Addendum – added January 2020

Executive summary

This addendum should be read in conjunction with the main report on this project: "Writing About Values Evaluation Report and Executive Summary, November 2018". This report presents the results of Writing about Values, a self-affirmation intervention, based on the KS4 results for 5,188 Y10 pupils who ended KS4 in May/June 2018 (a year after the end of the intervention). This was to test the delayed and sustained effect of the self-affirmation intervention. The 2018 report presented the results for 5,619 Year 11 (age 15–16) cohort who ended KS4 in May/June 2017 immediately after the end of the intervention (https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/writing-about-values/).

The project

The intervention in brief

Writing about Values is a self-affirmation intervention based on the theory that certain disadvantaged groups have a perception that others perceive them in a negative way. These negative perceptions present a stereotype threat that encourages individuals to engage in defensive behaviours to protect their self-worth. Such behaviours are believed to hamper their learning at school. The self-affirmation intervention involves these individuals engaging in activities designed to promote self-affirmation, such as writing about values that are important to them. It is believed that such activities can help give them a sense of value and thus alleviate negative feelings associated with the negative perceptions (stereotype threat).

Impacts measured by the addendum

In this addendum report, we present the results of the impact of the intervention on Year 10 pupils' academic attainment and self-efficacy. Academic attainment was measured using KS4 Attainment 8 results, and self-efficacy was measured using a bespoke survey questionnaire adopting items from the subscales of the Motivated Strategies for Learning Questionnaire (Pintrich et al., 1993). The KS4 exams were taken in May/June 2018, while the self-efficacy survey was taken in 2017.

Summary of previous results

- For the Y11 cohort there was a very small but positive impact from the intervention for EverFSM6 pupils, but the size of the effect is very small, at +0.05. There is no evidence that the intervention benefitted all children. This is consistent with the theoretical understanding of the intervention, which is to reduce the adverse effects of being associated with membership of a disadvantaged group.
- Pupils who completed more writing exercises made more progress. This is the same for all pupils and for EverFSM6 pupils.

Summary of new results

Key conclusions	
1.	Among disadvantaged pupils (defined as EverFSM6), those who received the self-affirmation intervention made slightly more progress between the end of primary school and GCSEs than the comparison pupils, but the size of the impact was very small, at +0.05 (1 month progress) for the Y11 cohort and +0.04 (0 months progress) for the Y10 cohort.
2.	Pupils who completed more writing exercises made more progress – a reasonable correlation of +0.26. This is stronger than for the Y11 (0.09), perhaps suggesting that there is a sustained effect from the intervention. This may mean that the intervention can lead to better outcomes if implemented more thoroughly, but might also be because the kind of pupils who completed more exercises would make more progress anyway.
3.	The intervention takes 10 minutes, delivered 3 times a year, and costs nothing for schools to implement.
4.	In this trial neither pupils nor teachers were aware of the nature of the intervention or why the tasks were undertaken because there is some evidence that knowledge about the purpose of the intervention can reduce its effectiveness. Therefore, schools considering this approach should bear in mind the practical difficulty in replicating these conditions.

The impact of the intervention for the Year 10 cohort (+0.04) is similar to that for Year 11, considering EverFSM6 students only, suggesting that the positive effect of the intervention is maintained for at least a year even after the intervention has ended.

As with the Y11 cohort, results for the Y10 cohort show a correlation between the number of exercises completed and attainment scores for all pupils and EverFSM6 pupils regardless of whether gain scores or post-test scores were used. In fact, the correlation is considerably stronger for the Y10 cohort than for the Y11. This suggests that the number of exercises completed is important in maintaining the effects.

Table 1: Summary of impact on primary outcome

Outcome/ Group	Effect size	Estimated months' progress	EEF security rating	No. of pupils	EEF cost rating
Main Report (Y11 cohort) KS4 Attainment 8 EverFSM6	+0.05	1	🔒🔒🔒🔒🔓	1,351	**£**£££
Addendum Report (Y10 cohort) KS4 Attainment 8 EverFSM6	+0.04	0	N/A	1,372	**£**£££

Introduction

Intervention

The Writing about Values project aimed to improve the KS4 attainment outcome of disadvantaged pupils. It is based on research about 'self-affirmation', though this term was not used to describe the project publicly as some research suggests that knowledge of the intended outcomes could compromise its efficacy. The theory states that disadvantaged pupils are aware of stereotypical beliefs about how pupils like themselves perform academically, and that this awareness can itself detract from their performance. Undertaking self-affirming activities, such as writing about values that are important to them, can help to protect pupils' self-worth and free up cognitive resources to engage more effectively with their learning.

The intervention involved three writing activities delivered once at the beginning of the year, once before mock GCSEs and once before the actual GCSE exams commence. Students randomised to intervention group wrote reflective essays about core values, such as relationships with friends and family, sport or music. Control students were also given the writing tasks but were asked about values that are not important to them but that might be important to other people. These are open tasks and allow pupils to respond however they choose. These writing tasks were delivered three times in the year during regular English language lessons. All the teachers were trained to use the materials and administer the writing exercises, but not told what the intended outcome of the intervention was in order to avoid contamination.

Further details of the intervention and trial design can be found in the **2018 report**. The 2018 report presents the impact of the intervention on KS4 results for 5,619 Year 11 (age 15–16) pupils who took their KS4 exams in May/June 2017. This addendum report presents the results for 5,188 Year 10 cohort of pupils who took their KS4 exams a year after the intervention to test the sustained effect of the intervention.

The **main report** ("Writing about Values") summarised the findings of this randomised controlled trial on the KS4 attainment outcome for the Y11 cohort.

Evaluation objectives

The objective of the trial was to evaluate the impact of the self-affirmation intervention on the academic performance of Year 10 (age 14–16) and Year 11 (age 15–16) pupils in England. The 2018 report presents the attainment outcomes for the Year 11 cohort. This report presents the attainment and self-efficacy outcomes for the Year 10 cohort. Since the self-efficacy survey had to be taken after the final exams (to avoid compromising the effects of the intervention), the Year 11 did not take the post-survey after KS4 exams when most would have left school. Only the Year 10 cohort completed both the pre- and post-survey. For this reason, only the Year 10 self-efficacy outcomes will be analysed and published in this report.

The research questions are:

1. What impact does the self-affirmation intervention have on the academic attainment of disadvantaged pupils, defined as those who have been eligible for FSM at some point in the last six years (EverFSM6), using the individual pupils' Attainment 8 measure at KS4 after one year of treatment (for initial Year 11 group)?

2. Is there a sustained impact of the self-affirmation intervention on the Attainment 8 measure at KS4 for EverFSM6 pupils after two years (one year after the end of the intervention)? (This analysis is based on the initial Year 10 group.)

3. What impact does the self-affirmation intervention have on all pupils (EverFSM6 and non-EverFSM6) using the individual pupils' Attainment 8 measure at KS4 after one year of treatment (for initial Year 11)?

4. Is there a sustained impact of the self-affirmation intervention on the Attainment 8 measure at KS4 for all pupils (EverFSM6 and non-EverFSM6) a year after the end of the intervention (for initial Year 10)?

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

Research questions (1) and (3) are answered in the 2018 report. Research questions (2), (4) and (5) are answered in this report. These reflect those in the **protocol** and **statistical analysis plan** published alongside this report and the 2018 report.

Project team

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

Delivery team

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

Evaluation team

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

Methods

Trial design

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

Evaluation of impact for Year 11 was undertaken at the end of the first year following release of the KS4 results, whereas impact evaluation for the Year 10 cohort was completed in the summer of 2019[1] (a year after the release of the KS4 results). The outcomes of the Y10 cohort was to test the sleeper or sustained effect of the intervention.

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

Participant selection and sample size

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

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

Once schools confirmed participation, a Memorandum of Understanding (MOU) outlining the intervention and the trial was sent to school heads who then signed an agreement to comply with the requirements of the trial, including releasing teachers for training, delivering of the writing exercises and administering the self-efficacy survey (see Appendix F of the 2018 report). To minimise dropout after randomisation, all participating schools were offered an incentive payment of £1,000 for completing the trial, delivering the three writing exercises and administering the non-attainment surveys.

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

[1] Analysis for the Y10 cohort had been delayed due to the introduction of the new GDPR, the handing over of access to the data from DfE to ONS and technical problems with using ONS SRS to access and analyse data.

Addendum outcome measures

Primary outcomes

The primary outcome was Attainment 8 at KS4 for EverFSM6 pupils after two years (one year after the end of the intervention) for pupils who received the intervention when they were in Year 10. The EEF is interested in the outcomes of disadvantaged pupils, and it was agreed that the definition of 'disadvantaged' would be 'pupils ever eligible for FSM in the last six years' as this is the definition underlying the Pupil Premium eligibility. Therefore, the primary outcome is the KS4 results of EverFSM6 pupils who took the KS4 exams in May/June 2018. This data was acquired from the DfE National Pupil Database.

Secondary outcomes

However, based on prior work conducted by the developers and others (for example, Bowen et al., 2012; Cohen and Sherman, 2014; Stephens et al., 2012), there is some evidence that the intervention may be more relevant to historically low performing groups, such as ethnic minority groups in the U.S. and those from lower income groups. Stephens et al., for example, found that the intervention closed the achievement gap between students from lower socio-economic backgrounds and their more advantaged peers by 50%. Bowen et al.'s study suggests that the intervention can prevent a decline in performance among low-income pupils. In line with this evidence, the intervention was expected to have a positive effect on the traditionally negatively stereotyped group, such as the FSM pupils in U.K. It was therefore decided to evaluate also the impact of the intervention on pupils who were currently eligible for free school meals.

The secondary outcomes were therefore:

- Attainment 8 at KS4 for all pupils (both EverFSM6 and non-Ever FSM6) who received the intervention when they were in Year 10
- Attainment 8 KS4 scores for FSM pupils (based on current FSM status) for pupils who received the intervention when they were in Year 10.

In the 2018 report, only the attainment outcomes for the Year 11 cohort were analysed.

Non-attainment outcome

The non-attainment outcome was pupils' perceived self-efficacy, one of the four subscales of the Motivated Strategies for Learning Questionnaire (Pintrich et al., 1993). The four subscales are effort and perseverance, perceived self-efficacy, control expectation and academic self-concept. The instrument has been tested for validity and reliability by Pintrich et al. These non-attainment surveys were designed, collected and electronically marked by the delivery team (who are blinded to randomisation), but analysed by the independent evaluators.

This report presents the attainment and self-efficacy results for the original Year 10 pupils. Further details of the design can be found in the 2018 report.

Analysis

A copy of the Statistical Analysis Plan is available in the 2018 report, Appendix H.

The report analysed the Year 10 and Year 11 cohorts separately for two reasons. First, we wanted to see if there was a lasting effect of the intervention. Therefore, results for Year 11 are analysed immediately after the intervention and that for Year 10 are analysed a year after when their KS4 results are available. Second, changes in regulations regarding access to NPD data and additional security and privacy measures in the year between the Y10 and Y11 results made it necessary to analyse the

Year 10 data separately as Y11 data was outside the ONS SRS system. This is different to the Statistical Analysis Plan, which combined the sample for both Year 10 and 11 pupils.

Primary analysis

Primary intention-to-treat (ITT) analysis

The primary analysis is intention-to-treat. Those pupils randomised to receive the intervention were analysed regardless of whether they received the intervention or not. This means that a school that was included in the randomisation but then decided not to participate in the trial was also included in the analysis. The primary impact evaluation is based on the difference between groups in terms of the gain scores between KS2 results for maths and reading and KS4 Attainment 8 outcomes for those who were eligible for free school meals in the previous six years (EverFSM6). The differences are expressed as effect sizes (Hedge's g) and converted to progress in months. Significance tests and confidence intervals are not reported here but the latter can be calculated given that the means, standard deviations, and effect sizes are reported. We have not calculated any p-values as they are not relevant (for further explanation, please refer to Gorard, 2016; Cohen, 1994; Colquoun, 2014, 2016; Trafimow and Rice, 2009; Perezgonzalez, 2015; Pharoah et al., 2017).

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

Imbalance at baseline

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

Missing data

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

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

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

Fidelity analysis

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

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

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

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

Table 2: Estimation of Complier Average Causal Effect

Participants	Compliers		Non-compliers		All	
	N who completed first writing task (proportion of those who complied)	Mean	N who did not complete first writing task (proportion who did not comply)	Mean	Total N	Mean
Treatment	A	B	C	D	E	F
Control	G	x	H	I	J	K

Secondary outcome analyses

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

- all pupils (that is, both EverFSM6 and non-EverFSM6) after two years (for Year 10 cohort)
- FSM pupils (based on current FSM status) after two years (for Year 10 cohort)
- Self-efficacy for the Year 10 cohort only [2]

Additional analyses

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

An additional regression analysis was performed as a test of fidelity using KS2 maths and reading results and the number of exercises completed or dosage (with zero for control pupils) as the predictors and KS4 Attainment 8 scores as the dependent variable. This additional analysis was specified in the Statistical Analysis Plan (see Appendix H in the 2018 report).

Effect size calculation

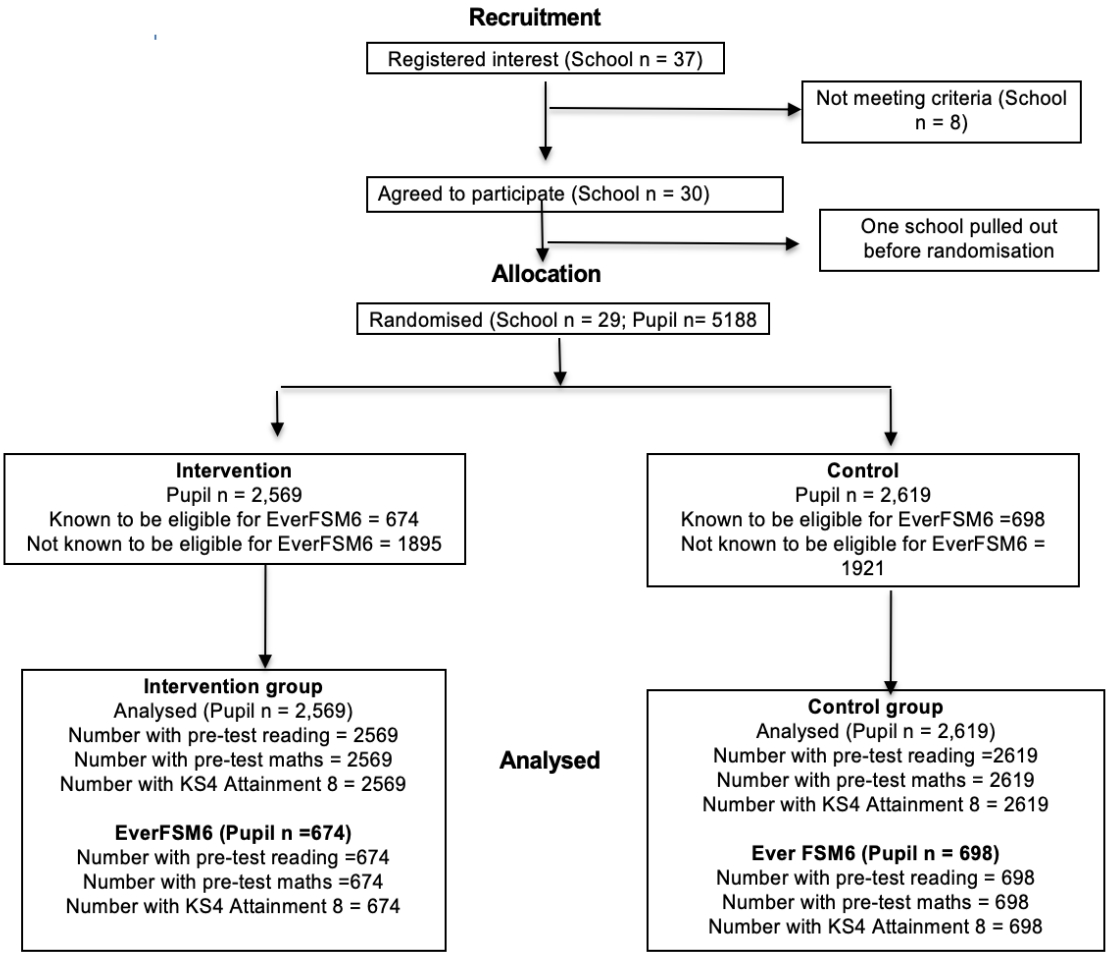
'Effect' sizes are calculated as the difference between mean post-test (and gain scores) for each variable, expressed as Hedges' *g*. We do not report 'confidence intervals' but this can be easily computed if any reader wishes to do so as the number of cases for each group, and the effect size for each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

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

[2] Since the self-efficacy survey was to be administered after the final exams, post survey data was not available for the Year 11 as many had left school immediately after KS4 exams.

Impact evaluation

Figure 1



Attrition

There were a total 5,188 pupils in the Year 10 cohort (Figure 1). All had pre-test data.

Table 3: Minimum detectable effect size at different stages (Year 10 cohort)

Stage	N [schools/pupils] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Power	Alpha	Minimum detectable effect size[3] (MDES)
Randomisation (All)	29 schools 5,188 pupils (2,569; 2,619)	0.67	0.00	80%	0.05	0.06
Randomisation (EverFSM6)	29 schools 1,372 pupils (674; 698)	0.60	0.00	80%	0.05	0.12
Analysis (EverFSM6)	29 schools 1,372 pupils (674; 698)	0.60	0.00	80%	0.05	0.12

Pupil and school characteristics

Table 4 shows that the two groups are balanced at pre-test. There is no difference between groups for both reading (ES = +0.01) and maths (ES = 0). The EverFSM6 pupils in the intervention, however, are slightly behind at pre-test for maths (ES = -0.04). For this reason, gain scores are also estimated, in addition to post-test scores.

Table 4: Comparison of pupils' baseline characteristics (Year 10 cohort)

Characteristics of pupils at randomisation	Intervention		Control	
	n/N (missing)	Percentage	n/N (missing	Percentage
Proportion of boys	1225/2521 (0)	49	1296/2521 (0)	51
Proportion of pupils eligible for EverFSM 6	674/1372 (0)	49	698/1372 (0)	51
Proportion of current FSM pupils	307/635 (22)	48	328/635 (25)	52
Proportion of pupils with SEN	401/806 (22)	50	405/806 (25)	50
Proportion of pupils whose first language is not English	229/466 (14)	49	237/466 (16)	51

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

	Raw means				
	Intervention		Control		
All pupils	n/N (missing)	Mean (sd)	n/N (missing	Mean (sd)	Effect size
KS2 Maths (point scores)	2569/5188 (0)	4.68 (0.78)	2619/5188 (0)	4.68 (0.81)	0.00
KS2 Reading (point scores)	2569/5188 (0)	4.70 (0.77)	2619/5188 (0)	4.69 (0.79)	+0.01
KS2 Reading and Maths combined	2569/5188 (0)	9.38 (1.42)	2619/5188 (0)	9.37 (1.47)	+0.01
EverFSM6 pupils					
KS2 Maths (point scores)	674/1372 (0)	4.40 (0.76)	698/1372/ (0)	4.43 (0.80)	-0.04
KS2 Reading (point scores)	674/1372 (0)	4.44 (0.83)	698/1372/ (0)	4.44 (0.83)	0.00
KS2 Reading and Maths combined	674/1372 (0)	8.84 (1.45)	698/1372/ (0)	8.85 (1.47)	-0.01
FSM pupils					
KS2 Maths (point scores)	307/635 (0)	4.32 (0.74)	328/635 (0)	4.32 (0.81)	0.00
KS2 Reading (point scores)	307/635 (0)	4.39 (0.85)	328/635 (0)	4.33 (0.88)	+0.07
KS2 Reading and Maths combined	307/635 (0)	8.71 (1.45)	328/635 (0)	8.65 (1.52)	+0.04

The primary analysis was conducted for those known to be eligible for EverFSM6. A total of 1,372 pupils are known to be eligible for EverFSM6. Of these, 674 (49%) are in the treatment group and 698 (51%) in the control group. There were no pupils missing KS4 results.

Analysis

Primary outcome analysis

Comparison of gain scores (EverFSM6 pupils only)

Analysis for the headline findings is performed for the 1,372 EverFSM6 pupils who have both pre-test scores for reading and maths and post-test scores. For comparability, all scores were converted to Z-scores before analysis.

Table 5 shows that the groups were not exactly balanced at the baseline, with the treatment group slightly behind the control. Both groups made less progress between KS2 and KS4 than their non-FSM peers, but the treatment group made slightly more progress than the control. This suggests that the intervention may have a small influence in improving the performance of the EverFSM6 pupils.

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)

	Pre-score mean	SD	ES	Post-score mean	SD	ES	Gain score	SD	ES
Treatment (n = 674)	-0.37	1.01		-0.47	0.91		-0.10	0.85	
Control (n = 698)	-0.36	1.02		-0.49	0.88		-0.13	0.84	
Overall (n = 1,372)	-0.37	1.01	-0.01	-0.48	0.89	+0.02	-0.11	0.85	+0.04

The number of counterfactual cases that would be needed to eliminate the headline finding of +0.04 is 27 (0.04 multiplied by 674). This means it would take approximately 27 missing cases with counterfactual scores (see methods) in the opposite direction for the findings to change. Since there are no cases with pre-tests missing post-test scores, this number means that the finding is reasonably secure, and cannot be due to attrition. These results are very similar to those for the Y11 cohort presented in the **2018 report**. Effect size is +0.04 for Y10 and +0.05 for Y11 cohort.

Secondary outcome analysis

Secondary analysis is performed for all pupils (EverFSM6 and not EverFSM6 together). Impact analysis was conducted for the 5,188 pupils who were randomised.

Table 6 shows that there is a very small difference between the two groups at pre-test, with the treatment group ahead, and also ahead at post-test. Overall, the intervention appears to have no impact on the overall pupils' gain scores (ES =0).

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 (n = 2,569)	0.005	0.98		0.015	0.99		0.009	0.79	
Control (n = 2,619)	-0.005	1.02		0.004	1.00		0.009	0.80	
Overall (n = 5,188)	0.000	1.00	+0.01	0.009	0.99	+0.11	0.009	0.80	0.00

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

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

Impact analysis (Table 7) shows that the intervention benefitted FSM pupils to about the same extent as all pupils in general. As with the Year 11 cohort the intervention did not specifically benefit FSM pupils as it does with EverFSM6 pupils.

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

	Pre-score mean	SD	ES	Post-score mean	SD	ES	Gain score	SD	ES
Treatment (n = 307)	-0.46	1.01		-0.59	0.85		-0.130	0.86	
Control (n = 328)	-0.50	1.05		-0.63	0.88		-0.130	0.80	
Overall (n = 635)	-0.48	1.03	+0.04	-0.61	0.87	+0.05	-0.130	0.83	0.00

Additional analysis

Fidelity analysis

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

A total of 5,188 pupils with KS2 and Attainment 8 scores were meant to complete the writing exercises. Of these, around half (n = 2,569) were in the intervention group who wrote about values important to them. The other half in the control group were given an alternative writing task.

Table 8 shows the over 60% of the intervention pupils completed all three writing exercises.

Table 8: Number of exercises completed by the intervention group

Number of exercises completed	Intervention (%)
0	128 (5.0%)
1	211 (8.2%)
2	626 (24.4%)
3	1581 (61.5%)
Missing	23 (0.9%)
Total	2569

Fidelity analysis comparing the number of exercises completed and outcomes shows a positive but small correlation for all pupils together as well as for EverFSM6 pupils (Tables 9 and 10). This is larger for post-test scores than for gain scores. Both sets of figures suggest that compliance is stronger for students with higher absolute levels of attainment, irrespective of the impact of the intervention. The results are similar to those for both cohorts although there is a stronger correlation for the Year 10 cohort (0.34 compared to 0.14 for the Year 11 cohort). This could be an indication of the longer term effect.

Table 9: Correlation between gain scores and number of exercises completed (all pupils)

	Gain scores using KS2 maths & reading combined	KS4 Attainment 8 score
Number of exercises completed	+0.24	+0.34

N = 5,188

Table 10: Correlation between gain scores and number of exercises completed (EverFSM6 pupils only)

	Gain scores using KS2 maths & reading combined	KS4 Attainment 8 score
Number of exercises completed	+0.26	+0.36

N = 1,372

The second method of looking at compliance is the Complier Average Causal Effect Analysis (CACE). Instead of intention to treat, this estimates the likely treatment effects for compliers where not all pupils in the treatment group complied with the intervention. Compliance is defined here as completion of the first writing task. Since the control group was already ahead at pre-test, CACE analysis was conducted using the standardised gain scores (Table 11).

Analysis of compliance (defined as completion of first writing task) using the Complier Average Causal Effect Analysis (CACE) based on the standardised gain scores and using the overall standard deviation shows a positive but very small effect (see Table 11) – ES= +0.05. This is similar to but slightly larger than the simple gain score ES of +0.04 (and the same as the simple ES divided by 0.81, which is the level of compliance in the treatment group).

Table 11: 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	549	-.005	125	-0.56	674	-0.10	
Control	569	-0.03	129	-0.56	698	-0.13	+0.05

Note: the N in red are based on there being the same proportion of compliers in the control group as in the treatment group (549/674), and the mean scores in red are based on the non-compliers in the control group having the same mean as those in the treatment group.

Regression analysis

To determine how much the results could be attributed to the intervention, two multivariate regression models were created using Attainment 8 KS4 scores as the dependent variable and total prior test scores (total KS2 maths and reading) and membership of treatment group as predictors. One model uses all pupils and the second uses only EverFSM6 pupils. As usual, the best predictor is pupil prior attainment at KS2 (Table 12). Once this is accounted for, there is little or no impact from the treatment (ES = 0.01). The result is similar to that for the Y11 cohort.

Table 12: Regression results with KS4 Attainment 8 as the dependent variable

	All initial Year 10 pupils	EverFSM6 pupils
R	0.68	0.61
Coefficient for KS2 maths and reading points as pre-test	0.68	0.61
Coefficient for treatment group	0.00	0.01

Another way of looking at compliance is using the number of exercises completed with zero for control pupils. The results are similar to Table 12, but R is slightly larger (see Table 13). Prior attainment is still the best predictor of outcomes but number of exercises completed has a coefficient of 0.30. This may mean that the intervention led to better outcomes if implemented well, or that the kinds of pupils who completed more exercises also made more progress. Results are similar for both cohorts although the correlation is stronger for the Y10 than for the Y11 (0.30 vs 0.15).

Table 13: Regression results with number of exercises completed and KS4 Attainment 8 as the dependent variable

	All pupils Year 11	All pupils Year 10	EverFSM6 pupils (Y11)	EverFSM6 pupils (Y10)
R	**0.67**	**0.73**	**0.61**	**0.69**
Coefficient for KS2 maths and reading points as pre-test	0.66	0.65	0.59	0.58
Coefficient for completed exercises	0.10	0.25	0.15	0.30

Impact on pupils' self-efficacy

Results are available for only the Year 10 cohort who provided both the pre-survey and post-survey data. As agreed with the funder and the developer, only pupils that completed the baseline survey are included in the analysis. One school did not complete the survey, and in some schools a few classes missed the baseline survey.

Since we exclude those without the baseline pre-survey scores, we want to check if excluding them would in any way bias the results. To do this, we first compare the post-survey scores for each of the four self-efficacy items for those missing pre-survey and not missing pre-survey (Tables 14a to 14d).

Comparing post-survey scores of those who completed pre-survey with those who did not.

Table 14a: Self-efficacy 1: I'm certain I can understand the most difficult material presented in texts

	Post-score mean	SD	ES
Not missing (n =3016)	4.24	1.37	
Missing (n = 810)	4.09	1.45	
Overall (n = 3826)	4.21	1.39	+0.11

Table 14b: Self-efficacy 2: I am confident I can understand even the hardest things taught by the teachers

	Post-score mean	SD	Effect size
Not missing (n=3037)	4.10	1.45	
Missing (n = 813)	3.95	1.49	
Overall (n = 3850)	4.06	1.46	+0.12

Table 14c: Self-efficacy 3: I'm confident I can do an excellent job on assignments and tests

	Post-score mean	SD	Effect size
Not missing (n=3040)	4.27	1.41	
Missing (n = 816)	4.08	1.47	
Overall (n = 3856)	4.23	1.43	+0.13

Table 14d: Self-efficacy 4: I am certain I can master the skills being taught

	Post-score mean	SD	ES
Not missing (n = 3031)	4.56	1.31	
Missing (n = 810)	4.42	1.36	
Overall (n = 3841)	4.53	1.32	+0.11

For the four items on the self-efficacy measures, those with pre-test scores had higher post-test scores than those missing pre-test (average ES = 0.12). This may bias the results if there is an imbalance between the two treatment groups. To check if this was the case, we compare the post-survey scores of those missing pre-survey in the two treatment groups (see Table 15).

Table 15: Comparison of post-test scores between treatment groups for those missing pre-test

	Treatment group	Control group	Overall	Effect size
Self-efficacy 1 Mean (sd)	4.06 (1.39) n = 424	4.12 (1.49) n =386)	4.09 (1.45) n = 810	-0.04
Self-efficacy 2 Mean (sd)	3.93 (1.45) n = 426	3.93 (1.52) n =387)	3.93 (1.49) n = 813	0
Self-efficacy 3 Mean (sd)	4.09 (1.42) n = 428	4.08 (1.53) n =388	4.08 (1.47) n = 816	0.006
Self-efficacy 4 Mean (sd)	4.42 (1.29) n = 423	4.43 (1.43) n =387	4.42 (1.36) n =810	-0.007

Table 15 shows no difference in post-test scores between treatment groups for those missing pre-test, suggesting that the missing pre-test is unlikely to affect the overall results.

Table 16: Improvement in self-efficacy scores between treatment groups

	Pre-survey mean (SD)	ES	Post-survey mean (SD)	ES	Gain score (SD)	ES
Treatment (n = 1522)	4.51 (1.12)		4.28 (1.16)		-0.228 (1.15)	
Control (n = 1526)	4.56 (1.11)		4.31 (1.19)		-0.25 (1.19	
Total (n = 3048)	4.53 (1.11)	-0.05	4.29 (1.17)	-0.02	-0.24 (1.17)	+0.02

Comparison of mean gain scores show a very small but positive gains in pupils' self-efficacy. The post-survey was taken immediately after the last writing task. Therefore, the long-term effect of the intervention on self-efficacy could not be ascertained here. This could be explored in future studies.

Conclusion

Key conclusions
1. Among disadvantaged pupils (defined as EverFSM6), those who received the self-affirmation intervention made slightly more progress between the end of primary school and GCSEs than the comparison pupils, but the size of the impact was very small, at +0.05 (1 month progress) for the Y11 cohort and +0.04 (0 months progress) for the Y10 cohort.
2. Pupils who completed more writing exercises made more progress – a reasonable correlation of +0.26. This is stronger than for the Y11 (0.09), perhaps suggesting that there is a sustained effect from the intervention. This may mean that the intervention can lead to better outcomes if implemented more thoroughly, but might also be because the kind of pupils who completed more exercises would make more progress anyway.
3. The intervention takes 10 minutes, delivered 3 times a year, and costs nothing for schools to implement.
4. In this trial neither pupils nor teachers were aware of the nature of the intervention or why the tasks were undertaken because there is some evidence that knowledge about the purpose of the intervention can reduce its effectiveness. Therefore, schools considering this approach should bear in mind the practical difficulty in replicating these conditions.

Interpretation

The results of the Year 10 cohort show that the positive effects of the intervention are sustained over one year, and that the intervention has a small impact on pupils' self-efficacy too. Disadvantaged pupils who received the intervention made slightly more progress between KS2 and KS4 (ES = +0.05 for Y11 cohort; +0.04 for Y10 cohort) than pupils who did not receive the intervention. Because of differences in the prior attainment of the groups, the results may be sensitive to how these differences are accounted for in the analysis. When prior attainment is controlled for in regression analysis, the small effect is even smaller, but the number of completed exercises remains a useful predictor. In line with theory, the intervention shows no effect for all pupils in general (that is including non-disadvantaged pupils) (ES = -0.01 for Y11; 0 for Y10). This is consistent with previous research suggesting that the intervention can help to mitigate against the negative effect of being stereotyped for being a member of a group that is often performing poorly academically (Oyserman et al., 2006; Cohen et al., 2006; Miyake et al., 2010). The theory is that getting pupils to write positive statements about themselves can give pupils a sense of value and thus alleviate any negative perceptions that they may have of themselves. This improves their self-concept and confidence which in turn can affect their peers' and teachers' expectations of them to do well.

The results of the self-efficacy survey show that treatment pupils also made a small improvement in their self-efficacy compared to control pupils who were not exposed to the intervention.

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 only 10 minutes, delivered 3 times a year and costs literally nothing, there is hardly any opportunity cost for schools to practice this approach. Although the impact is small, the positive correlation between number of exercises completed and the outcomes, plus the fact that the impact was sustained, it might still be valuable in practice (Gorard, 2006).

There is currently much policy and practice activity on raising individual's attitudes, aspirations, and self-concept to improve academic outcomes, but much of the evidence so far has been correlational (for example, Marsh and Martin, 2011; Marsh, 1990; Pinxten et al., 2010; Valentine and Dubois, 2005; Skaalvik and Valås, 1999; Marsh and Craven, 2006; Marsh and O'Mara, 2008; Grabowski et al., 2001). Previous research by Gorard, See and Davies (2012) indicates the difficulty in demonstrating the causal mechanism of these attitudinal concepts on academic attainment. The self-efficacy survey measures pupils' self-concept, motivation, self-belief/self-esteem, and locus of control. And the results suggest a small positive benefit of the intervention on pupil's self-efficacy. However, the post-survey was taken immediately after the last writing task. Therefore, the long-term effect of the intervention on self-efficacy could not be ascertained here. This could be explored in future studies.

Limitations

In this trial, the agreed prior attainment was the KS2 reading and maths scores, and the post-test was the KS4 Attainment 8 scores. Analyses were presented using KS2 reading and maths combined and the gain scores for the post-intervention outcome, because of the imbalance at pre-test, although post-test only figures are also presented for each result. The results vary slightly whether using reading (ES = +0.04) or maths (ES = +0.02), with the combined maths and reading showing a bigger effect (ES = +0.05; 0.04). However, we cannot be sure that the same effects would be achieved if different measures (for example, English rather than reading) were used for both the prior attainment and the post-test.

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

Another limitation is the use of EverFSM6 as a proxy for disadvantage. As have been shown by many (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 impact of such an intervention, which is to address negative experiences associated with membership of a disadvantaged group. Perhaps a more accurate measure would be parental income or occupational status, or permanent FSM status.

Future research and publications

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

- separate analyses of English and maths KS4 results for all and EverFSM6 students;
- assessment of intervention for various subgroups (for example, gender, ethnicity, low attainment groups);
- moderation of the intervention by class- and school-level variables (proportion of FSM, size, OFSTED rating, and so on);

- mediation analyses of non-attainment measures

It is envisaged that there will be at least one publication from this project in a peer reviewed journal based on the impact evaluation.

Future research could address the following questions:

- Is the improvement in attainment outcomes a result of the transfer mechanism via improvement in self-efficacy?
- Is there a differential effect for subgroups of pupils (for example, gender, ethnicity and low attainment groups)?
- Do the class- and school-level variables (such as Ofsted rating, proportion of FSM, and size) have a moderation effect?
- Can the results be sustained if rolled out, and there is a likelihood that participants will be aware of the purpose?

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gorard, S. (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, pp. 1-15.

Gorard. S. (2018) *Education policy: Evidence of equity and effectiveness*. London: Policy Press.

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

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

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

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

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

Hobbs, G., & Vignoles, A. (2010) 'Is children's free school meal 'eligibility' a good proxy for family income?', *British Educational Research Journal*, 36, pp. 673–690.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Stephens N. M, Fryberg S. A, Markus, H. R, Johnson, C.S., Covarrubias, R. (2012) Unseen disadvantage: how American universities' focus on independence undermines the academic performance of first-generation college students. *J Pers Soc Psychol*, 102, pp. 1178–1197.

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, pp. 29-51.

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

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

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

Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP