

# **A conceptual replication study of a self-affirmation intervention to improve the academic achievement of low-income pupils in England**

Beng Huat See<sup>1</sup>  
Rebecca Morris<sup>2</sup>  
Stephen Gorard<sup>1</sup>  
Nadia Siddiqui<sup>1</sup>

Matthew J. Easterbrook<sup>3</sup>  
Marlon Nieuwenhuis<sup>3,4</sup>  
Kerry Fox<sup>3,5</sup>  
Peter R. Harris<sup>3</sup>  
Robin Banerjee<sup>3</sup>

<sup>1</sup> School of Education, Durham University, UK

<sup>2</sup> Centre for Education Studies, University of Warwick, UK

<sup>3</sup> School of Psychology, University of Sussex, UK

<sup>4</sup> Faculty of Behavioural, Management and Social Science, University of Twente, The Netherlands

<sup>5</sup> School of Applied Social Science, University of Brighton, UK

## **Abstract**

Self-affirmation theory suggests that some potentially stigmatised groups of learners, such as those from ethnic minority or poor families, face possible stereotype threats which could undermine their academic performance. Engaging in value affirmation writing activities at times when such threats are most salient can give individuals a positive sense of value, negating harmful feelings, and fostering academic learning. An important test of the usefulness of such a theory is the replicability of evidence concerning its predictions. This paper describes an independently evaluated randomised control trial of a self-affirmation intervention, replicating earlier studies which were mostly conducted in the US with ethnic minority students. The present study, involving 5,619 Year 10 and 5,188 Year 11 pupils (age 14 to 16), assesses whether the promising results can be replicated in England with pupils from low socio-economic backgrounds. The intervention involved pupils writing about self-affirming values, delivered at three crucial time points before a key school assessment. The results showed that pupils from lower socio-economic background in the intervention group made slightly more progress between their KS2 scores (end of primary education exam) and KS4 (national exam at the end of secondary education) results than similar pupils who did not receive the intervention. There was a small, but positive effect (+0.05) for the Year 11, and a sustained effect for the Year 10 pupils a year after the intervention (+0.04). Pupils who completed more exercises also performed better. Consistent with theory and previous studies, the replicated intervention had no effect for the majority of pupils who are not labelled as disadvantaged. While the effects are small, the intervention may still be worth considering, given that it does not cost anything, does no harm, and could help reduce the poverty attainment gap.

**Keywords:** randomised control trial, self-affirmation, conceptual replication, stereotype threat

## **Introduction**

Replications are important for validating the trustworthiness of scientific findings, but they are rare in education. While there has been a recent increase in the number of randomised control trials and meta-analyses in education, few have been fully replicated. In the UK, the Education Endowment Foundation (EEF), a What Works centre, has commissioned over 160 trials - 10% of all known trials in education around the world since its inception in 2011 (EEF 2018). Although some of the more promising interventions have been scaled up from efficacy trials to effectiveness trials, few of these are direct or even conceptual replications. In the US, Slavin (2018) noted that half of all programmes in the What Works Clearinghouse are single evaluations. Relying on the evidence of single studies to accept or reject a programme is premature (Morrison 2020). Replications of single studies is needed to corroborate the initial findings, to overcome possible bias and errors in the original research, and confirm generalisability (Johnston and Pennypacker, 2009; Travers et al., 2016) to other contexts and populations. This is especially important for programmes that report positive results and have the potential to benefit pupil outcomes.

This new paper reports a conceptual replication of previous studies conducted in the US (Cohen et al., 2006; Sherman et al., 2013) to test the impact of self-affirming values on the academic attainment of low-income students in England. The new study replicates the conditions in terms of implementation and delivery as described in the initial studies from the US. It attempts to keep almost all the known conditions (outlined by Sherman et al. 2013) the same in terms of the timing, setting and “stealth” with which the intervention is delivered. In this respect, this is a conceptual rather than a direct replication (Morrison 2020; Hunter 2001). Such conceptual replications are useful in addressing generalizability (Morrison 2020; Earp and Trafimow 2015; Makel and Plucker, 2014).

## **Background**

Closing the attainment gap between rich and poor students is a policy issue relevant to many education systems in the world. The relationship between socio-economic status (SES) and students’ academic attainment is well-established. Young people’s PISA (Programme for International Student Assessment) outcomes for maths, science and reading can be predicted to a considerable extent by their SES (OECD 2019). In England, an ambitious policy initiative was introduced in 2011, which gave schools an initial £625 million of extra funding to close the attainment gaps for disadvantaged children (Gorard et al. 2021). This increased to £2.45 billion in the 2015-16 financial year (DfE 2015). Alongside this Pupil Premium (PP) funding was the establishment of the Education Endowment Foundation (EEF), a What Works Centre, to evaluate and identify promising programmes that can raise the attainment of the poorest children. These linked approaches were one of the most important recent developments in education in England. The PP funding is for schools to use for programmes or interventions to support the academic development of disadvantaged children (mainly children who are eligible for free school meals, but also those who had been in care or with parents in the armed forces). In England, children are eligible for free school meals if they live in households on income-related benefits, such as Universal Credit, Child Tax Credit, income-related employment and support allowance, income support, jobseeker’s allowance and asylum seeker support.

Evidence from studies, mostly conducted in the US, suggests that self-affirmation interventions can have positive and long-term results improving academic achievement, especially of those from ethnic minority backgrounds (Cohen et al., 2009; Good et al., 2003; Miyake et al., 2010;

Oyserman, Bybee and Terry, 2006; Sherman and Cohen, 2006; Steele, 1988; Wu, Spreckelsen, and Cohen, (2021). More recent studies showed that these effects persist through high school (Borman, Choi and Hall 2021; Borman et al., 2018) and right up to college (Goyer et al., 2017). Treatment students were more likely to enrol in college and competitive colleges than their untreated peers.

In light of such findings, the EEF funded an evaluation of self-affirmation writing exercises, aimed at improving the academic attainment of disadvantaged students at Key Stage 4 (KS4), principally the GCSE (General Certificate of Secondary Education) assessment. GCSE is a national standardised exam taken by 15/16 year olds at the end of their compulsory secondary education. The purpose of the trial was to see if similar results were produced for students in England who are eligible for free school meals (FSM), an indicator of low social economic status (SES). The theory suggests that the intervention is effective only for groups that experience stereotype threat. The focus of this evaluation is to replicate the conditions of the implementation of the self-affirmation intervention as used in the original studies by Cohen et al. (2006).

According to the notion of ‘stereotype threat’, students from some potentially stigmatised groups (e.g. students from disadvantaged backgrounds) are aware of the negative stereotype people have of them 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 the self-concept from being devalued by the negative stereotype (Spencer, Logel and Davies 2016). Disidentification results in academic achievement being discounted or devalued (Woodcock et al., 2012), and can reduce learning and motivation.

Self-affirming activities, such as writing positive statements about the values that are important to oneself, are believed to help protect students’ self-worth and free up cognitive resources so that they can engage more effectively with their learning (Oyserman, Bybee and Terry, 2006; Miyake et al., 2010; Sherman and Cohen, 2006; Steele, 1988). The theory is that such writing activities reinforce pupils’ sense of value, alleviating negative feelings they may have about themselves. The advantage of this approach is that no stigma is attached to individual pupils and the cost of delivery is minimal - initial training of teachers and any costs in printing exercise booklets or teacher manuals. If this approach is found to be effective in raising attainment for disadvantaged children it could prove to be attractive as it is almost cost-free, simple to implement, and would appear to generate few, if any, contra-indications. However, one needs to be cautious in how the intervention is implemented and who it is applied to. There is evidence that such an approach may be counterproductive for some groups where the factors affecting their academic performance are not psychological or social, or if it is not properly implemented (Binning and Browman 2020; Easterbrook and Hadden 2020; Walton and Yeager 2020).

Most of the studies conducted so far have been based in the US. The results are mixed but promising, and suggest that the intervention is particularly effective in raising the attainment of ethnic minority groups (Cohen et al., 2009; Cohen and Sherman, 2014; Sherman et al., 2013). Cohen et al. (2006), 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, African American students in the treatment group improved their Grade Point Average (GPA) score by 0.24 points, and low-achieving African-American students by 0.41. The intervention also appeared to reduce the likelihood of grade retention for lower-achieving African American

students. A longitudinal experiment (Sherman et al., 2013) showed that a self-affirmation intervention also benefitted Latino American students. Borman, Grigg and Hanselman (2015) also reported a positive impact on minority pupils' standardised maths test scores, while Mikaye et al. (2010) showed that self-affirmation can help to close the gender attainment gap.

However, other studies have shown no effects on either academic or other outcomes (Bratter, Rowley and Chukhray, 2016; de Jong et al., 2016; Hanselman et al., 2017; Protzko and Aronson 2016). There are subtle differences between these less promising studies and the ones by Cohen et al. (2006, 2009). Such differences provide hints about the delicate nature of delivering the intervention. Tweaks to the procedures of implementation from the original study can change the replication. For example, Simmons' (2011) study showed that students trained to use the self-affirmation strategy did not do better and were no more psychologically engaged than the control students. One important difference was that Simmons administered the intervention after the beginning of the term, whereas Cohen and his colleagues typically administered the intervention very close to the start of the term before students have 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, whereas Simmons administered the intervention in a different setting from their regular lesson (for example, a cafeteria or another room). Also, students were offered monetary incentives to complete the post-measure and this may have affected the apparent stereotype threat for participants. These differences could be important and suggest that the intervention is not simply about writing self-affirming statements. To be effective, these activities have to be carried out immediately prior to stressful events, such as before an exam, and as routinely as possible. This suggests that it is the conditions of delivery as much as the writing exercise that is the driver.

In the study by Protzko and Aronson (2016) study, the writing instructions were handed to students by researchers rather than teachers. The knowledge that it was a research exercise may have altered its impact. de Jong et al. (2016) also found no effects on school attainment of migrant children in the Netherlands despite close replication of the conditions of the earlier American studies (Cohen et al. 2009; Cohen and Sherman, 2014). One explanation could be the cultural context. Unlike in the US, where the ethnic minority students are largely either African American or Latino American, those in the Netherlands are of Turkish or Moroccan descent. They were likely to be Muslims and often chose 'religion' as an important value to reflect on in the self-affirming activity. Writing about their religion which has attracted negative media attention may have sometimes heightened their negative stereotype rather than reduced it. de Jong et al. also implied that cultural power distance (defined as the degree to which members of society accept their position in a hierarchical society) may explain why self-affirmation intervention may not work for certain cultural groups. Moroccan and Turkish students have a relatively high power distance (Hofstede, Pedersen and Hofstede, 2002), making it difficult for them to believe that they can change their situation. 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).

These studies suggest that the effectiveness of the intervention can depend on how the intervention is delivered. First, according to Sherman et al. (2013), it is important that the intervention should be seen as part of a normal classroom activity, and not billed as a stress-reduction or academic performance enhancement exercise. Awareness of the intent of the activity could reduce its efficacy. Second, the intervention should be administered at a period

when identity issues pose the biggest threat to the students. In the case of low-income and low-performing students, this threat is often associated with exams when students know that they will be judged by how well they perform in the exams (Hadden et al. 2019). Therefore, intervening just before students take their exams can help to break the cycle of negative self-belief. Finally, Sherman et al. stressed that it is important to consider the social and psychological context within which the intervention takes place. In some contexts (e.g. very disadvantaged schools) the stereotype threat may be less important than other structural barriers in students' academic performance, and in other contexts the threat may contribute less to the performance of ethnic minority groups than to other groups. Therefore, depending on the context, intervening with self-affirming values may not work as well.

Most of the studies cited were conducted in the US and focused on African American or Latino American students, and the results for white students were less promising (Cohen et al., 2006; Sherman et al., 2013). Although there is often a close relationship between socio-economic status (SES) and ethnicity, this link is perhaps weaker in England than in the US. Where academic disparities exist in the UK, it is more often examined along the lines of socio-economic status. The underachievement of white working class boys in England, for example, is well documented (e.g. Demie and Lewis, 2011; Strand, 2012). In the UK, for the 2019 GCSE cohort, 24.7% of disadvantaged achieved at least a grade 5 (considered a standard pass) in English and math compared to 49.9% for non-disadvantaged pupils, a difference of 25.2 percentage points. The difference between the lowest performing major ethnic group (Black students) and the White majority students, on the other hand was only 4.6% percentage points (DfE, 2019).

Although the interactions between SES and ethnicity are closely related both in the US and UK, in the UK, SES is a stronger predictor of attainment than ethnicity at all levels at school (Brannon, Higginbotham and Henderson, 2017; DfE, 2019, Harackiewicz, et al., 2016; Strand 2014). Hadden et al. suggest that an important factor for this might be the stereotype threat (2019). There is empirical evidence suggesting that early failure, such as experience by low SES children, can influence students' future performance and psychological state (Cohen et al., 2019; Fell and Hewstone 2015). A review of psychology research reported that studies show 'self-stereotyping' effects, in which the individual's perception of their membership of a particular group can influence their self-evaluation and psychological performance (Fell and Hewstone, 2015). Teachers' expectations of their students' academic potential can also result in self-fulfilling prophecies of students' academic performance (Rosenthal and Jacobson, 1968), known as the 'Pygmalion Effect'. And since students from low-income families tend to perform less well (on average) than their more privileged peers, teachers may underestimate their abilities. These are not conscious acts, but they can perpetuate students' feeling of inadequacy. A number of studies in the US have examined stereotype threat among students of different SES groups (e.g., Crozier and Claire, 1997; Spencer and Castano 2007). Using experiments, researchers compared the test outcomes of low and high SES students when tests are presented as being diagnostic of intelligence. They found that low SES students performed as well as other students when the test was not presented as a measure of their intellectual ability. This finding suggests that the disparities in performance between high and low SES observed in other studies might actually be the result of stereotype threat for low SES students.

Hadden et al. (2019) conducted a randomised control trial involving 562 pupils in England, showed that the self-affirmation approach works in raising the attainment of low SES pupils, reducing the attainment gap by 62%. This replication study tested if the intervention also

benefits poor children in England who may be stigmatised by their group's low academic performance, by simulating the exact conditions of administration immediately prior to exams. It was the first large-scale independently evaluated randomised control trial of the self-affirmation theory conducted in the UK, replicating the conditions used in the earlier studies in the US in its implementation and delivery.

## **The intervention**

Throughout our new trial, the intervention was referred to by its pseudonym, "Writing About Values" (WaV) to help mask the nature of the intervention somewhat. This is an important element of the intervention, because previous research has shown that knowledge of the purpose of the intervention can interfere with its efficacy (Yeager and Walton 2011). Therefore, every effort was made to keep the primary intention of the writing activity from both teachers and students (as agreed with the ethics panel).

The intervention was proposed, developed and implemented 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 the US (e.g. Sherman et al. 2013). The evaluation was conducted by independent evaluators from Durham University using national assessments at KS2 (a test taken at the end of primary school) and KS4 (Key Stage 4 - a test taken at the end of secondary education). While efforts were made to keep the intervention as close to those used in the US by Borman, Sherman and Cohen, there were some slight variations.

The intervention comprised three writing activities, each lasting 10 to 15 minutes, in which all students wrote short essays during their regular English lessons. These writing exercises were presented in booklets that were placed in named envelopes and distributed to pupils individually. For the first writing task the treatment group within each class wrote about values that were important to them, such as friendships and honesty. A list of values was provided for the pupils from which they could choose two or three to write about. Examples of such values included enjoying sports, being honest, and relationships with friends. The control pupils, on the other hand, wrote about values that might be important to other people. For the second writing activity, the treatment pupils wrote about things/people that mattered 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 were important to them and were asked to write about what they would do to show that these were important to them. or how much they enjoyed doing them. These values could be about relationships with friends, having a sense of humour, being with family, and following government and politics. An example of the writing task for the treatment and control groups is available in Appendix A1 and A2.

Short scripted instructions (Appendix B) were provided for teachers to introduce the task and to explain to pupils that the purpose of the activity was to get them to write freely. It was suggested that teachers say something like: "We will check to see if you've engaged with it properly, but it won't be marked. It's the process of writing about your own thoughts and ideas rather than me providing feedback. The exercises will be stored away". Pupils were told that there were no right or wrong answers, and that they did not need to focus on grammar or spelling; content and ideas were more important. Instructions were also available on the booklets. Pupils were encouraged not to talk to each other or look at their neighbours' writing

task while writing. Teachers were given strict instructions to use the prescribed answers to pupils' queries about the purpose of the exercises.

The exercises were delivered by English language teachers as a whole-class activity as part of their regular English lessons and collected in the envelopes by the teachers at the end of the session. Efforts were made to ensure that these exercises were delivered as naturally as possible to avoid pupils linking them to a research project. The researchers who conducted a light observation of the delivery of the intervention in some classes were blind to treatment conditions in that they would not know which were treatment or control pupils because all pupils were involved in a writing exercise that differed only in terms of content seen only by themselves.

In line with the theory of self-affirmation, the writing exercises were delivered during three crucial time points: once at the beginning of the academic year (before the experience of negative stereotype was established), and again before potentially stressful events. In this study these were the mock GCSEs and the actual GCSEs exams later in the year. The timeline of the three writing activities is presented in Table 1. The original Cohen et al., 2006 paper used 3-5 exercises, the Sherman et al. 2013 used 4 or 5, and Borman et al., 2018 did 'up to four' exercises. We opted to implement the interventions at the most obvious high stress times – the beginning of the year, before mocks and before exams – which is when the interventions are supposed to be most effective, while keeping the administrative load on schools to a minimum. Ultimately, the number of doses does not impact the effectiveness of the intervention (Wu et al. 2021). It is the timing of the intervention that is crucial.

The baseline survey was conducted in the first two weeks of the term. The aim was for all exercises to be completed within the following two weeks of the term. More than two thirds of pupils completed between 19<sup>th</sup> and 23<sup>rd</sup> of September, and just 5% were completed within the first week of October. Because the exercises are theorised to be most effective when they are delivered before a stressful event, we tied the second exercises to be implemented within the two weeks before the start of each school's mock GCSEs rather than a specific timeframe, and this varied across schools. Ninety percent of schools completed the second exercise before 23<sup>rd</sup> February. For the final exercises, the aim was to deliver them in the two weeks before the start of the GCSE exams. Eighty percent of pupils completed them between 24<sup>th</sup> April and 24<sup>th</sup> May, and 97% by the end of June.

Table 1: Timeline of delivery of writing activities

<b>Date</b>	<b>Activity</b>
<b>5 September 2016</b>	Delivery of first writing task
<b>December 2016/January 2017</b>	Delivery of second writing exercise before mock GCSEs
<b>May 2017</b>	Delivery of third writing exercise before GCSEs

To protect the integrity of the intervention, a number of strategies were employed to safeguard the precise nature and purpose of the intervention. Observation visits, for example, were kept to a minimum to avoid pupils linking it to a research project. The presence of evaluators in the classroom was explained as part of a programme to observe how English was being taught. Feedback from pupils about the intervention was obtained only from the Year 11 and only after

their KS4 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. The briefings to teachers were presented in a very general way and although occasional reference was made to the evidence-based nature of the intervention, teachers were not given the detailed background of the intervention at that stage. Instead, the briefing focused 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. All material was made available after the post-intervention tests.

### *Ethics*

There were no particular ethical issues raised with the trial because the intervention was introduced by teachers voluntarily as part of normal classroom activity. The ethics committees at Durham University and the University of Sussex agreed that the obscuring of the precise nature of the intervention was justified by its intent, and the extremely small chance of harm (based on prior studies). All pupils took part and all wrote about values (with only the precise nature of the task varying). This writing about values was how the intervention was explained and introduced.

### **Methods used in the evaluation**

This was a two-year, double-blind randomised control efficacy trial involving Year 10 (age 14–15) and Year 11 (age 15–16) pupils from 29 schools in the South East of England. Pupils (and their teachers) had no knowledge of whether they were in the treatment or control group. This was possible because control groups were given a writing activity as a placebo, which was also a viable alternative approach to writing about values. Teachers were not told what the intervention was, and students were not aware that it was a research activity.

Due to changes in the processes of access to the national pupil database (NPD) in the second year of the trial, data from both cohorts could not be merged. The two cohorts of pupils were therefore evaluated, and their results are presented, separately. Analysis for the Y11 cohort was undertaken at the end of the year while that for the Y10 cohort was taken a year later when they sat for their KS4 assessments. The inclusion of Y10 allowed us to assess the longer-term impact of the intervention.

### *Research questions*

1. What impact does the self-affirmation intervention have on the academic attainment of disadvantaged pupils, defined as those who were eligible for free school meals at some point in the last six years (EverFSM6), after one year of treatment? [This replicates the earlier studies from the US, but applying the intervention to disadvantaged pupils in England rather than potentially stigmatized groups, e.g. ethnic minority pupils in the US]
2. Is the impact for EverFSM6 pupils sustained after two years (one year after the end of the intervention)?
3. Does the self-affirmation intervention have any impact on the general pupil population (including not EverFSM6)?
4. Is the impact for all pupils (EverFSM6 and non-EverFSM6) sustained after two years?



EverFSM6 was used as a measure of socio-economic disadvantage because this was the definition underlying the distribution of Pupil Premium funding to schools (see above).

### Sample

The trial was conducted in 29 secondary schools across the South East of England with a total of 5,619 Y11 and 5,188 Y10 pupils. The schools recruited were those not in ‘special measures’ (i.e. at risk of failure), and that had a minimum of 10% of pupil population eligible for FSM. Pupils were individually randomised within schools, stratified first by year and then by FSM status to either the treatment or control conditions. This was to help ensure initial equivalence between the two groups.

Table 2 details the number of participants at randomisation and subsequently. The key figures are for the headline based on EverFSM6 pupils. There was no attrition from Year 10 pupils with pre-intervention scores, and just under 10% from Year 11 pupils with pre-intervention scores. Cases were only missing if they could not be found on the National Pupil Database (NPD), and so it did not matter if they moved to another school in England after the randomisation. As this is an intention-to-treat design, all pupils in the original design were included in the analysis even if they were no longer in one of the 29 schools in the trial. It was not clear why a number of Year 11 pupils could not be found by the DfE (Department for Education). One possible reason would be mistakes in the unique identifiers provided by the schools, or pupils who had just arrived in the country and not been given an identifier. Pupils who received their primary education overseas or in an independent school would also not have taken the KS2 exams. For the Y10 cohort, we excluded all pupils without KS2 scores. KS2 is a national exam that pupils take at the end of their primary school. We used KS2 test scores as the baseline assessment.

Table :2 Participants by year group and treatment conditions (N= 29 schools)

	Intervention pupils	Control pupils
Randomisation Y10 all	2,569	2,619
Randomisation Y11 all	2,809	2,810
Randomisation Y10 EverFSM6	674	698
Randomisation Y11 EverFSM6	706	800
Analysed Y10 EverFSM6	674	698
Analysed Y11 EverFSM6	640	711

This sample size of 1,380 (EverFSM6) individually randomised cases in the intervention groups (the smallest cells) is traditionally large enough to detect an effect size of just over +0.1 (which is small for a typical education intervention). However, we do not use traditional power calculations as these are based on an erroneous assumption (Gorard, See and Siddiqui, 2017). Instead, we calculated 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 2018). This number needed to disturb (NNTD) is 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). Therefore, the smallest detectable effect size for any NNTD is NNTD divided by the size of the smallest group. Based on Gorard (2018, pp. 12-13), NNTD of 50 can be considered a very strong and secure finding. Using this as a

working assumption, a sample of 1,380 might enable us to detect an effect size as little as 0.04 with considerable confidence.

### *Outcome measures*

To test the effect of the intervention on the academic attainment of disadvantaged groups of pupils, we used the KS4 Attainment 8 scores of pupils who were eligible for FSM at any point in the last six years (EverFSM6). Attainment 8 is used in England as a measure of students' academic performance in the last year of their compulsory secondary education. It is the student's average grade across the best eight subjects. Evaluation of impact for Year 11 was undertaken at the end of the first year following release of their results, whereas impact evaluation for the Year 10 cohort was completed a year later. This allows us to see if the effect (if any) was maintained one year after the intervention.

The KS2 results for maths and English (national tests taken at the end of primary school) were used as a pre-test measure of pre-intervention equivalence. Impact analysis was based on the combined reading and English scores as per pre-trial protocol, agreed with the funder.

The attitudes theory suggests that the intervention is effective only for groups that experienced stereotype threat (e.g. pupils from disadvantaged backgrounds). To test this theory, we also analysed the attainment outcomes of the general pupil population. This includes both EverFSM6 and not EverFSM6 pupils.

### *Analysis*

Pupil attainment was analysed using intention-to-treat. This means that all pupils randomised to receive the intervention were included in analysis regardless of whether they were known to receive the intervention or not. The impact of the intervention was measured as the difference between intervention and control groups in terms of the progress scores between average KS2 results for maths and reading and KS4 Attainment 8 outcomes. For comparability between phases, the test scores were converted to standardised z-scores. The pre and post-intervention differences are expressed as simple effect sizes (difference between means divided by their overall standard deviation). The advantage of using progress scores is that it addresses any initial imbalance in prior attainment created inadvertently by the randomisation. Significance tests and confidence intervals are not reported here as they are not relevant, because attrition means that the cases are no longer completely randomised. And even with complete randomisation such tests are still not appropriate because null hypothesis significant testing (NHST) states that assuming there is no difference between groups how likely are we to obtain the data as extreme as observed. The answer that most researchers want is: given the data how likely is there a difference between groups. Unfortunately, significant tests do not and cannot answer this question. Using such tests to judge if there is a difference between groups is therefore misleading. For further explanations, see Gorard, 2021; Colquoun, 2014, 2016; Perezgonzalez, 2015; Pharoah et al., 2017).

There should be no issue of clustering as randomisation was at the individual level within schools rather than at the school level. Analysis is of all pupils in the two groups and not by schools. The mean scores of all the pupils in the control group and treatment group in all schools or if computed class by class, would be the same as the mean scores of all treatment and control pupils in the whole trial, by definition.

To account for missing cases or missing data, which can potentially bias the results (Dong and Lipsey 2011; Foster and Fang, 2004; Little and Rubin, 1987; Puma et al., 2009; Shadish, Cook and Campbell, 2001) we presented differences in pre-test scores (KS2 maths and reading) between the cases dropping out from both groups (where these were available). Actually, this was not dropout as such, rather that the DfE did not find later results in NPD (see above). In addition, we also conducted a sensitivity analysis to see how much these missing cases would skew the results if they were included. To do this, we first calculated the number of counterfactual cases needed to disturb the headline finding (NNTD, as above). The number of counterfactual cases determines whether the number of missing cases is large enough to alter/explain the findings (see section on sample size above). The bigger this number is the more stable is the substantive result, as this means it will take this many counterfactual cases to reduce the effect size to zero.

### *Dosage and complier analysis*

Not all pupils completed the three writing activities. So, we carried out two further analyses to test the impact of dosage. The first was a correlational analysis comparing the outcomes of pupils with the number of exercises they completed. The latter would be zero for all cases in the control group. Information about dosage was collected by the project delivery team (the developers) who kept a log of the number of exercises completed by each pupil.

Further analysis was carried out to estimate the effects for the subgroup of treatment students who complied with their treatment assignment using the Complier Average Causal Effect (CACE) analysis. Minimal compliance here is defined as completion of the first writing exercise (according to the developers) because theoretically the first writing exercise is supposed to have the most impact (Cohen and Sherman, 2014; Garcia and Cohen, 2012) and 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.

CACE compares the average outcome of treatment pupils who complied with the control pupils who it is estimated would have complied if given the treatment (Nicholl, nd; Dunn, 2010). Table 3 illustrates how CACE is estimated. Given that we know the overall results for both groups (cells F and K) and the mean scores for those in the treatment group who complied and who did not comply (Cells A to D), we can calculate the average outcome for those in the control group who would have complied if given the treatment ( $x$ ). We assume that because of randomisation, the proportion of compliers in both arms of the trial would be 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). They are unaffected by the intervention.

The proportion in treatment group who complied will be  $A/E$ . The number who complied in the control group (Cell G) will be  $A/E * J$ . The number of non-compliers in the control group

(Cell H) will be  $J - G$ . The average outcome for compliers in the control group ( $x$ ) is thus  $((J * K) - (H * I)) / G$ .

Table 3: Estimation of Complier Average Causal Effect

	Compliers		Non-compliers		Overall	
	N who completed first writing exercise	Mean	N who did not complete first writing exercise	Mean	N	Mean
Treatment	A	B	C	D	E	F
Control	G	$x$	H	I	J	K

### *Process evaluation*

We also carried out a light touch process evaluation to collect information about teachers' delivery of the intervention, staff and student's views of the intervention and indications of any possible contamination or diffusion. This is not the focus of this paper. The main method of data collection was classroom observations. These were 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, the extent to which they adhered to the instructions for delivering the exercises, that the right pupils were given the correct writing exercise, and if there was any possibility that pupils could swap exercises with their classmates. Observation visits were made to classes in five schools. The number of visits was deliberately kept small to avoid alerting pupils to the research element of the writing activity.

We also had a number of informal conversations with teachers in schools to find out if they had observed any changes in pupil behaviour and to gather their views on the writing exercises. However, because of the nature of the intervention and the restrictions in what teachers knew about the overall project, these conversations were limited in scope and focused predominantly on teachers' views of the writing tasks and the children's reactions to them. The process evaluation was also intended to find out indirectly if teachers and pupils had any knowledge of the intervention.

Due to the nature of the research, pupils were not interviewed while the trial was still running. A small number of Y11 pupils were contacted by their teachers after their GCSE exams via emails through their parents inviting them to respond to a short questionnaire asking for their views on the writing activity.

### **Impact results**

To evaluate the impact of self-affirmation on the academic outcomes of EverFSM6 pupils we compare the gain scores for the control and treatment pupils between KS2 and KS4 Attainment 8 for EverFSM6 pupils only. Since KS2 scores and Attainment 8 scores are not on the same metric, for comparability we converted all to standardised z-scores before analysis. The negative scores show that FSM-eligible pupils in general perform below the average for their cohort. Analysis was performed for pupils who have both pre-test scores for reading and maths and post-test scores.

Table 4 shows the number of pupils with pre-test scores and the number of missing pre-test in both intervention and control group. A total of 10,807 pupils (Year 11 = 5,619; Year 10 = 5188) had post-test scores, but some were missing pre-test KS2 maths and some missing KS2 reading scores. Impact analysis was conducted for 10,274 pupils (Y11 = 5,086; Y10 = 5169) who have both KS2 maths and KS2 reading scores. The table shows that the two groups were not closely balanced at pre-test. Therefore, the gain score results are used for the headline findings.

Table 4: Comparison of pupils' baseline characteristics

Characteristics of pupils at randomisation	Intervention		Control		
	n/N (missing)	Percentage	n/N (missing)	Percentage	
Proportion of boys					
Year 11	1430/2823 (0)	50.7	1393/2823 (0)	49.3	
Year 10	1225/2521 (0)	49.0	1296/2521 (0)	51.0	
Proportion of pupils eligible for EverFSM 6					
Year 11	640/1351 (0)	47.4	711/1351 (0)	52.6	
Year 10	674/1372 (0)	49.0	698/1372 (0)	51.0	
Proportion of current FSM pupils					
Year 11	305/ 678 (15)	45.0	373/678 (21)	55.0	
Year 10	307/635 (22)	48.0	328/635 (25)	52.0	
Proportion of pupils with SEN					
Year 11	421/843 (0)	49.9	422/843 (0)	50.1	
Year 10	401/806 (22)	50.0	405/806 (25)	50.0	
Proportion of pupils whose first language is not English					
Year 11	314/652 (2)	48.2	338/652 (9)	51.8	
Year 10	229/466 (14)	49.0	237/466 (16)	51.0	
<b>Raw means</b>					
	Intervention		Control		
	n/N (missing)	Mean (sd)	n/N (missing)	Mean (sd)	Effect size

KS2 Maths (point scores) Year 11 Year 10	2621/5619 (189) 2569/5188 (0)	4.65 (0.79) 4.68 (0.78)	2620/5619 (189) 2619/5188 (0)	4.63 (0.80) 4.68 (0.81)	+0.03 0.00
KS2 Reading (point scores) Year 11 Year 10	2547/5619 (263) 2569/5188 (0)	4.66 (0.74) 4.70 (0.77)	2539/5619 (270) 2619/5188 (0)	4.64 (0.78) 4.69 (0.79)	+0.03 +0.01
KS2 Reading and Maths combined Year 11 Year 10	2547/5619 (263) 2569/5188 (0)	9.31 (1.41) 9.38 (1.42)	2539/5619 (270) 2619/5188 (0)	9.28 (1.45) 9.37 (1.47)	+0.03 +0.01
<b>EverFSM6 pupils</b>					
KS2 Maths (point scores) Year 11 Year 10	679/1431 (27) 674/1372 (0)	4.45 (0.70) 4.40 (0.76)	752/1431 (48) 698/1372/ (0)	4.50 (0.75) 4.43 (0.80)	-0.07 -0.04
KS2 Reading (point scores) Year 11 Year 10	640/1351 (66) 674/1372 (0)	4.48 (0.69) 4.44 (0.83)	711/1351 (89) 698/1372/ (0)	4.52 (0.68) 4.44 (0.83)	-0.05 0.00
KS2 Reading and Maths combined Year 11 Year 10	640/1351 (66) 674/1372 (0)	8.92 (1.26) 8.84 (1.45)	711/1351 (89) 698/1372/ (0)	9.02 (1.31) 8.85 (1.47)	-0.11 -0.01
<b>FSM pupils</b>					
KS2 Maths (point scores) Year 11 Year 10	290/678 (15) 307/635 (0)	4.31 (0.79) 4.32 (0.74)	357/678 (16) 328/635 (0)	4.32 (0.79) 4.32 (0.81)	-0.01 0.00
KS2 Reading (point scores) Year 11 Year 10	271/678 (34) 307/635 (0)	4.35 (0.80) 4.39 (0.85)	333/678 (40) 328/635 (0)	4.33 (0.78) 4.33 (0.88)	+0.03 +0.07
KS2 Reading and Maths combined Year 11 Year 10	271/678 (34) 307/635 (0)	8.66 (1.47) 8.71 (1.45)	333/678 (40) 328/635 (0)	8.65 (1.46) 8.65 (1.52)	+0.01 +0.04

Table 9 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 5 shows that both FSM-eligible groups made less than average progress between KS2 and KS4 (compared to the full cohort), but compared to the treatment group, the control group made even less progress. This suggests that the intervention may have a small influence in improving the performance of the EverFSM6 pupils. The effect sizes for both Year 10 and Year 11 cohorts are positive (+0.05 for Y11, +0.04 for Y10). Even one year after the intervention ceased a small positive “effect” has been sustained. Earlier field experiments (e.g. Cohen et al. 2009) suggest that these alterations in psychological states and performance provide the initial trajectory for a recursive process, and the changes in attributions and information processing it prompts can become self-reinforcing or self-sustaining over time.

Table 5: 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)

	<b>Pre-score mean</b>	<b>SD</b>	<b>ES</b>	<b>Post-score mean</b>	<b>SD</b>	<b>ES</b>	<b>Gain score</b>	<b>SD</b>	<b>ES</b>
	Pre-score mean	SD	ES	Post-score mean	SD	ES	Gain score	SD	ES
Treatment Year 11	-0.32	0.98		-0.42	0.93		-0.10	0.89	
Year 10	-0.37	1.01		-0.47	0.91		-0.10	0.85	
Control Year 11	-0.26	0.99		-0.39	0.94		-0.14	0.84	
Year 10	-0.36	1.02		-0.49	0.88		-0.13	0.84	
Overall Year 11	-0.29	0.99	-0.06	-0.41	0.94	-0.03	-0.12	0.87	+0.05
Year 10	-0.37	1.01	-0.01	-0.48	0.89	+0.02	-0.11	0.85	+0.04

Note: The scores are standardised z-scores, not raw scores. The negative signs indicate that the students are performing worse (in relative terms) than the average for the cohort. The comparison is with the cohorts, not between the randomised groups. As expected, FSM-eligible pupils perform worse than their non-FSM peers.

We also calculated the number of counterfactual cases (i.e. number of cases with counterfactual results) that would be needed to eliminate the positive effect. For the Year 11 cohort, this number is 32 (0.05 multiplied by 640). This means it would take approximately 32 missing cases with strong counterfactual scores (see methods) in the opposite direction for the findings to change. For the Year 10 cohort, the number of counterfactual cases is 27 (0.04 multiplied by 674). However, since there were no cases with pre-tests missing post-test scores, this means that the finding cannot be due to attrition even in the worst case scenario. Although the effects are small, they are therefore reasonably secure.

To examine whether self-affirmation had any impact on the general pupil population (not just FSM pupils), we compared the gain scores of treatment and control for all pupils. The analysis shows no differential benefit for either group, indicating that the intervention has no impact on the overall pupils' Attainment 8 scores (-0.01 for Year 11 and 0.00 for Year 10) (Table 6). This is consistent with the theory that self-affirmation only works with pupils experiencing stereotype threat.

Table 6: 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 Year 11	0.005	1.005		0.047	0.97		0.042	0.80	
Year 10	0.005	0.98		0.015	0.99		0.009	0.79	
Control Year 11	-0.005	0.995		0.042	0.97		0.047	0.81	
Year 10	-0.005	1.02		0.004	1.00		0.009	0.80	
Overall Year 11	0.000	1.000	+0.01	0.045	0.97	+0.01	0.045	0.81	-0.01
Year 10	0.000	1.000	+0.01	0.009	0.99	+0.11	0.09	0.80	0.00

### Dosage and complier analysis

To test whether the number of exercises completed made a difference to the outcomes, we compared the number of exercises completed (dosage) with the gain scores as well as the Attainment 8 scores (as post-test only). The number of exercises is treated as a continuous variable. For the control group this will be zero as they did not complete the intended writing activity. Over 60% of the Year 10 intervention pupils completed all three writing exercises, while only 53% of Year 11 cohort did (Table 7).

Table 7: Number of exercises completed by intervention group

Number of exercises completed	Year 10	Year 11
0	128 (5.0%)	111 (4.4%)
1	211 (8.2%)	299 (11.8%)
2	626 (24.2%)	775 (30.4%)
3	1581 (61.5%)	1362 (53.5%)
Missing	23 (0.0%)	
Total	2,569	2,547



Correlation analysis shows a small positive relationship between number of exercises completed and the gains made between pre-test (KS2 scores) and post-test (Attainment 8). The results are similar whether using gain scores or Attainment 8 post scores (Table 8). The relationship is stronger for the Year 10 pupils (+0.36) than for the Year 11 pupils (+0.16). This may suggest that the lasting effect of the intervention is stronger the more exercises pupils complete.

Table 8: Correlation between gain scores and number of exercises completed (EverFSM6 pupils only)

Number of exercises completed	Gain scores using KS2 maths and reading GCSE Attainment 8 score combined			
	Year 11	Year 10	Year 11	Year 10
	+0.09	+0.26	+0.16	+0.36

For the impact evaluation, we use an intention to treat analysis, meaning that all those randomised to treatment are analysed as being in the treatment group, even if they did not receive the intervention. This helps to preserve the prognostic balance afforded by randomisation (McCoy 2017) as those who adhere to the protocol differ in some ways from those who do not (Montori and Guyatt 2001), and ensures an unbiased estimate of the efficacy of the intervention on the primary outcome at the level of adherence observed in the trial. However, in reality not all pupils who received the intervention complied with the intervention. Compliance is defined here as completion of the first writing task because it is deemed most impactful (see above). Complier analysis, therefore, is to see if pupils who complied with the intervention do better than those who did not.

We analysed the effect of compliance using the Complier Average Causal Effect Analysis (CACE) based on the standardised gain scores and using the overall standard deviation in Table 5 (0.87 for Year 11 and 0.85 for Year 10). The result shows a small positive effect size (Table 9), exactly the same as the impact evaluation result in Table 4.

Table 9: 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							
Year 11	526	+0.01	114	-0.60	640	-0.10	
Year 10	549	-0.005	125	-0.56	674	-0.10	
Control							
Year 11	<i>583</i>	<i>-0.04</i>	<i>128</i>	<i>-0.60</i>	711	-0.14	+0.05
Year 10	<i>569</i>	<i>-0.03</i>	<i>129</i>	<i>-0.56</i>	698	-0.13	+0.04

Note: the figures in italics are based on there being the same proportion of compliers in the control group as in the treatment group, 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.

## **Process evaluation summary**

The process evaluation was conducted primarily to ensure that the intervention was implemented with fidelity. It does not form part of the impact evaluation, although its findings may help explain the results. For example, if teachers or pupils became aware of the purpose of the intervention or if the writing tasks were not delivered at the three crucial time points, it could affect the results. For this reason, we did not think it was necessary to do frequency count of number of teachers or pupils in each event mentioned below, but simply to capture the general views of staff and pupils.

### *Teacher briefings*

To ensure that the intervention was carried out as intended, and its covert nature was observed, we attended three teacher briefing sessions to understand how the intervention was explained to the teachers. These briefing sessions were hour-long meetings attended by Y10 and Y11 English language teachers during which the project developers presented a short introduction 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 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. One teacher, however, did question the premise of the research, to which the project team simply repeated information from the introduction revealing no extra information. The focus of the briefing session was to explain to teachers how they should go about delivering the writing exercises in English classes. Teachers were clear about the task and the need to ensure that named envelopes were given to the right pupils.

### *Classroom observations*

We observed a total of ten classes across five schools for the first writing exercise, which was delivered at the start of the academic year, one class for the second exercise, and a further five groups were observed across another two schools for the final writing exercise. The number of classes and schools we could observe depended on school's availability.

No evidence of diffusion in terms of pupils' swapping writing tasks was observed. In schools where we could not observe, teachers reported no issues with the administration of the writing task. However, there was one school where the pupils became suspicious with one pupil saying, 'this is so random and confusing, it's a conspiracy'. Another questioned whether it was 'some kind of social experiment'. Other pupils noted that they had different questions to their neighbours. Although they were told to work individually, some students were observed talking to their peers. The 'secret envelopes' also aroused some suspicion. One pupil commented that it was 'very dodgy' and there was vigorous questioning about who was going to read their work. The class teacher, an experienced head of department, stuck closely to the guidance provided and emphasised the whole-school nature of the project and that this was something that other schools were doing as well. For details about the standard response that teachers were told to use, see Appendix B.

One thing that most concerned pupils was whether their work would be marked and who was going to look at their work. Similar concerns were noted in all the schools we visited. Many

were puzzled as to why their work would not be marked and whether they would get feedback on their writing. In all such situations, the teachers explained to the pupils that their work would not be graded and no one would be looking at what they had written, and their writing would simply be stored away. 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 contradictory to the usual advice they received from their English language teachers, particularly in the lead-up to GCSEs where they were preparing to be assessed on these skills. In all cases the teachers responded to these queries as per the instructions from the sheet provided by the developers (Appendix B). Fortunately, teachers were not able to betray the true intention of the intervention because they had also not been told.

In a few instances teachers attempted to link the writing exercise and the topic being studied in that term so the writing activity seemed like a natural part of their curriculum. One teacher, for example, adeptly fitted in the writing exercise into 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.

### *Teachers' views*

There was also no evidence that teachers had knowledge of the real purpose of the intervention. 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. Most of the teachers we spoke to thought the intervention was a 'free writing' exercise and welcomed the opportunity as they felt it provided a pleasant alternative to the very structured, exam-focused work that Key Stage 4 pupils usually undertake. 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 were 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. Tying the intervention in with the English lesson worked well.

### *Pupil views*

To capture the views of pupils, a short questionnaire was sent out via emails through the school to pupils after their GCSEs. As pupils have already left school after their GCSE, these emails were sent to their parents' email addresses. Only six pupils responded, and these presented quite mixed perspectives. 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 were '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. Given the small number of responses, we could not read much into these comments, but they do give us some food for thought.

*What are the challenges teachers faced in delivering the intervention?*

One of the biggest challenges was in scheduling and delivering the writing exercises around Key Stage 4 mock exams and actual GCSE exams. One school did not complete the third writing exercise as the English teachers felt that the time was needed for revision before the GCSE. Despite several attempts 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 exercise before the final exercise. A small number of classes within some schools also did not complete at least one of the tasks. Ensuring tasks were completed if students were absent from the original English lesson was another challenge mentioned by some teachers.

For the majority of the teachers, the task was not seen as too much of an imposition as it took only 10 to 15 minutes and was delivered only three times in the year. It was simple, quick, and easy to deliver. Generally, the intervention fitted really well within the curriculum and its covert nature made it easier to be seen as part of the regular English lessons. On the whole, the intervention appeared to have been delivered as intended. Teachers closely followed the guidance and instructions provided in the way they handled pupils' queries. Therefore, for effective implementation of the intervention, it is important that teachers are thoroughly briefed. In this trial, the developers gave very clear verbal and written instructions to ensure that teachers adhered to the protocol. Additional telephone and email briefings were offered for teachers.

## **Conclusions**

This trial shows that disadvantaged pupils who received the intervention made very slightly more progress between KS2 and KS4 than pupils who did not receive the intervention. In line with theory, the intervention shows no benefit for the general pupil population (that is including non-disadvantaged pupils). 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 (Cohen et al., 2006; Hadden et al. 2019; Miyake et al., 2010). The effects, while positive, are small.

Previous evidence also suggests that the effects of the intervention could last for several years. We tested this with the Year 10 pupils a year after the intervention ended. The results show that the small positive effects of the intervention are sustained over one year.

There is no standard interpretation of effect sizes, and any effects must be considered in relation to costs, opportunity costs, and unintended outcomes. Given that the intervention takes under 20 minutes, may be useful in its own right, is delivered three times a year and costs almost nothing, there is hardly any opportunity cost for schools. Although the impact is small, the positive correlation between number of exercises completed and the outcomes, plus the fact that the impact was sustained, all suggest that the intervention remains worth considering as

there are currently no contra-indications or side effects. However, caution needs to be taken in deciding on the groups to which the intervention is administered. First, it does not benefit all children. Second, the evidence so far is that the intervention is beneficial only for groups in which the negative stereotype effect on their academic performance is psychological or social (Binning and Browman 2020; Easterbrook and Hadden 2020).

There is a problem that would need to be addressed if this intervention were to be rolled out more widely. As students and staff become more familiar with it, and more aware of the benign intentions, it may become even less effective. Scaling it up effectively becomes a new project in itself. Replication of the study will no longer be tenable as awareness of the purpose of the activity could reduce its efficacy. This is the challenge of such an intervention.

What this study demonstrates is that post-hoc conceptual replications are feasible with an intervention like self-affirmation where the intention and nature of the intervention have to be concealed from the participants, and its delivery is highly prescribed to maintain its integrity. The study closely adheres to the conditions of implementation in terms of its stealth, timing and setting. In line with the theory of self-affirmation, the writing exercises were delivered once at the beginning of the academic year (before the experience of negative stereotype was established), and prior to a stressful event before the final mock GCSEs (for the Y11 pupils) and the actual GCSEs exams (for the Y10 pupils) in normal classroom conditions. The findings suggest (weakly) that the intervention works with disadvantaged pupils in England just as it did with ethnic minority pupils in the US. In other words, the benefits of such value affirming activities can be effectively generalised to other contexts outside the US with other groups facing stereotype threats. Consistent with the initial studies of Cohen et al. (2006; 2009) and Sherman et al. (2013), the intervention has no benefit for the general population.

### *Limitations*

As with any research there are limitations and compromises. 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 SEND (special educational needs or disability) 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 have lower-attainment, on average, for the same reason. They have a lower proportion of pupils achieving five A\*–C at GCSE, or equivalent, compared to the national average. They are also more likely to have a higher proportion of White British pupils and a lower proportion of EAL (English as an additional language) pupils. Therefore, the results may not be as applicable to all other schools, such as those in London or the Midlands, where the demographics may be different.

Another limitation is the use of EverFSM6 as a proxy for disadvantage. As shown by many studies (e.g. Gorard, 2012; Hobbs and Vignoles, 2010; Taylor, 2018), snapshot FSM is not a reliable measure of disadvantage for a number of reasons. There is also a big disparity between those who are long term eligible and those who are temporarily eligible (Gorard, 2018). Short-term eligible pupils, while labelled disadvantaged, have higher average attainment than pupils with longer-term eligibility. The long-term FSM-eligible are more clearly disadvantaged. Therefore, using EverFSM6 as a measure of SES may not accurately reflect the full impact of such an intervention, which is to address negative experiences associated with enduring membership of a disadvantaged group. Perhaps a more accurate measure would be parental

income or occupational status, or permanent FSM status (pupils who have been eligible for FSM for most of their school life), but these figures were not available here.

Thirdly, the agreed prior attainment used for the analysis was the combined results for reading and maths. Our analysis shows that the results vary slightly whether using reading ( $ES = +0.04$ ) or maths ( $ES = +0.02$ ), with the combined maths and reading showing a slightly bigger effect ( $ES = +0.05$ ). We cannot be sure that the same effects would be achieved if different measures (e.g. English rather than reading) were used for both the prior attainment and the post-test.

There are also a number of differences between the current study and previous ones. One is that studies conducted in the US invariably use GPA as the outcome measure. Because GPA is a continuous assessment throughout the year, it provides students with almost continuous feedback on their performance, which could reinforce the cycle of adaptive potential through which recursive effects are thought to operate (Cohen & Sherman, 2014). There is no equivalent measure to GPA in the UK. The process through which the intervention operates could be different for GCSEs, which are high stakes, stressful, and one-shot exams. This may have dampened the effect. Indeed, other studies that use test scores as the outcome, e.g. Borman et al. (2016), have also shown effects that were small.

Unlike previous US studies where the final exercise was personalised, we were unable to personalise the values for the final exercise simply because we did not have the capacity with the very large sample. This could have diluted effects - a realistic limitation for scaling up.

## *Discussion*

Despite the increase in experimental studies in education in the past two decades, few studies have been replicated so far. Makel and Plucker (2014) noted that only 0.13% of around 16,000 studies in top 100 education journals were replications. Of these, 29% were direct replications and the rest were conceptual replications. One reason for this could be that journals, or indeed reviewers, look for articles that are deemed “original” in terms of concepts and analysis. Replication studies tended to be viewed as lacking in originality and so not contributing to new ideas. We think this widely held view is flawed and, as Makel and Plucker argued, this is a serious misunderstanding of science and creativity, privileging novelty over trustworthiness. Being able to verify the results of previous studies is a cornerstone of scientific rigour. However, this does not mean that every study needs to be replicated. But it is important that studies reporting positive results or that have the potential to influence student outcomes are replicated.

Initial studies on self-affirmation effects in education were conducted by researchers who are self-affirmation theorists. It is important that other researchers also conduct such tests to ensure that the findings are not open to the accusation that they come from those who have vested interests in the intervention. They were also conducted in the US with ethnic minority students. And where experiments were replicated showing different results (e.g. Bratter, Rowley and Chukhray, 2016; de Jong et al., 2016; Hanselman et al., 2017; Protzko and Aronson 2016; Simmons 2011), they were often because there had been changes in the way the intervention was implemented in terms of timing, setting and cultural context. In Protzko and Aronson and Simmons’ studies, the research nature of the study was not concealed from the students. This

could have compromised the integrity of the intervention. It is, therefore, essential to re-affirm the earlier positive findings replicating the conditions in terms of timing and settings.

Our study replicated these conditions changing only the study participants (poor students in England instead of ethnic minority students in the US) to see if similar results can be generalised to other populations. Stereotype threats based on ethnicity or race could be more apparent because they are based on physical attributes, SES-based stereotypes could be less obvious to all observers. However, children from low income families also experience stereotype threats through early failure in school, which can in turn influence their aspiration, attitudes and behaviour. Empirical research in psychology shows that children living in poverty also self-stereotype, and can see themselves as ‘failures’ (Fell and Hewstone 2015), which can affect their performance at school. The theory is that the value-affirmation 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, and more confident and this can affect peers’ and teachers’ expectations. The results of this study confirm that value-affirming activities might help overcome stereotype threats of low-performing disadvantaged pupils in England. This corroborates the findings of earlier studies, which will give confidence, to those who wish to apply this intervention, in the efficacy of this approach in overcoming the detrimental effects of negative stereotype for disadvantaged pupils.

We can conclude that our study has successfully replicated the early studies in the US, and this is possible partly because of the good study design (e.g. randomised control trial) and large sample sizes in the earlier studies (Patil, Peng and Leek 2016; Shadish et al. 2008; Steiner, Wong and Anglin 2019). Promising approaches from the EEF trials conducted so far generally have strong designs and involve large samples, and should be replicated before they are adopted more widely. To encourage such replication work, funders and government should require that research to inform policy and practice be directly replicated, preferably by an independent research team, different to the one which conducted the initial research. The moment has come in education research to demand such replications at a much larger scale.

While the effects are smaller than earlier studies, possibly due to muted effects because of the use of test scores rather than continuous teacher assessments and the fact that the final writing exercise was not personalised, this intervention is still worth considering given that it does not cost anything, does no harm, and could help reduce the poverty attainment gap.

## References

- Binning, K. R., and Browman, A. S. (2020). Theoretical, ethical, and policy considerations for conducting social – psychological interventions to close educational achievement gaps. *Social Issues and Policy Review*. <https://doi.org/10.1111/sipr12070.12066>
- Borman, G. D., Grigg, J., and Hanselman, P. (2016). An effort to close achievement gaps at scale through self-affirmation. *Educational Evaluation and Policy Analysis*, 38(1), 21-42.
- Borman, G.D., Grigg, J., Rozek, C., Hanselman, P., and Dewey, N.A. (2018). Self-affirmation effects are produced by school context, student engagement with the intervention, and time: Lessons from a district-wide implementation. *Psychological Science*, 29, 1773-1784.)
- Borman, G.D., Choi, Y., and Hall, G.J. (2021). The academic impacts of a brief middle-school self-affirmation intervention help propel African American and Latino students through high school. *Journal of Educational Psychology*, 113(3), 605–620.

- Bowen, N. K., Wegmann, K. M., & Webber, K. C. (2013). Enhancing a brief writing intervention to combat stereotype threat among middle-school students. *Journal of Educational Psychology, 105*(2), 427–435.
- Brannon, T. N., Higginbotham, G. D., and Henderson, K. (2017). Class advantages and disadvantages are not so Black and White: intersectionality impacts rank and selves. *Current Opinion in Psychology, 18*, 117–122.
- Bratter, J. L., Rowley, K. J., and Chukhray, I. (2016). Does a self-affirmation intervention reduce stereotype threat in black and Hispanic high schools?. *Race and Social Problems, 8*(4), 340-356.
- Cohen, G. and Sherman, D. (2014). The Psychology of Change: Self-Affirmation and Social Psychological Intervention. *Annual Review of Psychology, 65*, 333–371.
- Cohen, G. L., Garcia, J., Apfel, N., and Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*(5791), 1307-1310.
- Cohen, G. L., 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*(17) 400–403.
- 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>
- de Jong, E. M., Jellesma, F. C., Koomen, H. M., and de Jong, P. F. (2016). A values-affirmation intervention does not benefit negatively stereotyped immigrant students in the Netherlands. *Frontiers in psychology, 7*, 691.
- Dee, T. S. (2015). Social identity and achievement gaps: Evidence from an affirmation intervention. *Journal of Research on Educational Effectiveness, 8*(2), 149-168.
- Demie, F. and Lewis, K. (2011). White working class achievement: an ethnographic study of barriers to learning in schools. *Educational Studies, 37*(3), 245-264.
- DfE (2015). *2010-2015 government policy: Education of disadvantaged children. Policy paper*. London: DfE.
- DfE (2019). *KS4 performance, 2019 revised*. Retrieved from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/863815/2019\\_KS4\\_revised\\_text.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/863815/2019_KS4_revised_text.pdf)
- 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.
- Earp, B. D., and Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in psychology, 6*, 621.
- Easterbrook, M. J., and Hadden, I. R. Tackling Educational Inequalities with Social Psychology: Identities, Contexts, and Interventions. *Social Issues and Policy Review*. 1-57. DOI: 10.1111/sipr.12070
- EEF (2018). *Annual report*. London: EEF.
- Fell, B., and Hewstone, M. (2015). *Psychological perspectives on poverty*. York: Joseph Rowntree Foundation.
- 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.



- 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).
- Gorard, S. (2012). Who is eligible for free school meals? Characterising free school meals as a measure of disadvantage in England. *British Educational Research Journal*, 38, 6, 1-15.
- Gorard, S. (2018). *Education policy: Evidence of equity and effectiveness*. Policy Press.
- Gorard, S. (2021) *How to make sense of statistics: Everything you need to know about using numbers in social science*, London: SAGE
- Gorard, S., See, B. H. and Morris, R. (2016). *The most effective approaches to teaching in primary schools*. Saarbrücken: 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.
- Gorard, S., Siddiqui, N. and See, BH (2021) Assessing the impact of Pupil Premium funding on primary school segregation and attainment, *Research Papers in Education*, <https://doi.org/10.1080/02671522.2021.1907775>
- Goyer, J. P., Garcia, J., Purdie-Vaughns, V., Binning, K. R., Cook, J. E., Reeves, S. L., Apfel, N., Taborsky-Barba, Sherman, D. K. and Cohen, G. L. (2017). Self-affirmation facilitates minority middle schoolers' progress along college trajectories. *Proceedings of the National Academy of Sciences*, 114(29), 7594–7599.
- Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., and Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of Personality and Social Psychology*, 111(5), 745–765.
- Hadden, I. R., Easterbrook, M. J., Nieuwenhuis, M., Fox, K. J., and Dolan, P. (2020). Self-affirmation reduces the socioeconomic attainment gap in schools in England. *British Journal of Educational Psychology*, 90(2), 517-536.
- Hanselman, P., Rozek, C. S., Grigg, J., and Borman, G. D. (2017). New evidence on self-affirmation effects and theorized sources of heterogeneity from large-scale replications. *Journal of educational psychology*, 109(3), 405.
- Hobbs, G., and Vignoles, A. (2010). Is children's free school meal 'eligibility' a good proxy for family income? *British Educational Research Journal*, 36, 673–690.
- Hofstede, G. J., Pedersen, P., and Hofstede, G. (2002). *Exploring Culture: Exercises, Stories and Synthetic Cultures*. Maine: Intercultural Press.
- Hunter, J. E. (2001). The desperate need for replications. *Journal of Consumer Research*, 28(1), 149-158.
- Johnston, J.M., and Pennypacker, H.S. (2009). *Strategies and tactics of behavioral research* (3rd ed.). Routledge.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley
- Makel, M. C. and Plucker, J. (2014). Facts are more important than novelty: Replication in the education sciences. *Education Researcher*, 20 (10), 1-13.
- McCoy, C. E. (2017). Understanding the intention-to-treat principle in randomized controlled trials. *Western Journal of Emergency Medicine*, 18(6), 1075.
- Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., and Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, 330(6008), 1234-1237.
- Montori, V. M., and Guyatt, G. H. (2001). Intention-to-treat principle. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, 165(10), 1339–1341.

- Morrison, K. (2020). Realising the promises of replication studies in education. *Educational Research and Evaluation*, 25(6-7), 412-441.
- Nicholl, J. J. (n.d.). Complier average causal effect analysis. Retrieved from [https://www.sheffield.ac.uk/polopoly\\_fs/1.4187111/file/JNicholls.pdf](https://www.sheffield.ac.uk/polopoly_fs/1.4187111/file/JNicholls.pdf)
- Nye, P. and Thomson, D. (2019, December 5). Who's left 2019, part one: The disadvantage gap is bigger than we thought. *FFT Education Datalab*. Retrieved from <https://ffteducationdatalab.org.uk/2019/12/whos-left-2019-part-one-the-disadvantage-gap-is-bigger-than-we-thought/>
- OECD (2019). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. Paris: OECD.
- 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, 188-204.
- Patil, P. Peng, R.D. and Leek, J. (2016). What should we expect when we replicate? A statistical view of replicability in psychological science. *Perspectives on Psychological Science*. 11(4), 539-544.
- Perezgonzalez, J. D. (2015). The Meaning of Significance in Data Testing. *Frontiers in Psychology*, 6, 1293, 1-4.
- 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.
- Protzko, J. and Aronson, J. (2016). Context moderates affirmation effects on the ethnic achievement gap. *Social Psychological and Personality Science*, 7(6), 500-507.
- 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.
- Rosenthal, R., & Jacobson, L. F. (1968). Teacher expectations for the disadvantaged. *Scientific American*, 218(4), 19-23.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2001). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Shadish, W. R., Clark, M. H., and Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American statistical association*, 103(484), 1334-1344.
- Sherman, D. K. and Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology*, 38, 183-242.
- Sherman, D. K., Hartson, K. A., 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 Personality and Social Psychology*, 104(4), 591-618.
- Simmons, C. M. (2011). *Reducing Stereotype Threat in Academically At-Risk African-Americans Students: A Self-affirmation intervention*. (Doctoral dissertation, UC Berkeley).
- Slavin, R. (2018, Feb 28). *What kinds of studies are likely to replicate* [Web log post]. Retrieved from <https://robertslavinsblog.wordpress.com/2018/02/28/what-kinds-of-studies-are-likely-to-replicate/>
- Smith, E. N., Rozek, C. S., Manke, K. J., Dweck, C. S., & Walton, G. M. (2021). Teacher-versus researcher-provided affirmation effects on students' task engagement and positive perceptions of teachers. *Journal of Social Issues*, (December 2020), 1-18.

- Spencer, B., and Castano, E. (2007). Social class is dead. Long live social class! Stereotype threat among low socioeconomic status individuals. *Social Justice Research*, 20(4), 418-432.
- Spencer, S. J., Logel, C., and Davies, P. G. (2016). Stereotype Threat. *Annual Review of Psychology*, 67, 415–437. <https://doi.org/10.1146/annurev-psych-073115-103235> [doi]
- Steiner, P.M., Wong, V.C., and Anglin, K. (2019). A causal replication framework for designing and assessing replication efforts. *Zeitschrift für Psychologie*, 227(4), 280-292.
- Steele, C.M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in experimental social psychology*, 21(2), 261-302.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American psychologist*, 52(6), 613.
- Strand, S. (2014). Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and ‘Getting it’ for the white working class. *Research Papers in Education*, 29(2), 131-171.
- Taylor, C. (2018). The reliability of free school meal eligibility as a measure of socio-economic disadvantage: Evidence from the Millennium Cohort Study in Wales. *British Journal of Educational Studies*, 66, 1, 29-51.
- The big classroom experiment: England has become one of the world’s biggest education laboratories (2018, March 31). *The Economist*. Retrieved from <https://www.economist.com/britain/2018/03/31/england-has-become-one-of-the-worlds-biggest-education-laboratories>
- Travers, J.C., Cook, B. G., Therrien, W.J., and Coyne, M.D. (2016). Replication research and special education. *Remedial and Special Education*, 37(4), 195–204.
- Walton, G. M., and Yeager, D. S. (2020). Seed and soil: Psychological affordances in contexts help to explain where wise interventions succeed or fail. *Current Directions in Psychological Science*, 29(3), 219–226. <https://doi.org/10.1177/0963721420904453>
- Woodcock, A., Hernandez, P. R., Estrada, M., and Schultz, P. (2012). The consequences of chronic stereotype threat: domain disidentification and abandonment. *Journal of personality and social psychology*, 103(4), 635.
- Wu, Z., Spreckelsen, T. F., & Cohen, G. L. (2021). A meta-analysis of the effect of values affirmation on academic achievement. *Journal of Social Issues*, (December 2020), 1–49.
- Yeager, D. S., and Walton, G. M. (2011). Social-psychological interventions in education: They’re not magic. *Review of Educational Research*, 81, 267–301.

## APPENDIX A1: 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 A2: 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*

### 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)