupils					
KS2 Maths (point scores)	679/1431 (27)	26.2 (5.15)	752/1431 (48)	26.5 (5.56)	-0.05
KS2 Reading (point scores)	640/1351 (66)	27.5 (9.62)	711/1351 (89)	28.1 (9.53)	-0.06
KS2 Reading and Maths combined	640/1351 (66)	54.4 (12.67)	711/1351 (89)	55.2 (12.81)	-0.06
FSM pupils					
KS2 Maths (point scores)	290/678 (15)	26.0 (5.13)	357/678 (16)	26.1 (5.07)	-0.02
KS2 Reading (point scores)	271/678 (34)	27.3 (9.73)	333/678 (40)	27.2 (9.34)	+0.01
KS2 Reading and Maths combined	271/678 (34)	54.0 (12.67)	333/678 (40)	53.9 (12.03)	+0.01

The primary analysis is conducted for those known to be eligible for EverFSM6. A total of 1,506 pupils are known to be eligible for EverFSM6. Of these, 706 (47%) are in the treatment group and 800 (53%) in the control group. There were no pupils missing KS4 results.

Outcomes and analysis

Primary outcome analysis

Table 6 shows that control pupils were ahead of treatment pupils at pre-test (ES = -0.05 and -0.06 for maths and reading respectively). Since the two groups were not closely balanced at pre-test, the gain score results are used for the headline findings.

Comparison of gain scores (EverFSM6 pupils only)

Analysis for the headline findings is performed for the 1,351 EverFSM6 pupils who have both pre-test scores for reading and maths and post-test scores. For comparability, all scores were converted to Z-scores before analysis.

Table 7 shows that the control pupils were already ahead of their intervention counterparts at pre-test. Although both groups made less progress between KS2 and GCSE compared to their non-FSM peers, the gap between treatment groups closed because the decline was greater for the control than for the treatment group. This suggests that the intervention may have a small influence in improving the performance of the EverFSM6 pupils. But the difference is small.

Table 7: Comparison of pre, post and standardised gain scores using KS2 maths and KS2 reading combined as pre-test and Attainment 8 as post-test (EverFSM6 pupils only)

	Pre-score mean	SD	ES	Post-score mean	SD	ES	Gain score	SD	ES
Treatment (n = 640)	-0.32	0.98		-0.42	0.93		-0.10	0.89	
Control (n = 711)	-0.26	0.99		-0.39	0.94		-0.14	0.84	
Overall (n = 1,351)	-0.29	0.99	-0.06	-0.41	0.94	-0.03	-0.12	0.87	+0.05

The number of counterfactual cases that would be needed to eliminate the headline finding of +0.05 is 32 (0.05 multiplied by 640). This means it would take approximately 32 missing cases with counterfactual scores (see methods) in the opposite direction for the findings to change. Since there are no cases with pre-tests missing post-test scores, this number means that the finding is reasonably secure.

Secondary outcome analysis

Secondary analysis is performed for all pupils (EverFSM6 and non EverFSM6 together). Table 6 shows the number of pupils with pre-test scores and the number of missing cases in both intervention and control group. A total of 5,619 pupils had post-test scores, but some were missing pre-test KS2 maths and some missing KS2 reading scores. Impact analysis was conducted for 5,086 pupils who have both KS2 maths and KS2 reading scores.

Table 8 shows that there is a very small difference between the two groups at pre-test, with the treatment group slightly ahead. Impact analysis shows that the intervention appears to have no impact on the overall pupils' Attainment 8 scores at post-test (+0.01), or for gain scores (ES = -0.01).

Table 8: Comparison of pre, post and standardised gain scores using KS2 maths and KS2 reading combined as pre-test and Attainment 8 as post-test (All pupils)

	Pre-score mean	SD	ES	Post-score mean	SD	ES	Gain score	SD	ES
Treatment (n = 2,547)	0.005	1.005		0.047	0.97		0.042	0.80	
Control (n = 2,539)	-0.005	0.995		0.042	0.97		0.047	0.81	
Overall (n = 5,086)	0.000	1.000	+0.01	0.045	0.97	+0.01	0.045	0.81	-0.01

Comparison of pre, post and standardised gain scores for FSM eligible pupils

We conducted an additional analysis on current FSM-eligible pupils because the self-affirmation theory, as well as previous work by the developers, suggests that the intervention may be more relevant to such pupils than the broader group of those eligible in the previous six years (EverFSM6).

Impact analysis (Table 9) shows that the intervention benefitted FSM pupils to about the same extent as EverFSM6 pupils (+0.05 for post-test analysis, given little imbalance).

Table 9: Comparison of pre, post and standardised gain scores using KS2 maths and KS2 reading combined as pre-test and Attainment 8 as post-test (FSM pupils)

	Pre-score mean	SD	ES	Post-score mean	SD	ES	Gain score	SD	ES
Treatment (n = 271)	-0.35	0.98		-0.46	0.92		-0.11	0.83	
Control (n = 333)	-0.36	0.93		-0.51	0.92		-0.14	0.84	
Overall (n = 604)	-0.36	0.95	+0.01	-0.49	0.92	+0.05	-0.13	0.84	+0.03

Additional analysis

Fidelity analysis

Fidelity to the intervention was assessed in two ways. The first compares the outcomes of pupils with the number of exercises completed (dosage). The number of exercises is treated as a continuous variable. This will be zero for all cases in the control group.

A total of 5,086 pupils with KS2 and Attainment 8 scores completed the writing exercises. Of these, half (n = 2,547) were in the intervention group who wrote about values important to them. The other half (n = 2,539) in the control group were given an alternative writing task.

Table 10 shows over half of the intervention pupils completed all the three writing exercises.

Table 10: Number of exercises completed by intervention group

Number of exercises completed	Intervention (%)
0	111 (4.4%)
1	299 (11.8%)
2	775 (30.4%)
3	1362 (53.5%)
Total	2547

There is a positive but small correlation between the number of exercises completed and the outcomes (Table 11). This is larger for post-test scores than for gain scores, and larger for EverFSM6 pupils than for all pupils overall (Table 12). These figures suggest that compliance is slightly stronger for students with higher absolute levels of attainment, irrespective of the impact of the intervention.

Table 11: Correlation between gain scores and number of exercises completed (all pupils)

	Gain scores using KS2 maths & reading combined	GCSE Attainment 8 score
Number of exercises completed	0.02	0.14

N = 5,080

Table 12: Correlation between gain scores and number of exercises completed (EverFSM6 pupils only)

	Gain scores using KS2 maths & reading combined	GCSE Attainment 8 score
Number of exercises completed	0.09	0.16

N = 1,351

The second method of looking at compliance is the Complier Average Causal Effect Analysis (CACE). Instead of intention to treat, this estimates the likely treatment effects for compliers where not all pupils in the treatment group complied with the intervention. Compliance is defined here as completion of the first writing task. Since the control group was already ahead at pre-test, CACE analysis was conducted using the standardised gain scores (Table 13).

The CACE analysis was performed by estimating the number of pupils in the control group who would have complied if offered the intervention and what their outcomes might be (Table 13). Such an estimate is possible given that we know the overall results for both groups and the results for those who complied.[4] The scores in red are estimated for the control group using the known figures for the treatment (scores in black). The 'effect' size difference in gain scores between treatment groups based on compliers is 0.06 (using the overall standard deviation in Table 7). This is similar to but slightly larger than the simple gain score ES of 0.05 (and is the same as the simple ES divided by the level of compliance in the treatment group, 0.82).

Table 13: CACE compliance based on completion of first writing task and standardized gain scores (EverFSM6)

	Completed first writing task		Did not complete first writing task		Overall		Effect size
	N	Mean	N	Mean	N	Mean	
Intervention	526	+0.01	114	-0.60	640	-0.10	
Control	583	-0.04	128	-0.60	711	-0.14	+0.06

Note: the N in red are based on there being the same proportion of compliers in the control group as in the treatment group (526/640), and the mean scores in red are based on the non-compliers in the control group having the same mean as those in the treatment group.

Regression analysis

In addition, two multivariate regression models were created using Attainment 8 KS4 scores as the dependent variable and total prior test scores (total KS2 maths and reading) and membership of treatment group as predictors. One model uses all pupils and the second uses only EverFSM6 pupils. As usual, the best predictor is pupil prior attainment at KS2 (Table 14). Once this is accounted for, there is little or no impact from the treatment (ES = -0.01).

[4] https://www.sheffield.ac.uk/polopoly_fs/1.418711!/file/JNicholls.pdf

Table 14: Regression results with KS4 Attainment 8 as the dependent variable

	All Year 11 pupils	EverFSM6 pupils
R	0.67	0.60
Coefficient for KS2 maths and reading points as pre-test	0.67	0.60
Coefficient for treatment group	0.00	-0.01

A third way of looking at compliance is regression analyses replacing treatment group with the number of exercises completed (with zero for control pupils). The results are similar to Table 14, but R is slightly larger. Prior attainment is still the best predictor of outcomes but number of exercises completed has a coefficient of 0.15. This may mean that the intervention led to better outcomes if implemented well, or that the kinds of pupils who completed more exercises also made more progress.

Table 15: Regression results with number of exercises completed and KS4 Attainment 8 as the dependent variable

	All Year 11 pupils	EverFSM6 pupils
R	0.67	0.61
Coefficient for KS2 maths and reading points as pre-test	0.66	0.59
Coefficient for completed exercises	0.10	0.15

Cost

Direct and marginal costs

The cost of the intervention is estimated at £1.89 per pupil.

Most of this cost is in relation to the printing of teacher manuals and pupil exercise booklets. On top of that there is the additional one-off cost of training staff to deliver the intervention. The total cost for each school will depend on the number of pupils and teachers involved. Here the cost estimate is based on the intervention being delivered in secondary schools with two year groups, and on the assumption that there is an average of five classes in each group and 30 pupils in each class.

	Year 1	Year 2	Year 3	Over 3 years
Cost of exercise booklet	£0.36	£0.36	£0.36	£1.08
***Cost of initial training (based on £300 for 300 pupils per school)**	£1.00	-	-	£1.00
Printing of teacher manual and teacher scripts	£1.20	£1.20	£1.20	£3.60
Total	£2.56	£1.56	£1.56	£5.68

**The cost of the training is the same regardless of the number of teachers. It is the cost of the expenses for the developers. As the number of teachers in each school varies it is not possible to estimate per teacher cost. It makes more sense to estimate a cost per school.*

Cost per pupil is £1.89 per year spread over three years. This is to allow for comparability across all EEF projects.[5]

A break down on the costs per school is given in the tables below.

Items	Cost per school
Printing of teacher manual script and accompanying teacher scripts for pupils—one for each of the 3 exercises @ 4p each (4x3x300)	£360
Printing of 2/3 page pupil exercise booklets estimated at 4p each (or 12p per booklet). Three exercises a year will be 36p per pupil. Assuming two year groups of 10 classes and 30 in each class. Total is estimated for 300 pupils	£108
Initial training of teachers (travelling expenses and time of trainer)	£300

Schools staff time

The intervention is delivered three times a year during regular English lessons. Each session lasts approximately 15–20 minutes. There is minimal preparation for the teachers apart from distributing and collecting the exercise booklets. This is all done within the regular English lessons. Therefore, no additional time is required in delivering the intervention. However, English teaching staff do need to attend an initial 20-minute briefing on implementing the intervention.

5 **https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_guidance_to_evaluators_on_cost_evaluation.pdf**

Process evaluation

The purpose of the process evaluation was to examine the delivery of the intervention and to assess the fidelity of its implementation. The process evaluation was designed to explore all aspects of the study from the initial recruitment and preparation of schools, through to the training and delivery of the intervention, and the final analyses of outcomes. Observations of teacher briefing sessions, the writing exercises and the survey were carried out across a range of schools during the period of the trial. In addition, we carried out a number of informal interviews and discussions with staff in schools. However, due to restrictions in what teachers' knew about the overall project, these were limited in their scope and focused predominantly on teachers' views of the writing tasks and the children's reactions to them. Due to the nature of the research, pupils were not interviewed as it was key that they were not aware that the writing activity and survey were associated or part of a research project. A small number of Year 11 pupils were contacted with a short questionnaire following the final writing task to ask for their views on the activity. Findings are structured using headings provided by the EEF: implementation, fidelity and outcomes.

The process evaluation also observes the administration of the self-efficacy survey (the results of which will be reported in Report 2 together with the attainment outcomes for the Year 10 cohort). This was deemed necessary as the effectiveness of the intervention is theoretically sensitive to the sequence of the timing of the survey. The intervention specifies that the survey must be administered after the writing task, the writing task must be delivered prior to stressful events (such as before an exam), and that the pupils must not be aware that the two activities are linked. Therefore, it is important to include this aspect in the process evaluation to ensure that schools adhered to the sequence of events and to see if there is any evidence that pupils were able to link the survey to the writing activity. It is important to keep the two activities separate so that pupils do not see the writing activity as a research project, which could potentially compromise the efficacy of the intervention.

Implementation

Training and preparation

The evaluation team attended three of the teacher briefing sessions for the main trial. These were approximately hour-long meetings, led by the project team from Sussex University and attended by English departments. The briefing sessions included a short presentation on the background to the project, including some reference to the evidence-based nature of the intervention and the success of similar trials in America. This information was provided in a very general way, though, so that teachers were not made aware of the full background to the intervention and the current project aims. The majority of teachers at the briefings seemed satisfied with the introduction that they were given to the trial. In one school, however, a teacher asked, 'What is the premise of this research?', stating that she did not feel that the purpose had been made clear enough. The project team repeated information from the introduction in response.

The main focus of the briefing session was explaining to teachers how they should go about delivering the writing exercises in English classes. The project team provided examples of the tasks and spent some time outlining the instructions for completion and the 'best practice' approach to ensuring that children completed them properly. In two schools there were a number of questions raised about whether lower ability pupils or children with English as an Additional Language or Special Educational Needs would be able to access and cope with the independent writing task. The project team reassured teachers that the task did not have to be completed in exam conditions and that one-to-one explanation was permitted, but that the pupils should be allowed to complete the task as independently as possible.

The project team reiterated to teachers that pupils should not be made aware that they are involved in a research project. As such, the writing tasks should not be associated with the surveys (which pupils would complete in form time rather than English lessons) and teachers should not give details about the trial. One teacher suggested that the named envelopes containing the writing tasks may lead pupils to think that this is not a normal school-based activity. The project team emphasised in the sessions that the task should be made to seem like an everyday classroom task and not something unusual or special. They used examples from the pilot to explain that the writing task seems to work best as a starter activity at the beginning of lessons. A head of department in one school supported this view and emphasised to his staff the small amount of time (20 minutes) that the exercise would take.

Further questions were raised in two of the briefing sessions about feeding back to the pupils on their completed writing tasks. One teacher felt that the more able students would be frustrated that nobody would read their work or provide comments on it. She asked if teachers could photocopy the students' responses and provide feedback. The project team said that this was not possible. Other teachers at this school felt that the pupils would be fine receiving no feedback and would 'forget about the task' very quickly.

Teachers at the three sessions we observed were generally positive about participating in the intervention. Some were pleased that children would have the opportunity to complete some 'free writing' and felt that this would provide a pleasant alternative to the very structured, exam-focused work that Key Stage 4 pupils usually undertake. The project team were confident and clear when responding to teachers' questions and concerns. They offered additional telephone briefings for any teachers that were absent, and also explained that there would be a full debrief at the end of the project for all participating schools and staff.

Covert nature of the study

The self-affirmation study is unusual in that staff at participating schools were deliberately not made aware of the full scope of the project. Some aspects relating to the trial objectives, the supporting evidence for the intervention, and the writing exercises/surveys were not fully explained to teaching staff. Teaching staff were also requested not to let students know that the writing tasks were part of a research study. In order to support teachers with this, the Sussex University team had put together a clear set of instructions for teachers to follow; these also gave details of how they should respond to certain questions from pupils.

From our observations, the covert nature of some aspects of the study did not appear to cause any substantial issues for implementation. In a few instances during observations of the writing tasks and surveys, teacher and pupil queries did arise; these are discussed in the subsections below.

Conducting the survey

Observation of the survey administration was necessary to ensure that there was no bias and also to identify challenges that pupils may have with regards to the questionnaire items. More importantly, it was to look for evidence of whether pupils had made the connection between the survey and the writing task, as knowledge of the writing exercise as a research activity could potentially negate the efficacy of the intervention. The results of the survey are not reported here because Report 1 concerns only the Year 11 cohort who did not take the post-survey. The survey results will be reported for the Year 10 cohort only in 2019 when the GCSE results are available.

We observed a total of nine tutor groups across five schools completing the survey in September 2016 during the trial. Our observations suggested that the survey was found to be accessible for the vast majority of Key Stage 4 pupils. Most students were able to understand the questions in the survey and completed them within the time allocated for the task.

There was some variation between schools—and between teachers in the same school—in relation to how much of an introduction was given to the survey. In one school, for example, surveys were placed on desks ready for the students' arrival to the classroom; they were then simply told to sit down, read the instructions and complete the survey independently. In another school students were provided with a PowerPoint slide of instructions which were talked through and explained to the students prior to them beginning the survey.

For a number of pupils across the different groups we observed, there was an issue with ticking responses rather than using an 'X' as indicated in the instructions. In one class, the teacher went round and corrected pupils where this was happening. Some pupils in other groups, however, ticked throughout the survey. The delivery team explained that in such cases, the responses were inputted manually.

The survey appears to have been taken seriously by pupils and teachers alike. Adequate time was provided for completion of the task and, where students had questions, teachers responded helpfully. In one school, for example, some children were unsure about the social background questions and also were not clear of the meaning of 'self-esteem'. The teacher explained the concepts and the questions to the students but ensured that she did not 'lead' the students to provide particular answers.

In two schools it was noted that pupils (and teachers) were enthusiastic about the potential for students to win cash prizes for signing-up to complete the second questionnaire. Our observations suggest that this prompted many students to provide their email addresses, indicating that offering the incentive was worthwhile.

Where students were absent from tutor groups, teachers explained that arrangements had been made for them to complete the surveys on their return to school. Again, this demonstrates commitment to the project and potentially helped to reduce the numbers of missing/incomplete surveys overall.

We observed no evidence that pupils connected the survey to the writing task, although in one school, one form teacher did make the connection even though teachers were instructed not to link the two activities. In this instance the developers immediately informed schools not to make this apparent to pupils.

Conducting the writing exercises

Our evaluation team observed the writing task being completed in a total of ten classes across five schools in September 2016. In February 2017 we observed one class complete the second exercise, and in May 2017, a further five groups were observed across another two schools.

Our observations showed that as per the recommendation of the developers, teachers embedded the writing exercise in to the start of a lesson. One teacher commented that it had been a good way to settle the students before continuing with the rest of the session. Instructions for the writing task were presented to pupils in different ways. All staff that we observed used the instruction sheet that had been provided by the Sussex team; however, some also reinforced their spoken instructions with the use of a PowerPoint slide or handwritten pointers on the whiteboard. Teachers were positive in the way that they introduced the task, often presenting it as an opportunity to 'do something different' or write in a more personal/expressive than was normally required.

Following the instructions, some pupils across the schools we observed asked questions about the writing tasks. In one school a number of pupils appeared confused as to whether the exercise was part of their English lesson or not. In another school pupils wanted to know who was going to look at the writing and whether it was going to be marked. The issue of spelling, punctuation, and grammar was also a source of discussion. Pupils had been informed via the instructions that they did not need to worry about technical accuracy in their writing but should focus on content instead. This appeared to be quite a contradictory message to that usually received from their class teacher, particularly in the

lead-up to GCSEs where they were preparing to be assessed on these skills. In response to the questions, teachers followed the instructions from the sheet provided by Sussex University and, in most cases, children were satisfied with the responses and began writing.

There was just one school where we observed considerable resistance to the task. A middle-ability Year 11 set was very suspicious about what they were being asked to do. One pupil said 'this is so random and confusing, it's a conspiracy'. Another questioned whether it was 'some kind of social experiment'. The 'secret envelopes' were deemed very 'dodgy' and there was vigorous questioning about who was going to read the work and why. These questions created a ripple effect, encouraging others to grumble and occasionally swear about the writing task. The class teacher, an experienced head of department, remained upbeat and positive throughout all of this questioning and stuck closely to the guidance provided. She emphasised the whole-school nature of the project and did eventually get the students to settle and write.

Teachers emphasised the need for pupils to work independently and not to discuss their work with others. In some classes that we observed, this was achieved, while in some groups there was discussion about the exercises and responses before pupils settled to their own writing. This also occurred in a couple of groups where some pupils finished the task quickly and became distracted. In one school the teacher had another task for early-finishers to complete which allowed the rest of the class to continue writing in silence.

There was some flexibility with the timings allowed to pupils. The majority of teachers seemed very keen that students had enough time to complete the task although were also aware that there was considerable variation in the time that some pupils wanted and needed to write. On average, classes were given around 20 minutes to complete their tasks. There were a couple of instances where pupils were given a little extra time in order to finish what they were doing—and because they seemed to be very engaged in completing the task. All classes that we observed completed the task within the first 30 minutes of the lesson beginning.

Following completion of the writing tasks, teachers collected the envelopes and rapidly moved on to the next part of the lesson. In some cases a link was made between the writing exercise and the topic being studied. One teacher, for example, segued in to the exam preparation on the play 'An Inspector Calls' by explaining that they were now going to think about the values being displayed by some of the lead characters in the text. In other schools, no further reference to the writing task was made once the envelopes had been collected.

Attractiveness of the intervention to stakeholders

Teachers

Although teachers and pupils were not aware of the specifics of the intervention, the aim of this process evaluation was to find out about their views regarding the writing exercise, for example, whether they saw any value in the writing activity itself.

Teachers that we spoke to in the initial briefings and during observation visits to schools were generally very positive and optimistic about the project. Most felt that the 15–20-minute time commitment required to carry out each writing exercise was manageable and could easily be slotted in to the start of English lessons. In one school the head of English commented that some teachers found administering the exercises a little challenging as some of the pupils would not fully engage with the tasks. The reason for this, she felt, was that teachers were not able to explain the purpose, rationale and intended outcomes for what they were doing; the independent nature of the tasks also meant that some pupils struggled to complete them. She stated that the more detailed instructions provided with the third writing task made it easier to explain to pupils why they needed to participate and the potential benefits.

A number of heads of department and English teachers commented that the opportunity for young people to write freely and be able to express their personal views was very important. One teacher commented that the children are so conditioned to focus on exams and meeting exam criteria that to do something different was refreshing and interesting. In another school, following the first writing task, one teacher said that being involved in the Writing about Values (WaV) project has made the faculty consider whether to teach more free and creative writing and to embed this within the Key Stage 4 curriculum. She felt that there could be opportunities to include WaV-style tasks within schemes of work, benefitting staff and students by making it a regular and expected part of English lessons.

For a small number of teachers, fitting the exercises in around exam timetables and other school-based activities was a challenge. Some also mentioned that it was often difficult to ensure tasks were completed if students were absent from the original English lesson.

Pupils

The evaluation team's interaction with pupils was limited as it was important that during the main phase of the project they were not aware of participating in a research study. Most students that we observed completing the surveys or writing exercises did so willingly. We only witnessed a very small number of pupils who refused to participate; this is supported by the small number of pupils noted as refusing on the teacher feedback sheets.

In one school the teacher asked her group of pupils if they enjoyed the writing task. One of the boys commented that he liked the writing and found it interesting. Another pupil was worried that she had 'done it wrong' and asked if she could redo the task. The teacher was reassuring and told the whole group that it was more about the process than the final product.

In addition to our observations, our team sent out a short questionnaire to some Year 11 pupils after completion of the final writing task to ask about their views on them. We received a very small number of responses (n = 6) with mixed perspectives from students. One pupil commented that it was 'helpful to be encouraged to see things in a different way but at the same time a lot of people felt as though the time spent on the exercise could have been better used by working towards our GCSEs.' Another student felt that the writing task made them realise that there are 'lots of things that I find valuable' while another said that 'doing something free and away from the prescribed GCSE was a relief.' Two students mentioned that there was considerable overlap between the values exercise and issues raised during their Religious Studies GCSE course. This is quite interesting and perhaps indicates the potential for the writing tasks to be delivered within this subject as an alternative to English lessons.

Fidelity

During the briefing sessions it was made clear to English teachers that the writing exercises needed to be delivered faithfully and consistently. This meant that:

- a total of three exercises had to be delivered over the course of the year;
- pre-determined instructions had to be shared and reinforced with pupils taking the tasks;
- adequate time had to be given for students to write; and
- records of numbers of pupils completing the exercises had to be returned to the developers via the teacher cover sheet.

In addition, the intervention also specifies that the survey must take place following the writing exercises. These were to be delivered by Key Stage 4 form tutors in the schools. Rather than in-person briefing sessions, a set of written instructions was provided to support the delivery of this aspect of the project. This lack of briefing may have explained why there was more variation in the delivery of the survey than the writing tasks. Some teachers that we observed provided little or no instruction to pupils on how to complete the survey, leading to some pupils completing it incorrectly.

A visit to one school also revealed that some form tutors were aware that the survey was part of the University of Sussex self-affirmation research, and was linked to the writing tasks that pupils were completing. Although the initial instructions also emphasised the importance of not linking the survey and the writing tasks, once this was relayed to the project team, they got in touch with schools to inform tutors not to associate the surveys with the written exercises in future.

Scheduling and delivery

The Sussex University team liaised closely and frequently with schools to support and ensure that all elements of the intervention were carried out within the timeframes required. One of the biggest challenges was in scheduling and delivering the written exercises around Key Stage 4 mock exams and actual GCSE examinations. One school, for example, did not complete the third Year 11 writing exercise as the English teachers felt that the time was needed for revision before the GCSE. Despite several attempts by the Sussex team to encourage them to complete the writing exercise, the school was just unable to do it. Another school did not complete the second writing exercise. Apparently the exercises got lost in the school's internal post and turned up eight weeks later, by which time the third exercise was due. So it was not possible to fit in the second Year 11 exercise before the final exercise. The table below summarises the number of pupils who were actually engaged with each of the writing activities.

	Number of pupils who took the writing activity
First writing activity	5,395
Second writing activity	5,108
Third writing activity	5,260

There were also instances of a small number of class groups not completing one of the tasks. We were not provided with the reasons for this. The most common reason for individual pupils not taking the writing activity was pupil absences.

For a number of schools, the survey was delayed due to having Year 10 mock exams scheduled later than the project team had initially anticipated. As the surveys had to be completed after the exams, this did mean some delay in the collection of data but this is unlikely to have had any impact on students' recorded responses.

The second writing activity was to be scheduled between mid-November 2016 to February 2017, depending upon when a school was holding their Year 11 mock exams. Schools were given a two-week window before the exams started for completion of the writing task. The project team from Sussex also aimed to provide a short briefing during this period in order to give feedback on the first exercises and instructions for the second task. Some schools were not available for a full team brief in person; the Sussex team were supportive and flexible with this, offering a video briefing in some cases, or providing a briefing for the lead contact at the school and follow-up emails for the rest of the team. Similar issues were encountered with the third writing activity because of the GCSE exam.

In terms of conducting the evaluation, it was also very challenging to reach schools and arrange visits to observe the second writing activity. Many were not able to confirm mock exam dates or the scheduled days/times for the WaV tasks. Two schools also stated that they did not wish for the evaluation team to be present for the second writing exercise. No reasons were given.

As noted above, the writing tasks on the whole appeared to have been delivered as originally intended. Our observations suggested that English teachers closely followed the guidance and instructions provided by the project team. This is also the case with the delivery of the surveys by form tutors

although the lack of initial briefing perhaps explains why there was more variation in how this activity was presented and delivered.

Outcomes

Due to the nature of the research and the fact that staff and pupils were not made fully aware of the intended outcomes of the project, there were very limited opportunities for the evaluation team to discuss participation in the project and its perceived impact. As noted above, some teachers and students felt that completing the writing exercises had been enjoyable and interesting. Others, however, struggled to see the purpose of the tasks and therefore did not feel that they had value, particularly when viewed within the context of more pressing matters such as GCSE preparation. Crucially, during our visits no staff or students commented on the potential for the tasks to improve attainment in English or to develop improved feelings of self-efficacy.

Formative findings

The intervention was simple, quick, and easy to deliver. The majority of teachers that we observed delivered the writing tasks as required. The covert nature of the intervention meant that teachers had to be very careful about how they responded to pupils' queries. This was handled well mainly because of the clear instructions given by the project team. Therefore, for effective implementation of the intervention, it is important that teachers are thoroughly briefed. In this trial, the project team gave very clear verbal and written instructions to ensure that teachers adhered to the protocol. Additional telephone and email briefings were offered for teachers.

The intervention was, on the whole, well-received by the schools. Teachers generally believed that the intervention gave pupils the opportunity to write freely without fear of mistakes. When asked what they thought of the writing task, almost all the teachers interviewed said they could see the value in the activity and one English head suggested that pupils should be given more opportunity to express themselves. Some teachers suggested that the writing activity allowed pupils to be more creative.

The 20 minutes given for the writing activity was deemed sufficient. All of the pupils we observed were able to complete the writing task within the first 30 minutes of the lesson. The majority of teachers were flexible with the timings allowed for pupils. They understood that some pupils needed more time to complete and were willing to be flexible. However, some pupils completed the exercise quickly and became distractive. One teacher had a planned activity for these early-finishers to keep them occupied. Future implementation could consider an additional planned activity for pupils who finish early.

Tying the intervention in with the English lesson worked well. In some cases a link was made between the writing exercise and the topic being studied. In one case the teacher adeptly segued into the literature text they were studying by asking pupils to think about the values being displayed by some of the lead characters in the text. Given the covert nature of the intervention, this was a clever idea to make the writing task relevant to their regular lessons.

There was some ambivalence about the use of the sealed named envelopes as this was not normal classroom activity and could raise the suspicions of pupils that this was a research activity. The use of the named envelopes was only for the purpose of the trial—to keep the control and intervention writing task separated. If the intervention was to be introduced in schools, there would be no need for the use of envelopes; teachers could simply treat the writing activity as a routine part of their English lessons.

Control group activity

This was a double-blind intervention where neither pupils nor teachers were aware of the nature of the intervention or which writing task was the intervention itself. Both control and intervention pupils were given a writing task at three points in the year. The only difference is in the content of the writing

exercise. There was no report of pupils being given the incorrect envelopes or writing task nor was there any incidence of pupils swapping envelopes with each other.

Conclusion

Key conclusions	
1.	Among disadvantaged pupils, those who received the self-affirmation intervention made slightly more progress between the end of primary school and GCSEs than the comparison pupils, but the size of the impact was very small and further analysis suggests the impact was close to zero. This result has a high security rating.
2.	Pupils who completed more writing exercises made slightly more progress. This may mean that the intervention can lead to better outcomes if implemented more thoroughly, but might also be because the kind of pupils who completed more exercises would make more progress anyway.
3.	Neither pupils nor teachers were aware of the nature of the intervention or why the tasks were undertaken because there is some wider evidence that knowledge about the purpose of the intervention can reduce its effectiveness. Schools considering this approach should bear in mind the difficulty of replicating these conditions.

Interpretation

This trial shows that disadvantaged pupils who received the intervention made slightly more progress between KS2 and KS4 (ES = +0.05) than pupils who did not receive the intervention. Because of differences in the prior attainment of the groups, the results may be sensitive to how these differences are accounted for in the analysis. When prior attainment is controlled for in regression analysis, the small effect disappears. In line with theory, the intervention shows no effect for all pupils in general (that is including non-disadvantaged pupils) (ES = -0.01). This is consistent with previous research suggesting that the intervention can help to mitigate against the negative effect of being stereotyped for being a member of a group that is often performing poorly academically (Oyserman et al., 2006; Cohen et al., 2006; Miyake et al., 2010). The theory is that getting pupils to write positive statements about themselves can give pupils a sense of value and thus alleviate any negative perceptions that they may have of themselves. This improves their self-concept and confidence which in turn can affect their peers' and teachers' expectations of them to do well. The self-efficacy survey results are not available for this cohort of pupils and so the linked mechanism between self-efficacy and attainment outcomes cannot yet be tested. There is no standard interpretation of effect sizes, which must be considered in relation to costs, opportunity costs, and unintended outcomes. This intervention is so cheap that it might still be valuable in practice despite a very small possible impact (Gorard, 2006).

Previous evidence also claims that the effects of the intervention could last for several years. This long-term effect will be tested with the initial Year 10 pupils and the results will be reported in 2019 when their GCSE results are available. The 2019 report will also include an analysis of the self-efficacy survey to establish if the intervention enhances pupils' self-efficacy and thus attainment.

There is currently much policy and practice activity on raising individual's attitudes, aspirations, and self-concept to improve academic outcomes, but much of the evidence so far has been correlational (for example, Marsh and Martin, 2011; Marsh, 1990; Pinxten et al., 2010; Valentine and Dubois, 2005; Skaalvik and Valås, 1999; Marsh and Craven, 2006; Marsh and O'Mara, 2008; Grabowski et al., 2001). Previous research by Gorard, See and Davies (2012) indicates the difficulty in demonstrating the causal mechanism of these attitudinal concepts on academic attainment. The self-efficacy survey measures pupils' self-concept, motivation, self-belief/self-esteem, and locus of control. Therefore, it is hoped that the survey results will provide some evidence on the causal impact of these attitudinal and behavioural measures on pupils' attainment.

Limitations

In this trial, the agreed prior attainment was the KS2 reading and maths scores, and the post-test was the GCSE Attainment 8 scores. Analyses were presented using KS2 reading and maths combined and the gain scores for the post-intervention outcome because of the imbalance at pre-test. The results vary slightly whether using reading (ES = +0.04) or maths (ES = +0.02), with the combined maths and reading showing a bigger effect (ES = +0.05). However, we cannot be sure that the same effects would be achieved if different measures (for example, English rather than reading) were used for both the prior attainment and the post-test.

The characteristics of the pupils in the trial schools are broadly representative of secondary schools in England although they have, on average, a higher proportion of disadvantaged pupils, including EverFSM6 and SEN pupils. This is not surprising as the schools targeted were those with a higher than national average proportion of pupils eligible for free school meals. The trial schools also tend to be lower-performing schools. They have a lower proportion of pupils achieving five A*–C at GCSE compared to the national average. They are also more likely to have a higher proportion of White British pupils and lower proportion of EAL pupils. This is largely because schools were recruited from the South East of England, which is largely a white majority area. So the results may not be applicable to other schools, such as those in the North East or North West, where the demographics may be different.

Future research and publications

This report is based on the attainment outcomes for Year 11 pupils. The EEF will publish a follow-up report in 2019 that will include both the impact evaluation for the initial Year 10 to test the long-term effect of the intervention as well as the effect on pupils' self-efficacy. This will provide some evidence of the link mechanism between the intervention, improvement in self-efficacy, and attainment.

The developers (University of Sussex) will report separately on the different outcomes and subgroups, including:

- separate analyses of English and maths GCSEs for all and EverFSM6;
- assessment of intervention for various subgroups (for example, gender, ethnicity, low attainment groups);
- moderation of the intervention by class- and school-level variables (proportion of FSM, size, OFSTED rating, and so on);
- mediation analyses of non-attainment measures for Year 10s; and
- moderation analyses of non-attainment measures for Year 11s.

It is envisaged that there will be at least one publication from this project in a peer reviewed journal based on the impact evaluation. The University of Sussex will publish the results from the separate analyses.

Future research could address the following questions:

- Is there a long-term impact of the intervention on pupils' attainment outcomes?
- Is the improvement in attainment outcomes a result of the transfer mechanism via improvement in self-efficacy?
- Is there a differential effect for subgroups of pupils (for example, gender, ethnicity and low attainment groups)?
- Do the class- and school-level variables (such as Ofsted rating, proportion of FSM, and size) have a moderation effect?

References

Borman, G. D., Grigg, J. and Hanselman, P. (2015) 'An effort to close the achievement gaps at scale through self-affirmation', Educational Evaluation and Policy Analysis, 0162373715581709, first published on 17 June 2015.

Bowen, N. K., Wegmann, K. M. and Webber, K. C. (2013) 'Enhancing a brief writing intervention to combat stereotype threat among middle-school students', *Journal of Educational Psychology*, 105, pp. 427–35.

Campbell, D. T. and Stantley, J. C. (1963) *Experimental and quasi-experimental designs for research*, Chicago: RandMcNally.

Cohen, J. (1994) 'The Earth is Round (p<.05)', *American Psychologist* 49,12, pp. 997–1003.

Cohen, G., Garcia, J., Purdie-Vaughns, V., Apfel, N. and Brzustoski, P. (2009) 'Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap', *Science*, 324, 5925, pp. 400–440.

Cohen, G. and Sherman, D. (2014) 'The Psychology of Change: Self-Affirmation and Social Psychological Intervention', *Annual Review of Psychology*, 65, pp. 333–371.

Colquoun, D. (2014) 'An Investigation of the False Discovery Rate and the Misinterpretation of *p*-values', Royal Society Open Science, http://rsos.royalsocietypublishing.org/content/1/3/140216.

Colquoun, D. (2016) *The problem with p-values*, Aeon: https://aeon.co/essays/it-s-time-for-science-to-abandon-the-term-statistically-significant

Dee, T. S. (2015) 'Social identity and achievement gaps: Evidence from an affirmation intervention', *Journal of Research on Educational Effectiveness*, 8, 2, pp. 149–168.

Dong, N. and Lipsey, M. (2011) 'Biases in estimating treatment effects due to attrition in randomised controlled trials: A Simulation study', SREE Conference, 2011.

Dunn, G. (2010) 'Estimation of treatment effects from randomised controlled trials in the presence of non-compliance with randomised treatment allocation', PowerPoint presentation slides. Manchester: The University of Manchester.

Foster, M. E. and Fang, G. Y. (2004) 'Alternatives to Handling Attrition: An Illustration Using Data from the Fast Track Evaluation', *Evaluation Review*, 28, pp. 434–464.

Garcia, J. and Cohen, G. (2012) 'A psychological approach to educational intervention', in E. Shafir (ed.) *Behavioural Foundations of Policy*, Princeton, NJ: Princeton Univ. Press (pp. 329–50).

Good, C., Aronson, J., Inzlicht, M. (2003) 'Improving adolescents' standardized performance: An intervention to reduce the effects of stereotype threat', *Applied Developmental Psychology*, 24, pp. 645–662. doi:10.1016/j.appdev.2003.09.002

Gorard, S. (2006) 'Towards a judgement-based statistical analysis', *British Journal of Sociology of Education*, 27, 1, pp. 67–80.

Gorard, S., See, B. H. and Davies, P. (2012) *The impact of attitudes and aspirations on educational attainment and participation*, York: Joseph Rowntree Foundation.

Gorard, S. and Gorard, J. (2016) 'What to do instead of significant testing? Calculating the

number of counterfactual cases needed to disturb a finding', *International Journal of Social Research Methodology*, 19, 4, pp. 481–490.

Gorard, S. and See, B. H. and Morris, R. (2016) *The most effective approaches to teaching in primary schools*, Saarbrucken: Lambert Academic Publishing.

Gorard, S., See, B. H. and Siddiqui, N. (2017) *The trials of evidence-based education: The promises, opportunities and problems of trials in education*, London: Routledge.

Grabowski, L., Call, K. and Mortimer, J. (2001) 'Global and economic self-efficacy in the educational attainment process', *Social Psychology Quarterly*, 64, 2, pp. 164–179.

Little, R. J. A. and Rubin, D. B. (1987) *Statistical analysis with missing data*, New York: Wiley.

Marsh, H. W. and Martin, A. J. (2011) 'Academic self-concept and academic achievement: Relations and causal ordering', *British Journal of Educational Psychology*, 81, pp. 59–77. doi:10.1348/000709910X503501.

Marsh, H. (1990) 'Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis', *Journal of Educational Psychology*, 82, 4, pp. 646–656.

Marsh, H. and O'Mara, A. (2008) 'Reciprocal effects between academic self-concept, self esteem, achievement, and attainment over seven adolescent years: unidimensional and multidimensional perspectives of self-concept', *Personality and Social Psychology Bulletin*, 34, 4, pp. 542–552.

Marsh, H. and Craven, R. (2006) 'Reciprocal effects of self-concept and performance from a multidimensional perspective: beyond seductive pleasure and unidimensional perspectives, *Perspectives on Psychological Science*, 1, 2, pp. 133–163.

Mikaye, A., Kost-Smoth, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L. and Ilto, A. (2010) 'Reducing the gender achievement gap in college science: A classroom study of values affirmation', *Science*, 330, pp. 1234–237. doi:10.1126/science.1195996

Nicholls, J. (undated) 'Complier Average Causal Effect analysis', https://www.sheffield.ac.uk/polopoly_fs/1.418711!/file/JNicholls.pdf

Oyserman, D., Bybee, D. and Terry, K. (2006) 'Possible selves and academic outcomes: How and when possible selves impel action', *Journal of Personality and Social Psychology*, 91, pp. 188–204. doi:10.1037/0022-3514.91.1.188

Perezgonzalez, J. D. (2015) 'The Meaning of Significance in Data Testing', *Frontiers in Psychology*, 6, 1293, pp. 1–4, doi: 10.3389/fpsyg.2015.01293.

Pharoah, P. D. P., Jones, M. R. and Kar, S. (2017) 'P-values and confidence intervals: not fit for purpose?', bioRxiv preprint first posted online 24 August 2017: doi: http://dx.doi.org/10.1101/180117.

Pintrich, P. R., Smith, D. A. F., Garcia, T and Mckeachie, W. J. (1993) 'Reliability and predictive validity of the Motivated Strateties for Learning Questionnaire (MSLQ)', *Educational and Psychological Measurement*, 53, 3, pp. 801–813.

Pinxten, M., De Fraine, B., Van Damme, J. and D'Haenens E. (2010) 'Causal ordering of academic self-concept and achievement: effects of type of achievement measure', *British Journal of Educational Psychology*, 80, 4, pp. 689–709.

Puma, M. J., Olsen, R. B., Bell, S. H. and Price, C. (2009) *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049), Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Skaalvik, E. and Valås, H. (1999) 'Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study', *The Journal of Experimental Education*, 67, 2, pp. 135–149.

Shadish, W. R., Cook, T. D. and Campbell, D. T. (2001) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Boston: Houghton Mifflin.

Sherman, D. K., Hartson, K., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., Tomassetti, S., Nussbaum, A. D and Cohen, G. L. (2013) 'Deflecting the Trajectory and Changing the Narrative: How Self-Affirmation affects academic performance and motivation under identity threat', *Journal of Personal and Social Psychology*, 104, pp. 591–618.

Sherman, D. K. and Cohen, G. L. (2006) 'The psychology of self-defense: Self-affirmation theory', *Advances in Experimental Social Psychology*, 38, pp. 183–242.

Simmons, C. M. (2011) 'Reducing stereotype thread in academically at-risk African American students: A self-affirmation intervention', PhD thesis, Berkeley: University of California.

Steele, C. M. (1988) 'The psychology of self-affirmation: 'Sustaining the integrity of the self', in L. Berkowitz (ed.), *Advances in experimental social psychology* , New York, NY: Academic Press (Vol. 21, pp. 261–302).

Steele, C. M. (1997) 'A threat in the air: How stereotypes shape intellectual identity and performance', *American Psychologist*, 52, pp. 613–629. doi:10.1037/0003-066X.52.6.613

Steele, C. M. and Aronson, J. (1995) 'Stereotype threat and the intellectual performance of African Americans', *Journal of Personality and Social Psychology*, 69, pp. 79–811.

Trafimow, D. and Rice, S. (2009) 'A Test of the Null Hypothesis Significance Testing Procedure Correlation Argument', *The Journal of General Psychology*, 136, pp. 261–269. doi: 10.3200/GENP.136.3.261-270.

Valentine, J. and Dubois, D. (2005) 'Effects of self-beliefs on academic achievement and vice-versa: separating the chicken from the egg', in Marsh, H., Craven, R. and McInerney, D. (eds), *International advances in self research*, Greenwich, CT: Information Age (2, pp. 53–78).

Appendix A: Padlock rating

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	Power	Attrition*					
5 🔒	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%	5				
4 🔒	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%			Adjustment for Balance **[-1]**		4
3 🔒	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%					
2 🔒	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%			Adjustment for threats to internal validity **[ ]**		
1 🔒	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	41-50%					
0 🔒	No comparator	MDES > 0.6	>50%					

- **Initial padlock score**: lowest of the three ratings for design, power and attrition = 5 padlocks
- **Reason for adjustment for balance (if made):** Imbalance of 0.05 and 0.06 in maths and reading respectively at KS2.
- **Reason for adjustment for threats to validity (if made):** None
- **Final padlock score:** initial score adjusted for balance and internal validity = 4 padlocks

Appendix B: Cost

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 C1: Second writing exercise for control group

Name:

Date:

English teacher:

Writing about your life

People begin their days in many different ways. Sometimes it can be interesting to think about the way we begin our own day.

In the space below, please write about what you did this morning before you started school. What time did you get up? How long did it take to get ready? Did you eat or drink anything? How did you get to school? What did you pass on the way to school?

Try to start with the very first thing you did this morning, then describe what happened afterwards.

Focus on writing down what happened, and don't worry about spelling, grammar, or how well written it is, or how much you can write.

Please turn over

Appendix C2: Second writing exercise for treatment group

Name:

Date:

English teacher:

Writing about your life

There are a lot of things that are important to people—things that make their lives better, more important, or special.

For example, some people find being honest important because other people can trust them. Some other people find their family important because they love and value them. Other people find being good at sport important because it makes them feel good to play well.

In the space below, please write about what **you** find important in **your** life. How important is it to you? Why is it important to you? What does it mean to you to have it in your life?

Focus on your thoughts and feelings, and don't worry about spelling, grammar, or how well written it is, or how much you can write.

Please turn over

Appendix D: Teacher information sheet

Writing about Values exercise – Instructions for teachers

You will receive a box of envelopes with the writing exercises, sorted by class, with your pupils' names on the front.

What to do:

☑ Ensure the class is settled. Introduce the exercise as you would any other in-class exercise using your own words, but please ensure you cover the 10 numbered points below:

1. **For the first part of today's lesson, we're going to be doing something a bit different - a free-expression exercise.**
2. **I'm going to hand you out an envelope with your name on.**
3. **DO NOT open them until I tell you.**

☑ Then, give each envelope to the corresponding pupil, but do not let them open them yet. If a pupil's envelope is missing, please write their name on one of the blank envelopes and use that. Now please cover the following points:

4. **Read the instructions carefully so you know what to do**
5. **There are no right or wrong answers**
6. **The exercise is a chance for you to spend some time writing about your own thoughts and ideas; it's about the process of doing the activity rather than me providing feedback so it's not going to be marked**
7. **You don't need to focus on spelling or grammar**
8. **It takes about 10-15 minutes**
9. **Work individually and silently**
10. **If you have a question, raise your hand and I will come over to your desk**

☑ If you would normally do so, you can now check for questions. Ensure pupils are silent and then ask them to begin. Please make sure the pupils complete the exercise individually. If a pupil has a question, approach them at their desk and talk to them quietly, using the FAQs below where possible.

☑ Give pupils 10-15 minutes of writing time to complete the exercise. If a pupil finishes earlier, please encourage them to go back over their work. After about 10 minutes, please say something like "You have a couple of minutes left to finish up, don't worry if you can't quite finish it". It doesn't matter if some take longer than others.

☑ Have the **pupils put their completed exercise back into the envelopes and collect them.** Please fill out the cover sheet at the back and give everything to your school contact at the end of the day. Please do *not* refer back to the exercise in class once it is completed.

☑ **If any pupils are absent,** please give the exercise to them when they are next in your class (within 2 weeks of original exercise date) and write the date that they completed the exercise on the envelope.

Suggested responses to **frequently asked questions** from pupils:

- **Why are we doing this?** — Pupils in other schools have found that spending some time thinking and writing about their own thoughts really helpful and we are keen to try them out. Everyone in Y10 and Y11 is doing the exercise (If a pupil refuses, please accept this and note it on your cover sheet).
- **Will I get marked on this?/ Who will read this?** — I will check to see if you've engaged with it properly, but it won't be marked. The exercises will be stored away.
- **What are you going to do with what I write?** – This is about the process of writing and giving you the chance to write your own ideas, so it won't be marked. We'll collect them up and store them away.
- **Why do we get envelopes?** — You're writing about your own personal thoughts and ideas, so it's important that they are private.
- **Why do I have different questions from him/her?** — Everyone's got their own task but there's not enough time for everyone to do them all, some people have different ones.
- **Is this for the whole school?** — All Y10 and pupils will be doing this at some point.
- **Does spelling/grammar matter?** — No, just focus on writing down your thoughts.
- **Can I write about a value that's not on the list?** — For now, just choose one on the list.
- **Is this part of the study/research?** — This is an exercise that our school is trying out this year. (If possible, address this question individually at their desk)

Teacher Cover Sheet

Please fill this out for every class the exercises are done in alongside the class list

Teacher: _____

Class: _____ Date: _____ Time: _____

Please indicate to what extent you agree with the following statements. Please circle.

Overall, the writing exercise went smoothly today:

Strongly disagree 1 2 3 4 5 6 7 Strongly agree

Overall, the pupils found it easy to complete the exercise:

Strongly disagree 1 2 3 4 5 6 7 Strongly agree

None A few Most All

About how many pupils

- Talked during the exercise……	1	2	3	4
- Did not complete the exercise seriously……..……..	1	2	3	4
- Thought it was a university/research exercise……...	1	2	3	4
- Were aware there were different versions of the exercise	1	2	3	4
- Needed help completing the exercise……..	1	2	3	4

Do you have any general comments about how the exercise went? (e.g. pupil reactions, pupil questions and how you answered, issues, whether pupils understood the exercise, whether they enjoyed it, etc.):

Thank you for your time!

Appendix E: Pilot

Pilot phase

Prior to the main trial, a pilot was conducted in five schools to develop and test the materials, manual or protocol, and training approaches. These schools did not form part of the main trial, and were offered £500 to assist with the pilot trial.

Purpose

The aim of the pilot was to test the intervention materials, such as the pupils' writing exercise booklets and the scripts used by the teachers. It assessed whether teachers were able to use the scripts with fidelity. The booklets were tested to make sure that they were age-appropriate, fitted the context, and that instructions were clear. The pilot also provided opportunities to rehearse the randomisation process, the intervention delivery procedures and the plan for teacher training and for trialling the non-attainment survey instrument. We also tested the feasibility/complexity of the process for a joint application for the NPD data.

Methods

Before the pilot two local secondary schools in Sussex were approached at the beginning of 2016 to test the intervention materials. The pilot involved only Year 11 pupils because the project team wanted to test whether it was feasible to introduce the intervention prior to the GCSE exams. The pilot intervention began with the training of teachers in April 2016. Focus groups with pupils and English teachers provided feedback that informed the development of the writing task. Teachers were provided with a one-hour training session to deliver the writing task.

The delivery of the treatment in the pilot phase was carried out in early May 2016 before the onset of the GCSE exams. Light touch process evaluation of the delivery from training of staff to implementation in the classroom was conducted in a random sample of three of the five pilot schools to test the fidelity of implementation, and assist where possible by providing suggestions and feedback. Formative feedback on the training, delivery of intervention, and teaching materials were relayed back to the project team. The pilot also helped ascertain whether the level of support and training was sufficient, identify potential hiccups and to suggest improvements needed to ensure that the main trial ran smoothly. The process evaluation was primarily in the form of participant observations, and informal chats with staff and pupils in the pilot schools to identify potential barriers to implementation, issues with data collection, possible resistance and also any potential risks of contamination. Lesson observations were very informal. More specifically we asked staff about issues relating to the delivery/implementation, resources/materials used, and if there were suggestions for improvement.

Formative feedback from the pilot

The teacher briefing sessions provided feedback on the kind of questions that teachers and pupils were likely to ask and this informed the final teacher instruction protocol. The survey

questionnaire was revised and the number of items reduced in light of the feedback from pupils and teachers.

The convergent and predictive validity of the scales, as well as their internal consistency were tested in this pilot phase and the self-efficacy measure was selected as the most appropriate outcome to be included in the final analyses.

The pilot schools that we visited were, on the whole, very positive about participating in the self-affirmation project. At the teacher briefings that we attended, senior leaders were also present, providing further support to the English Department, and highlighting the value of involvement. In schools that were more familiar with using or engaging with research, English teachers tended to ask more questions about the delivery of the writing exercises and surveys. Some wanted details about the underlying mechanisms of the intervention; the project team noted that they could not always share the full details of the study and its background with teachers, but did have prepared responses. In one school it was also noted that the teaching staff were understanding of the fact that not all details could be shared with them.

The pilot phase was useful for helping the project team to refine or develop their responses to teachers' questions. This was the case, for example, in relation to safeguarding and confidentiality issues linked to the writing exercises, and in terms of the amount of support that teachers should provide to lower ability/SEN students and the rationale for this.

Teachers delivered the writing tasks as requested by the Sussex University project team. While a small number of students asked why they were doing the exercise, or what the purpose was, overall, they were compliant and were satisfied with the responses that teachers gave.

The survey was delivered in all of the pilot schools. However, one school administered it to pupils prior to the writing task. This was a concern as the intervention protocol states clearly that the survey should come after the writing activity so that the content does not influence what students write about. The project team noted that ensuring that main trial schools adhered to this sequence was important. Staff and pupils at some of the pilot schools also informed the project team and the evaluation team that they found the survey too long; this was something that was reviewed and addressed prior to the main trial.

Appendix F: Memorandum of Understanding

Memorandum of Understanding: Writing about Values Project

This Memorandum of Understanding (MoU) is an agreement to participate in the Writing about Values Project and to support the evaluation of the trial.

This project is conducted by the University of Sussex and independently evaluated by Durham University. It is funded by the Education Endowment Foundation.

The Trial

This is a randomised controlled trial involving Y10 and Y11 pupils. The intervention involves writing about values. There is currently much research in the US indicating that such exercises can improve educational outcomes by improving an individual's aspirations, attitudes and behaviour. This project will be the first randomised trial in the UK to test this hypothesis.

The Project Team from **the University of Sussex** will conduct the Writing about Values project in the academic year **2016/2017** with all **Y10 and Y11 pupils** at your school. By participating in the project the school is agreeing to commit to the following:

- Distribute information letters: ***By June 30th 2016***

The Project Team will deliver to your school information letters and opt-out consent slips for the parents of participating pupils. **The school will distribute the information letters and opt-out slips to parents or carers, and keep any returned slips until the Project Team collects them.** The Project Team will take responsibility for ensuring that opted-out pupils do not participate in the trial.

- Provide pupil background data & class lists: ***By June 30th 2016***

The school will provide the Project Team with the following **pupil background data** from the School Information Management System (SIMS) for all Y10 and Y11 pupils:

- Unique Pupil Number (UPN)
- Pupil forename(s) and surname
- Free School Meal status (FSMever)
- Date of birth
- Ethnicity
- Gender
- Special Educational Needs (SEN)
- English as an Additional Language (EAL)
- Language spoken at home
- Name of English class & ability level/set
- Form/tutor group
- Y10 end-of-year assessments (provide this data in summer 2017 for Y10 only).

The most secure format and means of data transfer will be discussed with the school.

- Release teachers to attend briefing session: ***5th – 16th September 2016***

The Project Team will provide a **1-2 hour teacher briefing** at the school that must be **attended by all Y10 and Y11 English teachers.** Teachers will be briefed on the intervention and the delivery of the exercises.

❖ Conduct paper surveys: ***September 2016 and June/July 2017***

The school will allow time for Y10 and Y11 pupils to complete a **15-minute paper survey** outside of English classes, preferably in form or tutor time, on **two occasions**:

1. Y10 and Y11 pupils must complete the first survey within the first two weeks of the 2016/17 academic year.
2. Y10 pupils must complete the second survey in June or July 2017, after the interventions have taken place.

The Project Team will deliver the surveys to the school. **The school will administer the surveys** to the pupils and collect and store the surveys until they are collected by the Project Team.

❖ Deliver the interventions: ***September 2016; Before mock GCSEs; Before final GCSEs***

The school will **deliver the intervention in English classes** on **3 occasions** in the 2016/2017 academic year:

1. The first writing exercise must be completed within the **first three weeks of the academic year**, after the teacher briefing and first survey have taken place.
2. The second writing exercise must be completed **within the two weeks before your school's mock GCSEs** for Y11 and around the same time for Y10
3. The third writing exercise must be **completed within the last two teaching weeks before the GCSE exams** for Y11 and around the same time for Y10

Each of the three writing sessions takes approximately 20 minutes and must be delivered as a **normal classroom writing exercise** by **English teachers**.

The Project Team will place each intervention into a named envelope for each individual pupil, organise the interventions into whole-class groups, and deliver them to the school. The school will distribute the interventions to the pupils in English classes, and collect and store the completed interventions until they are collected by the Project Team.

Structure of the Evaluation

An Evaluation Team from **Durham University** has been assigned to independently evaluate the project in order to assess the rigour with which it is implemented and its scientific integrity. By participating in the project the school agrees to the following evaluation elements:

Pupil data will be collected from the NPD

- The Evaluation Team will collect KS2 and GCSE results from all pupils in the trial from the National Pupil Database (NPD). This will be done in October 2017 for Y11 pupils and in October 2018 for Y10 pupils.
- In a small sample of the schools that are participating in the trial:
    - A member of the Evaluation Team will sit in the classroom and observe the process of delivery and administration of the intervention and surveys.
    - A member of the Evaluation Team will conduct informal interviews with pupils and teachers after the trial (in June/July 2017) about their experience of and opinions about the writing activities.

The schedule of visits will be agreed with the Project Team and the school.

Use of Data

Pupils' survey responses, responses to the writing exercises, and any other pupil data will be treated with the strictest confidence.

- The completed surveys and interventions will be collected by the Project Team and accessed by the Evaluation Team. All data will be securely stored and anonymously processed.
- **Anonymised pupil data** will be matched with data from the National Pupil Database by the Evaluation Team and shared with the Project Team and the Education Endowment Foundation (EEF) and submitted to FFT Education to be part of the UK data archive.

No individual school or pupil will be identified in any report arising from the research.

Summary of Roles and Responsibilities

The School will:

- Distribute information letters and opt-out slips to parents/carers by June 2016 and pass onto the Project Team any returned opt-out consent slips
- Share pupil background data from the SIMS with the Project Team by June 2016
- Release English teachers to attend a 1-2 hour briefing session taking place between 5th – 16th September 2016.
- Conduct three intervention writing exercises in English classes at the above specified times (three exercises, 20 minutes each)
- Conduct two paper surveys in form time or other classes at the above specified times (two surveys, 15 minutes each)

The University of Sussex Project Team will:

- Conduct a 1-2 hour briefing session for English teachers between 5th – 16th September 2016
- Provide the schools with individually labelled paper surveys and pick these up when completed (on 2 occasions)
- Provide the schools with individually labelled envelopes containing the writing exercises and pick these up when completed (on 3 occasions)
- Provide the first point of contact for questions about the trial and evaluation
- Provide on-going support to the school via a dedicated contact person
- Send out regular updates (twice per term) on the progress of the project through a newsletter
- Organise payment of £1000 honorarium via the EEF once the school has completed the whole trial
- Fully debrief staff and pupils via a letter about the project after its completion

The Durham University Evaluation Team will:

- Observe the delivery and administration of a sample of the writing exercises and surveys
- Conduct informal interviews with pupils and teachers after the trial (June/July 2017)
- Retrieve pupil data (e.g. KS2, GCSE results) from the National Pupil Database using UPNs
- Analyse the data from the project to assess the effectiveness of the trial at raising pupil attainment
- Ensure all staff carrying out assessments have been trained and have received DBS clearance
- Disseminate research findings in an EEF report

Appendix G: Consent form

University of Sussex

RESEARCH PROJECT INFORMATION SHEET

Dear Parent/Guardian,

Your child's school is taking part in a research project in the 2016/2017 school year. The headteacher at the school has agreed that the project can take place, and the activities are being planned within normal class time.

This information sheet tells you a bit about why we are doing this research and what is involved for you and your child.

Please take your time to read the information carefully and keep this sheet. If you decide that you do **NOT** want your child to take part, then please sign the attached opt-out form and return it to reception at your school. If you are happy for your child to take part, you do not have to do anything.

WHAT IS THE PURPOSE OF THE PROJECT?

The purpose of the project is to evaluate some classroom activities. There is evidence from America that these classroom activities can improve students' work. This research is evaluating whether the classroom activities have the same effect in England. We are conducting this study across several schools in South East England.

WHY HAS MY CHILD BEEN INVITED TO PARTICIPATE?

Your child's school was approached to take part in the research project and they agreed. All Year 10 and Year 11 pupils in the school will be involved in the project.

DOES MY CHILD HAVE TO TAKE PART?

Taking part in the project is completely voluntary. If you are happy for your child to take part, then you do not have to do anything. If you decide that you do **NOT** want your child to take part, then please sign and return the attached opt-out consent slip by 10th of July 2016. Either way, your decision will have no effect on your child's mark, assessment or future studies.

WHAT WILL HAPPEN IF MY CHILD TAKES PART?

Every pupil that takes part will complete two short questionnaire (15 minutes each) in form or tutor time, one at the beginning and one at the end of the 2016/2017 school year and take part in three classroom activities in September 2016, December 2016 and May 2017. All the activities will be conducted in normal class time.

1

WHAT ARE THE POSSIBLE BENEFITS OF TAKING PART?
The classroom activities may improve students' work.

WILL MY INFORMATION IN THIS PROJECT BE KEPT CONFIDENTIAL?
Yes. All information given by pupils in this research would be treated confidentially, except in cases where there is specific information that indicates that the person or another specific person is experiencing, or is at risk of experiencing, significant harm. In that situation, we would communicate with the relevant contact person at the school. Your child's responses will be collected by the University of Sussex Project Team and accessed by a a research team from Durham University (the independent evaluator of the research). In the study dataset, your child will be identified only by a number in the study dataset and no names will be included. For the purpose of research, the responses will be linked with information about your child from the National Pupil Database (held by the Department for Education) and shared with the Department for Education, EEF, EEF's data contractor FFT Education and in a fully anonymised form to the UK Data Archive. Your child's data will be treated with the strictest confidence. We will not use your child's name or the name of the school in any report arising from the research.

WHAT WILL HAPPEN TO THE RESULTS OF THE RESEARCH PROJECT?
The results of this project will be reported by the Education Endowment Foundation. The report will be freely available on their website. The results may also be written up for publication in academic journals.

WHO IS ORGANISING AND FUNDING THE RESEARCH?
The research is being conducted by members of staff in the School of Psychology at the University of Sussex, and independently evaluated by the School of Education at the University of Durham. The project is being funded by the Education Endowment Foundation.

WHO HAS APPROVED THIS RESEARCH?
The study has been approved by the University of Sussex's Sciences and Technology Cross-Schools Research Ethics Committee and the Education Endowment Foundation.

CONTACT FOR FURTHER INFORMATION
If you would like any further information about the research, please contact Dr. Matthew Easterbrook at the University of Sussex on m.j.easterbrook@sussex.ac.uk. If you have any concerns about the way the study is conducted, you should contact the Chair of the Cross-Schools Research Ethics Committee at the University of Sussex on crecscitec@sussex.ac.uk (project reference ER/MJE23/15). The University of Sussex has insurance in place to cover its legal liabilities in respect of this study.

Thank you very much for your time!

University of Sussex

RESEARCH PROJECT OPT-OUT CONSENT SLIP

Please only complete if you do NOT want your child to participate in the research project

Details of the research project are included in the attached information letter. If you are happy for your child to take part, then you do not have to do anything. However, if you do NOT want your child to take part in the research project, then please sign the below slip and return it to the reception area at your school by the 10th of July 2016. If you have any questions, please contact Dr. Matthew Easterbrook on m.j.easterbrook@sussex.ac.uk or 01273 876597.

.....

I do NOT want my child (full name)

_____ at (school)

_____ in

(class/form)_____ to take part in the research project

being conducted by the University of Sussex and the University of

Durham, as described in the information letter.

Parent/guardian name: _____

Parent/guardian signature: _____

Date: _____

(completed slips can be returned to the reception area at your school)

3

Appendix H: Statistical Analysis Plan

INTERVENTION	Writing about Values
DEVELOPER	University of Sussex
EVALUATOR	Durham University
TRIAL REGISTRATION NUMBER	ISRCTN79754465
TRIAL STATISTICIAN	Stephen Gorard
TRIAL CHIEF INVESTIGATOR	Beng Huat See
SAP AUTHOR	Beng Huat See
SAP VERSION	
SAP VERSION DATE	6 November 2017
EEF DATE OF APPROVAL	
DEVELOPER DATE OF APPROVAL	

Introduction

"Writing about Values" is a one-year intervention comprising two phases: an initial pilot phase and the main trial. The main trial consists of two randomly controlled trials in the same schools. One trial involves Year 10 pupils and the other involves Year 11 pupils. The inclusion of the Y10 pupils enables evaluation of the long-term impact of the intervention – a year after the end of the intervention.

The trial is a double-blind experiment where both pupils and teachers are not told what the intervention involves. The writing exercises are completed during English lessons as part of the regular English class. The treatment and control pupils are given similar exercises with a slight variation.

The intervention is based on the hypothesis that students from some stigmatised groups are aware that they are the target of a negative stereotype regarding their academic performance (Steele 1997). The "Writing About Values" strategy has been employed to alleviate the effects of stereotype threat on low performing students, especially those from ethnic minority

backgrounds (Oyserman et al. 2006; Cohen et al. 2006; Miyake et al. 2010) by getting them to write positive statements about themselves (Cohen & Sherman 2006). It is believed that this can help ameliorate the detrimental effects of stereotype threat on academic performance. For this trial the focus is on EverFSM pupils rather than ethnic minorities because EEF's focus is on disadvantaged pupils.

Study design

This is an efficacy trial running for two years involving two randomised trials: one with Y10 pupils for two years, and another with Y11 pupils for one year. However, delivery of the intervention stops at the end of the first year (July 2017). Evaluation of impact for Y11 will be undertaken at the end of the first year after the GCSE results, while impact evaluation for the Y10 will be at the end of the second year to test the sustained effect.

This trial is conducted as a double-blind experiment in that both pupils and teachers are not given information about the intervention apart from the fact that it involves a writing task.

Randomisation

Pupils will be individually randomised within school, stratified by year group and free school meal status. Randomisation is carried out after all participating schools have submitted the pupil data. A random number generator on Excel will be used for this process which will be conducted in the presence of colleagues in the School of Education. Year groups will be randomised separately, and FSM and non-FSM separately within each year. There will be, in effect, four randomisations. Schools will not be informed of the results of the randomisation as it is a double-blind experiment. Since the treatment and control pupils are given different writing tasks, the developers will be informed of the result of the randomisation immediately in order for the named exercise booklets to be printed on time.

Calculation of sample size

The sample size calculation is based on the assumption that there would be 25 schools and two year groups (Year 10 and Y11). Assuming an average of 5 forms in each year group, there will be 125 forms (25 X 5) per year group. Working on the assumption of an average of 30 pupils per form, there will be 3,750 pupils (30 X 125) for each year group or 7,500 overall. Randomising individual pupils to treatment conditions, there will be 1,875 pupils in each arm for each year group (or 3,750 per arm overall). Assuming around 25% of EverFSM-eligible pupils overall, this would mean around 470 EverFSM pupils per arm per year group (or 940 per arm of the trial).

Traditional power calculations are based on the approach of significance testing (Gorard et al. 2017), which is misleading. They are therefore not included here.

Instead, we calculate the sample size needed for any effect' size to be considered secure by considering *a priori* the number of 'counterfactual' cases needed to disturb a finding (Gorard

and Gorard 2016). This number needed to disturb (NNTD) is calculated as the 'effect' size multiplied by the number of cases in the smallest group in the comparison (i.e. the number of cases included in either the control or treatment group, whichever is smaller). This approach allows for estimating ES and sample size using the formula as shown.

$$NNTD = ES*n$$

Therefore, $n = NNTD/ES$ and

$$ES = NNTD/n$$

This is a useful measure of the scale of the findings to chance (and their variability as represented by the standard deviation used to compute the 'effect' size), taking into account the scale of the study. It can then be extended to compare this sensitivity directly to other more substantial sources of error such as the number of missing values/cases. The number of cases actually missing a value can be subtracted from the NNTD to give an estimate of how large the 'effect' size would be even in the extreme situation that all missing cases had the "counterfactual" score hypothesised in the NNTD calculation. Here the 'counterfactual' score is one standard deviation away from the mean of the group with the largest number of cases. The standard deviation would be added if the mean of the smaller group (in scale) were smaller than the mean of the larger group, and subtracted if the mean of the smaller group was the largest. (Gorard et al. 2017).

Based on Gorard et al. 2016, NNTD of 50 can be considered a strong and secure finding. Using this as a working assumption, the number of cases needed in each group (assuming equal size) to detect an 'effect' size of 0.2 (which is typical for an education intervention) will be 250 (or 50/0.2). This is assuming no attrition.

Assuming 25% of pupils are EverFSM (n=940), we would expect to detect an 'effect' size of 0.05, or 50/940 (ES=NNTD/n), for the EverFSM pupils. In reality 29 schools and 11,978 pupils were recruited (n =5953 treatment and n = 6025 control). Of these 26.8% (or 3,131) had EverFSM status (or around 1,565 per arm). This will enable us to confidently detect an 'effect' size of +0.03.

The NNTD calculation concerns the security of a difference, and so is relevant to internal validity only. Issues such as clustering, concerned with whether the result may also occur among cases not in the RCT, are therefore irrelevant. In addition, as pupils are individually randomized within schools and analysis would be of all pupils in the two groups and not by schools, clustering effects, if there are any, should be evenly spread between the two groups across all schools.

Follow-up

One school pulled out of the intervention after randomisation, but it was agreed with the EEF and the developers that only schools that completed the baseline survey will be included in the trial and the analysis. This school was therefore not included in the evaluation.

Outcome measures

We propose using attainment 8 KS4 scores as the main attainment outcomes. We will use KS2 maths total marks and reading marks as the pre-intervention attainment measures.

Primary outcomes

- Attainment 8 KS4 scores for FSM pupils (based on EverFSM) after 1 year of treatment (for initial Y11 pupils).
- Attainment 8 KS4 scores for FSM pupils (based on EverFSM) after 2 years (one year after the end of the intervention) for pupils who received the intervention when they were in Y10.

Secondary outcomes

- Attainment 8 KS4 scores for All (EverFSM and non-EverFSM) pupils after 1 year of treatment (for initial Y11 pupils).
- Attainment 8 KS4 scores for All (EverFSM and non-EverFSM) pupils after 2 years (for pupils who received the intervention when they were in Y10).

Non-attainment outcomes

Pupils' perceived self-efficacy measured using the subscales from the Motivated Strategies for Learning Questionnaire (Pintrich et al. 1993).

Analysis

Primary intention-to-treat (ITT) analysis

The analyses for the impact evaluation will be based on the difference between groups in terms of their post-test mean scores for EverFSM pupils only. The differences will be expressed as effect sizes (Hedge's) and converted to progress in months. Given the number of cases per cell and the reported effect sizes, it will be possible for readers to construct a confidence interval.

Imbalance at baseline

To establish baseline equivalence we will use the 'effect' sizes for each measurement at the outset and also present the characteristics of schools in each group. To cater for any initial imbalances between groups we also present the gain scores analysis. For the benefit of readers we present the pre-, post- and gain scores regardless of imbalance.

Missing data

Dong and Lipsey (2011) demonstrated that any missing values can create bias, even if attrition is balanced between comparator groups. And where such attrition is not random (as is most often the case) it can bias the estimate of the treatment effect, and the bias can still be large even when advanced statistical methods like multiple imputations are used (Foster & Fang 2004; Puma et al. 2009). Such bias can distort the results of statistical significant tests and threaten the validity of any conclusion reached (Shadish, Cook & Campbell 2001; Campbell & Stanley 1963; Little & Rubin 1987).

Based on this, we should not use existing data to substitute for data that is missing, since we have little or no knowledge of the missing cases, and missing data/cases are seldom random. Doing so will increase the potential for bias. We therefore present differences in pre-test scores (KS2 Maths and Reading) between cases dropping out from both groups (where these are available).

In addition, we will report any missing data and compare the level of missing data to the number of hypothetical counterfactual cases needed to disturb the finding (Gorard et al 2017). The number of counterfactual cases will help determine whether the number of missing cases is large enough to alter/explain the findings (see explanation in section on Calculation of Sample Size).

Fidelity analysis

The fidelity to the intervention will be assessed by comparing the outcomes of pupils with the number of exercises completed (dosage). We will run an analysis using dosage (number of exercises completed) as the predictor.

The number of exercises will be used as a count variable in the analysis. This will be zero for all cases in the control group.

In addition, we will perform Complier Average Causal Effect (CACE) analysis to estimate the effects for the subgroup of treatment students who comply with their treatment assignment. Compliance is defined as completing the first writing task.

Secondary outcome analyses

Secondary outcome analyses will be comparisons of pre-, post- and gain score ‘effect’ sizes for:

- All pupils (i.e, both EverFSM and non-EverFSM) after 1 year
- All pupils (i.e. both EverFSM and non-EverFSM after 2 years
- Self-efficacy

Additional analyses

We will create two multivariate regressions, the first will use post-test scores (Atttainment 8 KS4 scores) as the dependent variables, and total prior test scores (KS2 maths and Reading) and membership of treatment group as predictors. The second model will include year group and FSM status as predictors.

Effect size calculation

'Effect' sizes will generally be calculated as Hedges' *g* based on the difference between mean post-test (and gain scores) for each variable. We will not report 'confidence intervals' but an interested reader can compute them if they wish as we will report the number of cases per group, and the effect size for each comparison.

'Effect' sizes for categorical variables (self-efficacy) will be based on post-intervention odds ratios – or changes in odds where the groups are clearly unbalanced at the outset. All will be presented with the number of counterfactual cases needed to disturb the results.

Protocol changes

1. Following the pilot trial, a few changes have been made to the initial protocol:
    - The non-attainment survey items have been revised and after testing the convergent and predictive validity of the scales as well as the internal consistency, it has been decided that the pupils' perceived self-efficacy will be the non-attainment outcome of interest. This will be measured using the subscales from the Motivated Strategies for Learning Questionnaire (Pintrich et al. 1993).

      The sub-group analysis now includes an analysis of self-efficacy as the non-attainment outcome

    - Pupils are individually randomised, stratifying by year group and FSM status only, but not by teaching class as originally planned.

2. Due to the recent changes in the GCSE exams we anticipate that there may be a delay in getting hold of the results. It is decided that the unconfirmed GCSE scores will be used to ensure that results are available on time for analysis. The expected date of completion for Report 1 will now be February 2018 (instead of January 2018) and February 2019 (instead of January 2019) for Report 2.

3. Because previous research and theory suggest that the intervention may be more relevant to those currently in receipt of FSM than those who have been eligible for FSM at some point in the last six years, it was decided to conduct an additional analysis for current FSM eligible pupils as well. The secondary outcomes will now include:

- Attainment 8 KS4 scores for FSM pupils (based on current FSM status) after 1 year of treatment (for initial Y11 pupils).

- Attainment 8 KS4 scores for FSM pupils (based on current FSM status) after 2 years (one year after the end of the intervention) for pupils who received the intervention when they were in Y10.

References

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago, IL: Rand McNally.

Dong, N. and Lipsey, M. (2011) *Biases in estimating treatment effects due to attrition in randomised controlled trials: A Simulation study*. SREE Conference, 2011.

Foster, M. E. and Fang, G. Y. (2004). Alternatives to Handling Attrition: An Illustration Using Data from the Fast Track Evaluation. *Evaluation Review, 28*:434-464.

Little, R. J. A., and Rubin, D. B. (1987) *Statistical analysis with missing data.* New York: Wiley

Mikaye, A., Kost-Smoth, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L. & Ilto, A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, 330, 1234-1237. doi:10.1126/science.1195996

Oyserman, D., Bybee, D., & Terry, K. (2006) Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology*, 91, 188-204. doi:10.1037/0022-3514.91.1.188

Pintrich, P.R., Smith, D.A.F., Garcia, T and Mckeachie, W.J. (1993) Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (Mslq). *Educational and Psychological Measurement*, 53, 3, 801-813.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629. doi:10.1037/0003-066X.52.6.613

This document is available for download at **www.educationendowmentfoundation.org.uk**

Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk