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

Beng Huat See[1]
Rebecca Morris[2]
Stephen Gorard[1]
Nadia Siddiqui[1]

Matthew J. Easterbrook[3]
Marlon Nieuwenhuis[3,4]
Kerry Fox[3,5]
Peter R. Harris[3]
Robin Banerjee[3]
[1] School of Education, Durham University, UK
[2] Centre for Education Studies, University of Warwick, UK
[3] School of Psychology, University of Sussex, UK
[4] Faculty of Behavioural, Management and Social Science, University of Twente, The Netherlands
[5] School of Applied Social Science, University of Brighton, UK

Abstract

Self-affirmation theory suggests that some potentially stigmatised groups, such as those from ethnic minority or poor families, face stereotype threats which 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 useful productiveness of any theory is the replicability of evidence concerning its predictions. This paper describes a randomised control trial of a self-affirmation intervention, replicating earlier studies, mostly conducted in the US with ethnic minority students. The present study, involving 5,116 Year 10 and 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 wasa small 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, and so helps reduce the poverty attainment gap. The findings are worth consideration because the intervention has no cost.

Introduction

Replications are crucial in validating the trustworthiness of scientific findings, but they are rare in education, and while there has been an increase in the number of randomized 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 are direct or even conceptual replications. In the US, Slavin noted that half of all programmes in the What Works Clearinghouse are single evaluations (Slavin 2018). 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 & 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 influence pupil outcomes.

The present 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, replicating the conditions in terms of implementation and delivery as described in the original studies. 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 & Trafimow 2015; Makel & Plucker, 2015).

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 performance in 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. 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).

Evidence from studies conducted largely in the US has suggested 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 & Terry 2006; Sherman and Cohen, 2006; Steele, 1988). Among the evaluations funded by the EEF was a trial of self-affirmation writing exercises, aimed at improving the academic attainment of disadvantaged students at Key Stage 4, 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 are produced with English students who are in receipt of free school meals (FSM). 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. (2003), on the academic outcomes of low-income pupils in England

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 defense mechanism, known as 'disidentification', in order to protect the self-concept from being devalued by the negative stereotype (Spencer, Logel & 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 & 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 apart from the initial training of teachers and the costs of printing any exercise booklets and 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 & Browman 2020; Easterbrook & Hadden 2020; Walton & Yeager 2020).

Most of the studies conducted on this 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 & Sherman, 2014; Sherman et al., 2013). Cohen et al. (2009), for example, found that although there were no overall gains in grade point averages across four core academic subjects in both treatment and control groups, African American students in the treatment group improved their Grade Point Average (GPA) score by 0.24 points, and the low-achieving African-American students by 0.41. The intervention also appeared to reduce the likelihood of 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 & Hanselman (2015) also reported a positive impact on minority pupils' standardised maths test scores, while Mikaye et al. (2010) showed the self-affirmation can help close the gender attainment gap.

However, other studies have shown no effects on either academic or other outcomes (Bratter, Rowley & Chukhray, 2016; de Jong et al., 2016; Hanselman et al., 2017; Protzko & Aronson 2016). There are subtle differences between these less promising studies and the ones by Cohen et al. (2006, 2009). These provide hints about the delicate nature of delivering the intervention. Such tweaks in 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 Protzko & Aronson's (2016) study, the writing instructions were handed to students by researchers rather than teachers. The knowledge that it was a research exercise *may* have nullified its effect. 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 & Sherman, 2014). One explanation is 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 & 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 theywill 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 recursive 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 likely to be along the lines of socio-economic status. The underachievement of white working class boys in England, for example, is well documented (e.g. Demie & Lewis, 2011; Strand, 2012). Hadden et al.'s (2019) study, which was 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 whose self-esteem is threatened by their 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 original studies in its implementation and delivery.

The intervention

Throughout the 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 as previous research has shown that knowledge of the purpose of the intervention can interfere with its efficacy (Yeager & 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 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 (a test taken at the end of secondary education)

As in the original study from the US, the intervention comprised three writing activities, each lasting 10 to 15 minutes, in which 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 and to reflect on their thoughts, beliefs and views, and that their work would not be marked or read by their teachers. The writing exercises would be stored away and there will be no feedback given. The aim was to get pupils to write freely. 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. Teachers were given strict instructions to use the prescribed answers to pupils' queries about the purpose of the exercises.

These 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 are treatment or control pupils because all pupils were involved in a writing exercise that differed only in terms of content seen only by themselves.

As a replication of the Sherman et al. study and 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 again before stressful events. In this study these were the mock GCSEs and the actual GCSEs exams later in the year.

Although not part of the intervention, a questionnaire survey was administered to measure the possible impact on pupils' self-efficacy (Appendix C). To ensure that pupils did not associate the survey with the intervention and also that the survey did not interfere with the intervention, the survey was delivered after the final writing task.

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, and no schools were visited for both observation of the delivery of the exercises and the administration of the survey questionnaire. 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. 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 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 committee 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 test 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 original studies by 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.

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 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 1 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. 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 1: 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 individually randomized cases in the interventions 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 & 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 & Gorard, 2016). 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 deteactabel 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.

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 Ever FSM6 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 received 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. The differences are expressed as simple effect sizes (difference between means divided by their overall standard deviation). For comparability between phases, the test scores were converted to standardised z-scores. 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 and liable to mislead (for further explanation, see Gorard, 2021; Colquoun, 2014, 2016; Perezgonzalez, 2015; Pharoah et al., 2017).

There should no issue of clustering as randomization 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 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 & Lipsey 2011; Foster & Fang, 2004; Little & Rubin, 1987; Puma et al., 2009; Shadish, Cook & Campbell, 2001) we presented differences in pre-test scores (KS2 maths and reading) between 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 estimated 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

Since not all pupils completed the three writing activities, 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 completed. This would be zero for all cases in the control group. Information about dosage was collected by the project delivery team 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. Compliance 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 & 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 2 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 2: 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 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 3 shows that both groups made less than average progress between KS2 and KS4, 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 effects 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 any 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 3: Comparison of pre, post and standardised gain scores using KS2 maths and KS2 reading combined as pre-test and Attainment 8 as post-test (EverFSM6 pupils only)

	Pre-score mean	**SD**	**ES**	**Post-score mean**	**SD**	**ES**	**Gain score**	**SD**	**ES**
	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

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 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 (-0.01for Year 11 and 0.00 for Year 10). This is consistent with the theory that self-affirmation only works with pupils experiencing stereotype threat.

Table 4: 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 wheter 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 11cohort did (Table 5).

Table 5: 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 6). The relationship is stronger for the Year 10 pupils (+0.36) than for the Year 11 pupils (+0.26). This may suggest that the lasting effect of the intervention is stronger the more exercises pupils complete.

Table 6: Correlation between gain scores and number of exercises completed (EverFSM6 pupils only)

	Gain scores using KS2 maths & reading combined		GCSE Attainment 8 score	
Number of exercises completed	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. 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 3 (0.87 for Year 11 and 0.85 for Year 10). The result shows a small positive effect size (Table 7), exactly the same as the impact evaluation result in Table 3.

Table 7: 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 Year 10	 526 549	 +0.01 -0.005	 114 125	 -0.60 -0.56	 640 674	 -0.10 -0.10	
Control Year 11 Year 10	 *583* *569*	 *-0.04* *-0.03*	 *128* *129*	 *-0.60* *-0.56*	 711 698	 -0.14 -0.13	 +0.05 +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 will not be graded and no one will be looking at what they have written, and their writing will 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 are so conditioned to focus on exams and meeting exam criteria that to do something different was refreshing and interesting. In another school, following the first writing task, one teacher said that being involved in the "Writing about Values" (WaV) project has made the faculty consider whether to teach more free and creative writing and to embed this within the Key Stage 4 curriculum. She felt that there could be opportunities to include WaV-style tasks within schemes of work, benefitting staff and students by making it a regular and expected part of English lessons. 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 written 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 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).

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 is 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 where the negative stereotype effect on their academic performance is psychological or social (Binning & Browman 2020; Easterbrook & 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 became more familiar with it, and more aware of the benign intentions, it may become 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 that the intervention works with disadvantaged pupils in England just as well as 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 original 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 & 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.

Discussion

Despite the increase in experimental studies in education in the past two decades, few studies have been replicated so far. Make & Plucker (2014) noted that only 13% of around 16,000 studies in top 100 education journals were replications. Of these, 28.5% 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 & Plucker (2014) 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.

Previous studies on self-affirmation were largely conducted by the same researchers who may have a vested interest in the intervention as they are strong proponents of the theory of value-affirmation. These were also conducted in the US with ethnic minority students. And where experiments were replicated by different authors, similar results were not produced (e.g. Bratter, Rowley & Chukhray, 2016; de Jong et al., 2016; Hanselman et al., 2017; Protzko & Aronson 2016; Simmons 2011). This is mainly because there had been changes in the way the intervention was implemented in terms of timing, setting and cultural context. In Protzko & 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. The results confirm that value-affirming activities can help overcome stereotype threats of low-performing pupils disadvantaged pupils in England. This verifies and 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 original studies, and this is possible only because of the good study design (e.g. randomised control trial) and large sample sizes in the original studies (Patil, Peng & Leek 2016; Shadish et al. 2008; Steiner, Wong & 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 replication work, funders and government should require that research that informs policy and practice be directly replicated, preferably by an independent research team, different to the ones who conducted the original research. The moment has come in education research to demand such replications.

References

Binning, K. R., & 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., & Hanselman, P. (2016). An effort to close achievement gaps at scale through self-affirmation. *Educational Evaluation and Policy Analysis*, *38*(1), 21-42.

Bratter, J. L., Rowley, K. J., & 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. & 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., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. S*cience*, *313*(5791), 1307-1310.

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science*, *324*(17) 400–403.

Colquoun, D. (2014). A*n 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., & 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. & 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.

Dong, N. & 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., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in psychology*, *6*, 621.

Easterbrook, M. J., & 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.

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

Garcia, J. & 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. & Gorard, J. (2016). What to do instead of significance testing? Calculating the number of counterfactual cases needed to disturb a finding, *International Journal of Social Research Methodology*, 19, 4, 481–490.

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. & Morris, R. (2016). *The most effective approaches to teaching in primary schools*. Saarbrucken: Lambert Academic Publishing.

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

Hadden, I. R., Easterbrook, M. J., Nieuwenhuis, M., Fox, K. J., & 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., & 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., & Vignoles, A. (2010). Is children's free school meal 'eligibility' a good proxy for family

Hofstede, G. J., Pedersen, P., and Hofstede, G. (2002). *Exploring Culture: Exercises, Stories and Synthetic Cultures*. Maine: Intercultural Press.

income? *British Educational Research Journal*, 36, 673–690.

Hunter, J. E. (2001). The desperate need for replications. *Journal of Consumer Research*, *28*(1), 149-158.

Johnston, J.M., & Pennypacker, H.S. (2009). *Strategies and tactics of behavioral research* (3rd ed.). Routledge.

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

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*(6), 537–542.

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, *330*(6008), 1234-1237.

Nicholl, J. J. (n.d.). Complier average causal effect analysis. Retrieved from https://www.sheffield.ac.uk/polopoly_fs/1.418711!/file/JNicholls.pdf

OECD (2019). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. Paris: OECD.

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

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. & Kar, S. (2017). *P-values and confidence intervals: Not fit for purpose?* bioRxiv preprint first posted online 24 August 2017.

Protzko, J. & 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. & 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.

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

Shadish, W. R., Clark, M. H., & 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. & 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 & 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/

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype Threat. *Annucal Review of Psychology*, *67*, 415–437. https://doi.org/10.1146/annurev-psych-073115-103235 [doi]

Steiner, P.M., Wong, V.C., & 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 Wale. *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., & Coyne, M.D. (2016). Replication research and special education. *Remedial and Special Education*, *37*(4), 195–204.

Walton, G. M., & 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., & Schultz, P. (2012). The consequences of chronic stereotype threat: domain disidentification and abandonment. *Journal of personality and social psychology*, *103*(4), 635.

Yeager, D. S., & 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

APPENDIX B: TEACHER INFORMATION SHEET

Writing about Values exercise – Instructions for teachers

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

What to do:

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

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

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

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

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

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

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

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

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

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

Name:
Tutor group:

Why is this survey being done?
This survey is being conducted by researchers from the University of Sussex who are investigating why it is that some people like school and other people do not. To help the researchers do this, the survey asks about how you feel about your school and your school work. There are no right or wrong answers, and all your answers will be completely confidential and will only be seen by the researchers – your teachers will not see your answers - so please answer honestly. The survey takes about 10 minutes to complete.

How to complete the survey
Please use a **black or blue pen or a pencil** to complete the survey. Please don't use any other colours, otherwise our computer can't detect your answers!

Please read each question carefully and **mark the box with an 'x'** to show your answer.

Example:	Strongly disagree			Neither agree nor disagree			Strongly agree
	1	2	3	4	5	6	7
I like my school					☒		

If you make **a mistake,** please cross out the wrong answer and mark another box with an 'x' and circle it to show your right answer.

Example when making a mistake:

	1	2	3	4	5	6	7
I like my school							

Please do **not** draw on or colour the black boxes in the corners of the survey or the code box on the top left corner of each page. We use a computer to automatically read your survey answers, which doesn't work if these boxes are tampered with!

Prize draw
The researchers are also very interested in asking you some similar questions after you have completed your GCSE exams at the end of Year 11. As a thank you to those who complete that survey, everyone who fills it out will be entered into a prize draw to win a top prize of £50 and two prizes of £25. If you would like to sign up to complete the survey and be entered into the prize draw after you have finished your GCSEs, then please write your email address below so that the researchers can contact you with the survey. Your email address will not be passed on or used for anything other than to contact you about the survey, and will be deleted from the records once it has been used to contact you. Signing up now does not commit you to anything.

My email address is: ……

Thank you for taking the time to complete this survey. You can now start with the survey on the next page.

11951

Welcome!
We are interested in your thoughts and opinions about yourself and school.
There are no right or wrong answers. We are only interested in what YOU think.

Please answer the following questions about school and school work.

A. How do you feel about school work?	Strongly disagree			Neither agree nor disagree			Strongly agree
	1	2	3	4	5	6	7
I work as hard as possible	☐	☐	☐	☐	☐	☐	☐
I keep working even if the material is difficult	☐	☐	☐	☐	☐	☐	☐
I try my best to learn what is being taught	☐	☐	☐	☐	☐	☐	☐
I apply my best effort	☐	☐	☐	☐	☐	☐	☐
I'm certain I can understand the most difficult material presented in texts	☐	☐	☐	☐	☐	☐	☐
I'm confident I can understand even the hardest things taught by the teacher	☐	☐	☐	☐	☐	☐	☐
I'm confident I can do an excellent job on assignments and tests	☐	☐	☐	☐	☐	☐	☐
I'm certain I can master the skills being taught	☐	☐	☐	☐	☐	☐	☐
I learn things quickly in most school subjects	☐	☐	☐	☐	☐	☐	☐
I'm good at most school subjects	☐	☐	☐	☐	☐	☐	☐
I do well in tests in most school subjects	☐	☐	☐	☐	☐	☐	☐
I have an uneasy, upset feeling when I take an exam	☐	☐	☐	☐	☐	☐	☐
When I struggle with my schoolwork, I usually step back and think about the bigger picture	☐	☐	☐	☐	☐	☐	☐

B. How do you feel about school?	Strongly disagree			Neither agree nor disagree			Strongly agree
	1	2	3	4	5	6	7
In general, I really feel like I belong at school	☐	☐	☐	☐	☐	☐	☐
In general, I feel people in school accept me	☐	☐	☐	☐	☐	☐	☐
In general, people in school like me	☐	☐	☐	☐	☐	☐	☐
Today, I feel anxious about school	☐	☐	☐	☐	☐	☐	☐

C. The following questions are about yourself	Strongly disagree			Neither agree nor disagree			Strongly agree
	1	2	3	4	5	6	7
I feel better about myself when I know I'm doing well at school	☐	☐	☐	☐	☐	☐	☐
My self-esteem is influenced by my performance at school	☐	☐	☐	☐	☐	☐	☐
Doing well in school gives me a sense of self-respect	☐	☐	☐	☐	☐	☐	☐
I feel like I have the support I need from the people close to me	☐	☐	☐	☐	☐	☐	☐
I have high self-esteem	☐	☐	☐	☐	☐	☐	☐

11951

Sometimes when we face problems in our daily lives we can find ourselves thinking about ourselves. We are interested in how often you think about yourself when things start to bother you.

D. When I have problems I find myself …	Strongly disagree			Neither agree nor disagree			Strongly agree
	1	2	3	4	5	6	7
... thinking about my values (the things I think are important in life)	☐	☐	☐	☐	☐	☐	☐
... thinking about the things I'm good at	☐	☐	☐	☐	☐	☐	☐
... thinking about the people who are important to me	☐	☐	☐	☐	☐	☐	☐

E. The next questions are about you and your social background

By *social background*, we mean people who are from the same social class or community as you, who live in the same types of places as you, and who do similar things as you, and whose family has similar amounts of money and do similar sorts of things as yours.

	Strongly disagree			Neither agree nor disagree			Strongly agree
	1	2	3	4	5	6	7
I feel strong ties with people who have my social background	☐	☐	☐	☐	☐	☐	☐
Working hard at school fits with my social background	☐	☐	☐	☐	☐	☐	☐
My background is compatible with someone who does well at school	☐	☐	☐	☐	☐	☐	☐
People with my social background usually get good grades at school	☐	☐	☐	☐	☐	☐	☐
People expect me to behave in a certain way at school because of the amount of money my family has	☐	☐	☐	☐	☐	☐	☐
People make assumptions about me at school based on my family's social background	☐	☐	☐	☐	☐	☐	☐
My school treats everyone the same regardless of their background	☐	☐	☐	☐	☐	☐	☐

F. Now please answer some general background questions.

1. Has your father (or person you consider to be your father) studied at university?
   ☐ Yes ☐ No ☐ I don't know ☐ I don't have a father or a person I consider to be my father
2. Has your mother (or person you consider to be your mother) studied at university?
   ☐ Yes ☐ No ☐ I don't know ☐ I don't have a mother or a person I consider to be my mother
3. Do you plan on going to university?
   ☐ Yes ☐ No ☐ I don't know

This is the end of the questionnaire. Thank you for your participation!

You can write any comments you have about this questionnaire in the box below.
Please make sure the questionnaire is returned to your teacher.