Case Study 1: Evidence Based Practice Report

Theme: School/setting Based Interventions for Learning.

How effective are self-affirmation writing interventions in raising the academic achievement of Black students in the context of susceptibility to stereotype threat?

<u>**Summary**</u>

Writing interventions which entail reflecting upon one's meaningful personal values, are conceptualised as self-affirmation exercises. They aim to support students by reinforcing their self-esteem (Cohen, Garcia, Apfel, & Master, 2006). This brief approach proposes to alleviate the stress and anxiety that Black students may face, when educated in contexts where a minority status may rehash negative stereotypes about their academic competency. This is known as "stereotype threat" (Steele & Aronson,1995). This systematic literature review critically appraises the effectiveness of school-based self-affirmation writing exercises, in improving the academic achievement of Black students. Six research studies including students aged 11-15 years old, were examined. Effect sizes were largely negligible, with one study reporting a medium effect. The findings propose that whilst self-affirmation writing exercises can demonstrate statistical effects, the impact of these effects are insubstantial. Educational Psychologists could alternatively explore self-affirmation interventions that are tailored to the academic outcomes they propose to improve.

Introduction

The Psychological Theory of Stereotype Threat

Historical theories of intelligence often highlighted differences in academic ability (via Intelligence Quotient (IQ) testing) between Black students and students of other ethnic groups (Neisser et al. 1996). Unfortunately, findings of lower performance on such tests by Black students, were attributed by Educational Psychologist Arthur Jensen to racial differences in genetics (Jensen, 1972).

Whilst such views remain unfounded (Neisser et al. 1996), they have had implications in the enduring existence of negative stereotypes about the academic competency of Black students. The extent of this impact has been examined within psychological literature as "stereotype threat" (Steele & Aronson, 1995). Stereotype threat has been conceptualised as the psycho-social experience of belonging to a group which has been the recipient of negative typecasts regarding theories of intelligence; whereby awareness of such theories can contribute to the development of "inferiority anxiety" when faced with threats in an academic environment (Steele & Aronson, 1995)

Regardless of whether an individual subscribes to the stereotype, simply sharing membership to their ethnic group can make them susceptible to its effects (Hanselman, Rozek, Grigg, & Borman, 2017). Such effects include impediments upon one's performance and achievement through the experience of arousal and stress (Cohen et al. 2006) and the development of a negative self-consciousness which prompts personal worries (Taylor & Walton, 2011). Both experiences are thought to inhibit a student's ability to effectively engage in a learning task; through impeding skills such as the ability to manage cognitive load (Taylor & Walton, 2011). The outcomes of these experiences can extend to attributions about one's belonging in academic environments and ability to achieve (Pennington, Heim, Levy, & Larkin, 2016; Taylor & Walton, 2011); which can aid disengagement and result in lowered academic performance.

How self-affirmation writing interventions alleviate the impact of stereotype threat

To reduce the harmful impact of stereotype threat, interventions with a socio-psychological foundation have been proposed to help students change the way they view themselves in relation to their academic environment and their ability. These interventions range from increasing the visibility of role models who negate negative typecasts (Shapiro, Williams, & Hambarchyan, 2013); promoting intergroup similarities (Rosenthal & Crisp, 2006), teaching students that intelligence can be malleable (Good, Aronson, & Inzlicht, 2003) and affirming students' values to promote belonging and reduce stress (Cohen et al. 2006; Shnabel, Purdie-Vaughns, Cook, Garcia, & Cohen, 2013).

Nevertheless, all interventions centre on the concept of increasing student belonging within academic environments and appear to draw upon theories of belonging (Shnabel et al. 2013). Such theories (Baumeister & Leary, 1995; Maslow, 1943) propose that belonging is a fundamental human psychological need which serves as a prerequisite to be able to self-actualise and achieve one's full potential. This also includes academic potential; whereby experimental evidence has shown that social exclusion can lead to large decrements in school test performance and on IQ tests (Baumeister, Twenge, & Nuss, 2002); through diminishing information processing, logic and reasoning skills.

Self-affirmation interventions are thought to be most simply and effectively employed prior to academic testing (Cohen et al. 2006), by having students select a list of values and then engage in affirmative writing regarding those values. This is thought to negate stereotypes and consequently reduce the cognitive load and defences (Bowen, Wegmann, & Webber, 2013). Most namely, Cohen et al. (2006) designed a self-affirmation intervention which demonstrated noteworthy effectiveness for Black students. The intervention was deemed to have long term effects of raising these students' Grade Point Averages (GPA) up to 0.41 points and reducing the Black-White student achievement gap by up to 40% (Cohen et al. 2006; Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009). Thus, this research

serves as the foundation of this review; whereby replicable studies were sought in order to establish consensus within the literature on the effectiveness of this intervention.

Rationale and Relevance to the practice of Educational Psychologists

Whilst stereotype threat has existed within literature since 1995, a surge in interventions in the United States coincided with the No Child Left Behind Act (2002-2015); which focused on addressing the achievement gap of minority ethnic groups. Extensive literature explores the achievement gap in the context of stereotype threat amongst Black students (Cohen et al. 2009; Ladson-Billings, 2006; Steele & Aronson, 1995).

Similarly, within the UK, long-standing concerns over the achievement of some Black students[1] have been raised (Rampton, 1981) and more recently regarding their exclusion from school (Rollock, Gillborn, Vincent, & Ball, 2015; Timpson, 2019) and overrepresentation in alternative educational provisions (Department for Education, 2019). In a review exploring the achievement gap over the last 25 years (Gillborn, Demack, Rollock, & Warmington, 2017) practice such as specialist provision placements and Ethnic Minority Achievement Grants (EMAG) have been highlighted as ineffective in ameliorating the gap. Black students remain one to two and a half times more likely to underachieve (Gillborn et al. 2017). This extends to university education whereby a 25.3% gap exists between Black and White students qualifying with a First or 2:1 degree (Higher Education Statistics Agency, 2017). With the SEND Code of Practice (2014) extending the remit of Educational Psychologists to work with children and young people until the age of 25, it is paramount to explore school-based interventions that can re-shape the educational trajectories and consequently life outcomes of Black students.

Additionally, another publication by the Department for Education (Graham, White, Edwards, Potter, & Street, 2019) has placed an emphasis on pupils' sense of belonging and wellbeing

---

[1] The concerns predominantly centre on children of Black Caribbean / West Indian heritage, although some data has used the term "Black" to also include children of Black African heritage.

and the proactive application of equality principles as avenues for school development. Feeling undervalued and the experience of being negatively racially stereotyped was highlighted as a factor that prompted disengagement amongst pupils; reiterating the literature on stereotype threat. Furthermore, Educational Psychologists Hamilton and Morgan (2018) found that Black students[2] who transferred from alternative provision to Further Education named belonging as a key factor to their success. The authors recommended that Educational Psychologists are best placed to work with schools to facilitate changes in practice, to increase learner attainment, pupil wellbeing and a sense of belonging. This work can manifest in Educational Psychologists implementing effective school-based interventions which target belonging and achievement.

Whilst systematic reviews in the UK exist on the concept of stereotype threat (Pennington et al. 2016), at present there are few systematic reviews that have specifically looked at interventions which target it. Where these exist, reviews have considered a range of different interventions as interchangeable or looked at the effectiveness of the interventions through combining samples of different ethnic groups (Shapiro et al. 2013). This approach makes it difficult to distinguish the particular effectiveness of an intervention for Black students. In one review where this has been done (Aronson, Cohen, & McColskey, &, 2009), just three interventions were considered, yet all were distinctly different. Thus, this review aims to specifically evaluate the effectiveness of self-affirmation writing interventions in raising the academic achievement of Black students.

Review Question:

How effective are self-affirmation writing interventions in raising the academic achievement of Black students in the context of susceptibility to stereotype threat?

---

[2] Contact was made with the authors who confirmed the demographics of their sample

Literature Search:

A systematic literature search was conducted on 31st December 2019. Table 1 indicates the search terms entered into ERIC, PsychINFO, Web of Science and The Cochrane Library databases and the filters applied. The terms include a variation of demographic descriptors that may refer to students whose ethnicity is Black. In addition, the terms include variations of the terminology used to measure academic outcomes. Within the literature on stereotype threat, measures of attainment and achievement have both been used; sometimes interchangeably (Selm, Peterson, Hess, Beck, & McHale, 2019; Steele & Aronson, 1995). Terms such as "attainment gap" or "achievement gap" have also been included to gather studies that consider the impact of interventions in raising attainment of Black students, relative to pupils belonging to other ethnic groups.

120 studies resulted from the search. Table 2 displays the inclusion and exclusion criteria for screening studies. Figure 1 – a PRISMA flow diagram illustrates the study selection procedure. Six studies were included in this review for critical appraisal (Appendix A, Table 7). Appendix A also includes tables of the excluded studies and the rationale.

Table 1.

Search Terms for Databases

Participants		Intervention		Outcome
Black Students OR Black American Students OR African American* Students OR Black Brit* Students OR Black Caribbean* Students OR Afro-Caribbean* students OR African-Caribbean* Students OR African* Students OR African Descen* Students Black Pupils OR Black American Pupils OR African American* Pupils OR Black Brit* Pupils OR Black Caribbean* Pupils OR Afro-Caribbean* Pupils OR African-Caribbean* Pupils OR African* Pupils OR African Descen* Pupils	AND	Stereotype Threat Intervention* OR Reducing stereotype threat*	AND	Attainment OR achievement OR Academic achievement OR Academic attainment OR academic perform* OR academic outcome*OR attainment outcome* Achievement gap OR Attainment gap OR academic success OR attainment success

Note. Filters were applied to limit the search to publications in the English language (as for the purpose of this review, it was not possible to access or translate literature in other languages) and from the year that the concept of stereotype threat was first published - 1995 (Steele & Aronson, 1995).

truncation signifies that searches can yield results for variations of the word. For example, descen can also yield results for 'descent' and 'descendants'.

Table 2.
Inclusion and Exclusion Criteria and Rationale for Screening Studies.

		Inclusion criteria	Exclusion criteria	Rationale
1.	Participants in the sample	Any self-identifying Black students including anyone of African descent such as African American, Black Caribbean, Black African, African-Caribbean etc.	Other ethnic groups that do not self-identify as Black or of African descent, studies that included Black students but have no focused analysis on Black students and studies that do not focus on a student population.	This review is focused on self-affirmation interventions for Black students.
2.	Intervention	Self-affirmation interventions carried out in an educational setting with pupils of school age (i.e. 5 years old up to and including 16 years of age – or the equivalent of Year 1 –Year 11).	Interventions carried out at home or in the community or with pupils who are not of school age e.g. college or university students.	Stereotype threat is conceptualised as a function of situational threats in an academic environment (Steele & Aronson, 1995). The review focuses on school-based interventions.
3.	Outcomes	Attainment or achievement outcomes in relation to academic e.g. GPA and other measures of test performance.	Non-academic outcome measures e.g. recall or matching tasks and studies which do not report any outcomes.	The review is focused on academic or attainment outcome measures. These have more ecological validity pertaining to school achievement.
4.	Design	Randomised Control Trials (RCT) or Quasi Experimental Studies which include a control condition.	Studies which do not implement experimental conditions or are qualitative or correlational. Reviews, articles and meta-analyses are also excluded. Follow up studies are also excluded if they include the same sample or data set as studies already selected for the review.	In order to establish effectiveness of an intervention, controlling for additional variables needs to be considered. Correlational designs and other non-experimental designs do not allow for this (Petticrew & Roberts, 2003)
5.	Country where the research	Anywhere in the UK, USA or any other country which is a member of the Organisation for Economic Co-operation and development (OECD) as they share educational policies and	Studies conducted in Non – OECD countries and countries where Black students do not make up a minority population.	Black students who reside in countries with educational contexts similar to the UK are included as they may also experience stereotype threat and be susceptible to the achievement gap

	Inclusion criteria	Exclusion criteria	Rationale
was carried out	may be more likely to share an educational system that is similar in structure and content to the UK. To be selected the country must also have a population of Black students who make up a minority ethnic group as they are more likely to experience stereotype threat than Black students in countries where they make up a majority of the population.		(Gillborn et al. 2017; Ladson-Billings, 2006; Pennington et al. 2016) and thus, the effectiveness of the intervention may also be likely to be replicable in a UK context.
6. Date of publication	Any studies from 1995 to present published in a peer-reviewed journal	Any studies that are prior to 1995 and publications that are not from peer-reviewed journals	Stereotype threat was first conceptualised after this period (Steele & Aronson, 1995). Studies prior to this period may have a different conceptualisation of the experience. Non-peer reviewed publications may contain evidence that has not been subject to a rigorous peer review scientific process.

Figure 1.

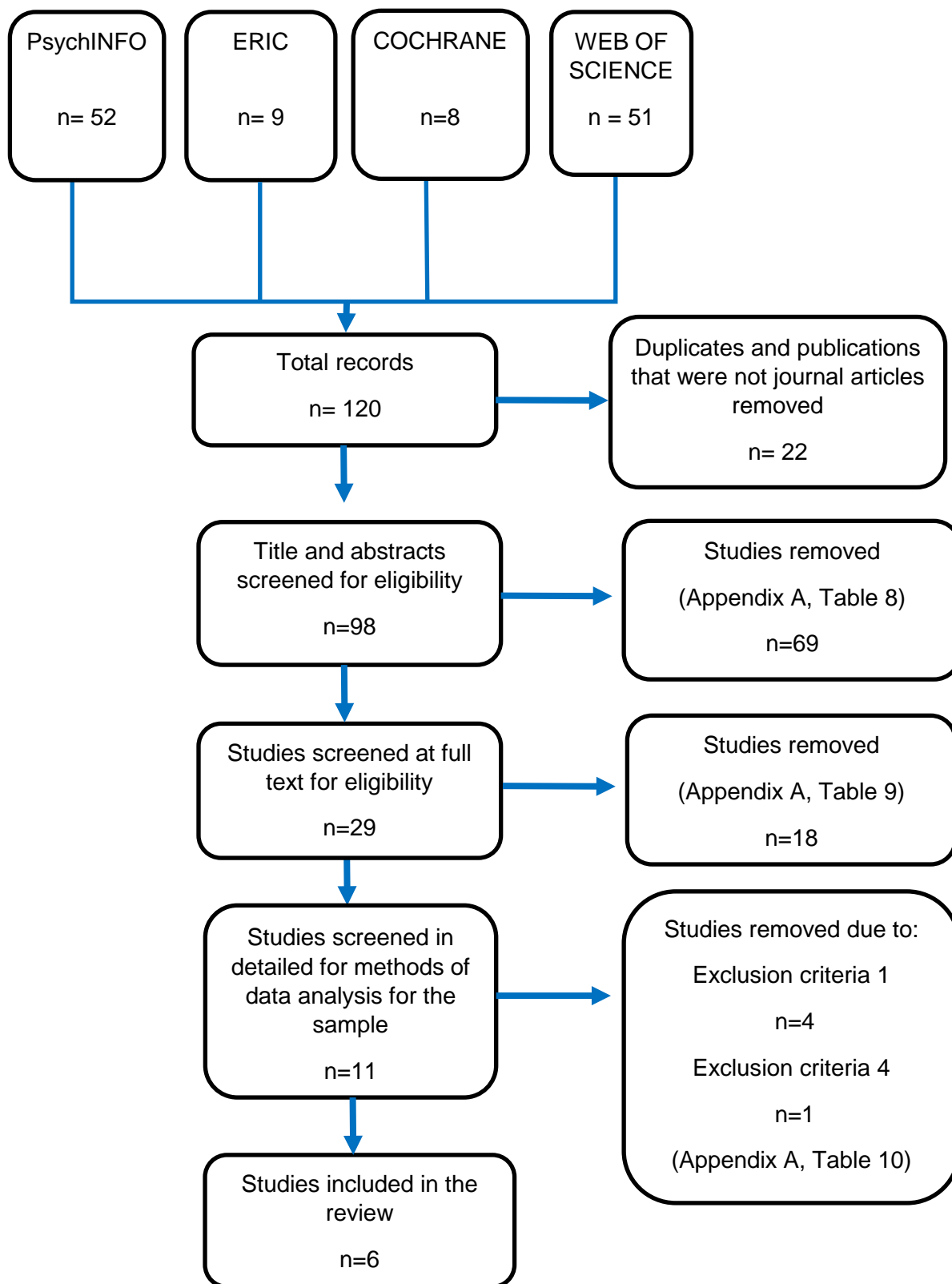*PRISMA Flow Diagram of the Study Selection Process*

PsychINFO	ERIC	COCHRANE	WEB OF SCIENCE
n= 52	n= 9	n=8	n = 51

Total records

n= 120

→ Duplicates and publications that were not journal articles removed

n= 22

Title and abstracts screened for eligibility

n=98

→ Studies removed

(Appendix A, Table 8)

n=69

Studies screened at full text for eligibility

n=29

→ Studies removed

(Appendix A, Table 9)

n=18

Studies screened in detailed for methods of data analysis for the sample

n=11

→ Studies removed due to:

Exclusion criteria 1

n=4

Exclusion criteria 4

n=1

(Appendix A, Table 10)

Studies included in the review

n=6

Table 3.

Mapping The Field: An Overview of the Included Studies

No.	Author/s	Participants (included in analyses)	Study Design	Description of the Intervention	Measures	Main Findings
1	Bancroft, Bratter & Rowley (2017)	N=514 (of which Black students = 118) 9th Grade students Age= 14-15 years old Gender = 52% Female and 48% Male Sampled from 3 public high schools. School 1 = predominantly Hispanic student body School 2 = predominantly Black student body School 3 = even representation of	Group-based experimental design Intervention group (value-affirmation writing task). Control group (value-free general writing exercise) Random assignment to conditions at student level.	The value-affirmation writing exercise replication Cohen et al's (2006) procedure (see below). 2 exercises were implemented per term before tests (this ranged from a few days before a test or up to a week before).	Standardised math performance (State of Texas Assessments of Academic Readiness - STAAR for algebra)	Effects of the intervention were not significant for Black students on maths measures.

No.	Author/s	Participants (included in analyses)	Study Design	Description of the Intervention	Measures	Main Findings
		Hispanic, Black and White students in the student body				
2	Binning, Wang & Amemiya, (2019).	$N = 598$ (of which Black students =152). 7th Grade and 9th Grade students Age 12-13 years old and 14-15 years old Gender = descriptive statistics not reported Sample from 2 Schools School 1 = predominantly Black student body School 2 = a public high school with a predominantly White student body	Group-based experimental design Intervention group (three intervention approaches which included growth mind-set training, struggle-story approach and a belonging mind-set writing exercise) Control group (study skills activities and general writing exercises) Random assignment to conditions at student level	Intervention approach 1: A Growth-Mind-set video, real-world application and writing exercise focusing on the idea that intelligence can be improved with effort. Intervention approach 2: A struggle-story lesson focused on learning strategies celebrities used to overcome their struggles and then writing reflections about the process. Intervention approach 3: Belonging mind-set testimonials from students at the school were shared and then students then wrote about their own challenges and how to overcome them.	Grade Point Average (GPA) (End of year) Pre and Post surveys on educational attainment expectations (based on a National Survey)	A significant effect of the group of 3 interventions such that Black students in the treatment group received higher GPA grades over the school year. (The authors note this was significant only for Black students with high educational expectations).

No.	Author/s	Participants (included in analyses)	Study Design	Description of the Intervention	Measures	Main Findings
		All participants were social-economically disadvantaged as indicated by free school meals or a reduced price lunch.		Each intervention takes 45mins to administer (approx. total. 2hrs 15mins). Administered at 2 time points: one initial session followed by a booster session approx. 5 weeks later.		
3	Bowen et al. (2013) (Experiment 1)	N=274 (of which 236 were Black students) 6th, 7th and 8th Graders Age = 11-14 Gender = 50.4% Female, 49.6% Male Students sampled from one school. 80% of the students received the lunch program	Group-based experimental design Treatment group (writing a self-affirmation values essay) Control group (writing a neutral essay) Random assignment to conditions at student level	The value-affirmation writing task was a replication of Cohen et al's (2006) procedure (see below). The writing exercise lasted 15 minutes. The intervention was administered 2 weeks prior to the end of the first school testing period.	Social Studies Grades (quarterly)	Significant effects found for all students (including Black students) such that self-affirmation intervention reduced underachievement in Social Studies Grades over time. However, there were no significant differences between ethnicities.

No.	Author/s	Participants (included in analyses)	Study Design	Description of the Intervention	Measures	Main Findings
		which was indicative of socio-economic disadvantage.				
4	Bratter, Rowley & Chukhray (2016)	N=886 (of which 227 were Black students) 9th graders Age= 14-15 years old Gender= 50% Female 50% Male 3 public schools School 1 = predominantly Black student body School 2 = predominantly Hispanic student body School 3 = student body was equally representative of	Group-based experimental design Treatment group (self-affirmation values writing task) Control group (neutral writing task) Random assignment to conditions at student level	The value-affirmation writing task was a replication of Cohen et al's (2006) procedure (see below). The writing exercises lasted between 15-20mins. 4 exercises were administered throughout the academic year, at 3 days or less, prior to school tests.	Student English grade for the term, STAAR Reading and Algebra scores.	No significant effects for the treatment on Black student's English grade or STAAR Algebra grade. However significant reverse effects were found for the treatment on Black student's STAAR reading grade (especially for those attending a school where the study body was predominantly Black).

No.	Author/s	Participants (included in analyses)	Study Design	Description of the Intervention	Measures	Main Findings
		Black, Hispanic and White students.				
5	Cohen et al. (2006) *(Experiment 1)*	N=133 (of which 51 were Black students 7th graders Age=12-13 years old Gender = 55% Female, 45% Male All the students were from middle to lower-middle class families.	Group-based experimental design Treatment group (self-affirmation values writing task) Control group (values writing task about other people) Random assignment to conditions at student level.	Students are presented with a list of 12 values from which they select 2 or 3. They then write about how meaningful each of the values are personally to them; giving two reasons why and then rating their agreement with the values. The control group repeat this process, instead writing about why the values are of importance to someone other than themselves. The writing exercises lasted 15mins.	GPA grades (termly)	Higher grades were achieved by Black students in the treatment condition vs control conditions. Those who benefitted the most were the lowest performing students.

No.	Author/s	Participants (included in analyses)	Study Design	Description of the Intervention	Measures	Main Findings
				One exercise was completed in Experiment 1.		
6	Shnabel et al. (2013) *(Study 1 only)*	N=355 (of which Black students =169) 7th graders Age = 12-13 years old Gender= 52% Female, 48% Male The student body of the school was around 50/50 between Black students and White students. Middleclass to lower-middle class socio-economic status	Group-based experimental design Treatment group (self-affirmation values intervention) Control group (values writing exercise focused on other people) Random assignment within classrooms	The value-affirmation writing task was a replication of Cohen et al's (2006) procedure (see above). The control group repeat this process, instead writing about why the values are of importance to someone other than themselves. 2 of the 3 cohorts completed a follow up exercise one month later - although only the first intervention exercise was included in the analyses.	Post-intervention change in GPA from their grades from the first term of the 7th Grade.	Significant effects found for improving the GPA of Black students in the treatment condition for the value-affirmation intervention. (The authors argue that this was specifically due to writing about the value of belonging).

Critical Appraisal of the Reviewed Studies:

In order to provide a synthesis of the research studies; Gough's (2007) Weight of Evidence (WoE) framework was employed. It consists of making three judgements (A, B and C) regarding WoE A - the quality of the evidence in relation to studies of the same type, WoE B - the appropriateness of the methodology in providing evidence in relation to the review question, and WoE C – an evaluation of the suitability of the characteristics of the research evidence to answering the review question. All criteria can be found in Appendices B, C and D.

In undertaking judgements for WoE A, the Gersten et al. (2005) coding protocol (Appendix B) was employed as all studies were experimental field designs carried out in schools. The protocol is suitable as it allows for an evaluation of the quality of the methodologies employed in experimental group and quasi-experimental designs within education. The protocol consists of two sub-sets of questions which entail "Essential Quality Indicators" and "Desirable Quality Indicators." In order to allow for more scrutiny between the quality of the evidence in the areas of comparability of the experimental and control groups, fidelity of the intervention and the use of appropriate outcome measures, the protocol was adapted to include further sub-questions outlined in the descriptive criteria in Gersten et al's (2005) publication. Scores from each sub-set of indicators were summed and weighted according to Gersten et al's (2005) criteria, to give an overall WoE A weighting of Low = 1, Medium = 2 and High= 3.

Each study was then appraised for the suitability of the methodology (WoE B) and the suitability of the characteristics of the research (WoE C) in providing evidence for the review question. WoE B judgements were made based on Petticrew and Roberts (2003) typology of evidence. WoE C was comprised of an average of judgements regarding the implementation of the intervention, sample, context and the outcome measures utilised. WoE A, B and C judgements are averaged. This gives an overall weight of evidence evaluation of the extent

to which each study contributes relevant quality research evidence, to the review question –

WoE D (Gough, 2007) (Table 4).

Table 4.

Weight of Evidence (WoE) Ratings

Studies	WoE A	WoE B	WoE C	WoE D
Bancroft et al. (2017)	1 (Low)	3 (High)	1.67 (Medium)	1.89 (Medium)
Binning et al. (2019)	1 (Low)	3 (High)	1.67 (Medium)	1.89 (Medium)
Bowen et al. (2013)	1 (Low)	3 (High)	1.33 (Low)	1.77 (Medium)
Bratter et al. (2016)	3 (High)	3 (High)	2.67 (High)	2.89 (High)
Cohen et al. (2006)	1 (Low)	3 (High)	1.67 (Medium)	1.89 (Medium)
Shnabel et al. (2013)	1 (Low)	2.5 (High)	2 (High)	1.83 (Medium)

Note. <1.4 = Low, 1.5-2.4=Medium, >2.5=High

Characteristics of the included studies:

Participants

The review consists of six studies conducted in the United States, published between 2006 and 2019 (Table 5). Taken together, the studies encompassed a total of 2,870 participants, of which 1,015 were Black students. The samples of Black students ranged from 51 (Cohen et al. 2006) to 236 (Bowen et al. 2013). Most of the studies reported roughly an equal balance of male and females in the sample. These studies received higher WoE A ratings as they reported whether such gender characteristics were comparable across conditions; increasing the likelihood of eliminating whether experimental effects were due to other differences in characteristics. One study (Binning et al. 2019) did not report the descriptive statistics by gender and thus, received lower WoE A ratings on criteria assessing salient differences between participants. The participants in the study ranged from 11-15 years old (6th-9th Grade); with four of the six studies replicating the methods of the original study by Cohen et al. (2006), by focusing on pupils in the 7th Grade (age 12-13) preparing for transition to High School; where academic challenges may be more prominent (Binning et al. 2019). Three studies consisted of pupils from public schools (Bancroft et al. 2017; Binning et al. 2019; Bratter et al. 2016). Two studies included a high proportion of students of socio-economic disadvantage (Bowen et al. 2013; Binning et al. 2019). Studies which included schools (or combined school samples) with a more heterogeneous ethnic composition, received higher ratings on WoE C (Bancroft et al. 2017; Binning et al. 2019; Bratter et al. 2016). This is because the school context of these studies may more readily reflect an environment of a minority student status; wherein stereotype threat most often occurs (Hanselman, Bruch, Gamoran, & Borman, 2014).

Study Design

All studies included in the review employed student-level randomisation procedures with the exception of the research by Shnabel et al. (2013) which was randomised at class level. Random assignment at individual level may be less likely to be confounded by existing group differences in the characteristics of a class (Hutchison & Styles, 2010) and can potentially heighten the confidence in the intervention's ability to produce true effects. Thus, the study by Shnabel et al. (2013) received a slightly lower rating of 2.5 for WoE B as it satisfied all but one of the criteria for a high rating. All the studies also included control groups which consisted of Black students. This enabled a comparison of the interventions effectiveness for Black students as an individual ethnic group and thus received higher ratings. In addition, all studies included a pre-measure of achievement prior to the intervention taking place and a post-measure of achievement. With the exception of the aforementioned critiques of Shnabel et al. (2013)'s randomisation procedures, all studies received high ranks on WoE B.

Intervention

Five studies replicated Cohen et al's (2006) intervention (Table 3); which entailed students writing for 15-20 minutes about values that were of importance to them. The control groups either wrote about the importance of these values to someone else (Cohen et al. 2006; Shnabel et al. 2013) or a general writing task that was not focused on values (Bancroft et al. 2017; Bowen et al. 2013; Bratter et al. 2016). One study (Binning et al. 2019) implemented a three-part self-affirmation writing intervention which focused on growth mind-set, struggle stories and belonging mind-set strategies to overcome difficulties. Each of the three interventions required a 45 minute delivery. To establish the effectiveness of self-affirmation interventions; delivery prior to an upcoming testing period was essential to all studies. The interventions ranged from being delivered between 3 days to 2 weeks prior to school tests.

Studies which implemented the intervention at more than one time point received higher ratings on WoE C (Bancroft et al. 2017; Bratter et al. 2016; Shnabel et al. 2013) as they

allow the long-term effects upon grade trajectories to be explored and can provide insight into critique that a brief one-time intervention can have dramatic effects upon academic achievement (Bowen et al. 2013). The time between the initial interventions and the booster sessions, ranged from between one month to one term. Studies which administered the interventions simultaneous to other interventions received lower ratings for WoE C (Binning et al. 2019), as it is difficult to disentangle the sole effects of the self-affirmation intervention.

Measures

Three studies (Binning et al. 2019; Cohen et al. 2006; Shnabel et al. 2013) replicated the use of Grade Point Average (GPA) as a global outcome measure of academic achievement as in Cohen et al's (2006) original study. GPA is a teacher rating which can be comprised of academic scores on assessments for a range of subject areas. However, some researchers (Siegal & Anderson, 1991) have noted there can be considerable variability across schools in which subjects are taken into account for the GPA score and whether these subjects are more heavily weighted. Consequently, some variations in its conceptualisation can exist meaning it is not a completely standardised measure of academic achievement. However, for use in field experiments, GPA can give a study high ecological validity. Although, the reliability of GPA as a measure of academic achievement has been found to range from just under the acceptable level (.64) to excellent (.94) (Bacon & Bean, 2006). Therefore, studies utilising GPA as measures received medium ratings on WoE C, compared to studies which utilised standardised global measures of academic performance (Bratter et al. 2016) such as the State of Texas Assessments of Academic Readiness (STAAR); which received high ratings. This measure assesses the student's application of knowledge and skills from the curriculum. The Human Resources Research Organization, (2016) reports the projected reliability of the STAAR from .79-.92 for the subscales of Writing, Reading, Maths, Science and Social Studies. Additionally, the inter-rater reliability ranges from 86.2-98.8%; demonstrating very good reliability. Similarly, Bancroft et al. (2017) utilised a singular Maths measure of the STAAR. One study used Social Studies grades (Bowen et al. 2013) though

information on the standardisation and reliability of this measure was not available. Thus, these studies received low ratings for WoE C outcome measures, as their ability to measure academic achievement is unknown.

Outcomes

Three studies reported significant effects in improving academic outcomes for Black students (Binning et al. 2019; Cohen et al. 2006; Shnabel et al. 2013). One study (Bowen et al. 2013) found significant effects of the intervention on all students in the treatment group, which also included Black students. This study was also the only publication where the authors reported the effect size (hedge's $g = .57$). In order to make a consistent comparison, effect sizes were calculated using Lipsey and Wilson's (2001) formula where the unstandardised Beta regression co-efficient is divided by the standard deviation to give a Cohen's d value. Cohen's (1988) d descriptors were used to describe the effect size (Table 5).

Bowen et al. (2013) reported a medium effect size ($d=.57$) for self-affirmation interventions on Social Studies Grades. However, this effect was found for all students in the intervention group and there were no significant differences by ethnicity; meaning the intervention was suitable for all students regardless of ethnicity. Even though this study had a medium WoE D rating, the outcomes of this study would be considered to be of higher value as it is likely that the findings can be applied in schools as a universal and/or targeted intervention.

Binning et al. (2019) also found significant effects upon GPA, but these were negligible ($d=.11$) and only pertained to Black students with high educational expectations. This study also had a medium WoE D rating. Whilst the outcomes of this study pertain to a specific group (i.e. Black students with high educational expectations), the findings from this study may be considered to be of minimal value in that the intervention shows outcomes for a group that may be least likely to present needs for an educational intervention. Thus, the intervention may be ineffective for the most vulnerable Black students who may be

underachieving. Nevertheless, the intervention maybe useful in supporting Black students with high educational expectations to continue to thrive in their progress.

The original intervention designed by Cohen et al. (2006) found significant effects of the intervention on GPA for Black students, but again, these effects were also negligible (d=.18). This study also received a medium WoE D rating, and despite receiving a high WoE B rating for its methodology, it seems that its ability to produce substantial targeted effects for Black students was minimal. Thus, the findings of this study may mean that the intervention may only have a very small real-world effect for Black students.

Shnabel et al. (2013) also found significant effects of the self-affirmation intervention upon GPA and this was a marginally small effect (d=.2) and only met Cohen's (1988) descriptive criteria after rounding up the original value of d=.195. Similarly, this study also had a WoE D rating of medium and again it is likely that the outcomes of the intervention are well exceeded by the level of investment required by students and teachers.

Lastly, Bratter et al. (2016) was the only study to report significant reverse effects of the intervention in reducing reading grades (d=-.08), although these were also negligible. Interestingly, this study was the only study to receive a high ranking for WoE D. It seems that despite high quality evidence and a sound methodology, the interventions ability to predict either an increase or a decrease in academic performance is not substantial enough to be recommended for use in educational settings.

Table 5.

Effect Sizes for Intervention Outcomes

Study	Sample Size	Measure	Outcomes	Reported Statistics	Power (Calculated with G* Power) and sample size required	Effect Size	95% CI	Descriptor	WoE D Descriptor
Bancroft et al. (2017)	*N*= 514	STAAR algebra grades	No significant effects of the intervention found for Black students.	*β*=71.505 *SEM* = 137.927 Calculated *SD* = 3127.02	0.95 *N*=863 Underpowered	*d*=.02	-0.15-0.20	Negligible	1.89 Medium
Binning et al. (2019)	*N* = 598	End of year GPA	A significant effect (*p*<0.01) of the group of 3 interventions was found, but accounted for by Black students with high educational expectations.	*β*=0.25 *SEM* = 0.09 Calculated *SD*=2.2	0.95 *N*=215 Sufficiently powered	*d*=.11	-0.05-0.27	Negligible	1.89 Medium
Bowen et al. (2013)	*N* = 274	Quarterly Social Studies Grades	Significant effects (*p*<0.01) of the intervention found for all students	Hedges *g* =.57	0.95 *N*=38 Sufficiently powered	Converted to Cohen's *d*=.57	-	Medium*	1.77 Medium

Study	Sample Size	Measure	Outcomes	Reported Statistics	Power (Calculated with G* Power) and sample size required	Effect Size	95% CI	Descriptor	WoE D Descriptor
			(including Black students), but this did not differ by ethnicity.						
Bratter et al. (2016)	$N = 886$	STAAR reading, algebra and English grades	No significant effects of the intervention found. Significant ($p<0.001$) reverse effects found for Black student's STAAR reading grades.	Intervention regressed on STAAR reading when interaction effects between school type and treatment are included in the model. β=97.80 SEM=41.98 Calculated SD=1249.57	0.95 N=281 Sufficiently powered	d= -.08	-0.21-0.05	Negligible	2.89 High
Cohen et al. (2006)	$N = 133$	Fall term GPA	Significant effects ($p<0.05$) of the intervention found for Black students	β=.29 SEM=.14 Calculated SD= 1.61 (Values obtained from the online	0.95 N=155 Underpowered	d=0.18	-0.16-0.52	Negligible	1.89 Medium

25

Study	Sample Size	Measure	Outcomes	Reported Statistics	Power (Calculated with G* Power) and sample size required	Effect Size	95% CI	Descriptor	WoE D Descriptor
				supplementary information)					
Shnabel et al. (2013)	$N = 355$	GPA	Significant effects ($p<0.001$) of the value-affirmation intervention when it focused on belonging for Black students	Intervention (focused on belonging values only) regressed on outcomes β=.347 *SEM* = .095 Calculated *SD* = 1.79	*0.95* *N*=130 Sufficiently powered	*d*=.2 (only significant through rounding up *d*=.195)	-0.01-0.40	Small	1.83 Medium

Note. Cohen's *d* effect sizes are typically assigned the labels of "small" if *d*=0.2, "medium" if *d*=0.5 and "large" if *d*=0.8.

Conclusions and Recommendations:

Summary of the Review

This review examined the effectiveness of self-affirmation interventions in raising achievement for Black students in the context of susceptibility to stereotype threat. Of the six studies reviewed, five studies were of medium quality and one (Bratter et al. 2016) was ranked as high. Five of the studies attempted to replicate the original intervention devised by Cohen et al. (2006) enabling the review to explore how replicable the findings of the intervention were.

Four studies reported significant effects. Of these four, three employed GPA as an outcome measure suggesting that the intervention has little effects on individual standardised measures such as the STAAR, but can be effective upon overall academic achievement measures such as GPA. However, only one of these reported a medium effect size, whilst the others were marginally small or negligible. Of note, Bratter et al. (2016) found evidence of self-affirmation interventions demonstrating reverse effects in reducing academic achievement for standardised reading outcomes (STAAR – Reading); although the effect of this was also deemed negligible.

Overall, the findings of this review suggest that whilst self-affirmation interventions are simply employable classroom strategies that have statistically significant results, the effect sizes generally suggest that the extent of the impact upon the academic achievement of Black students is negligible. This is supported by prior views (Bowen et al. 2013) that has raised scepticism over Cohen et al's (2006) proposal of a15 minute intervention contributing to dramatic changes in academic achievement.

As a socio-psychological based intervention, it is plausible that the medium effects demonstrated by Bowen et al. (2013), could have been due to a focus on Social Study Grades as an outcome. This measure of sociological knowledge could be said to be closely related to the socio-psychological content of the intervention. It is reasonable to hypothesise

for studies which utilised GPA as an outcome, whether the effects are due to higher weighting or inclusion (Seigal & Anderson, 1991) of socio-psychological subjects as part of their GPA; resulting in contributing more prominently to the overall effects found in measures of academic achievement. Null effects found on individual measures (e.g. STAAR algebra) may be due to the unlikelihood of the socio-psychological self-affirmation intervention targeting factors such as motivation or specific learning needs, which may prevail even if stereotype threat has been alleviated (Aronson et al. 2009).

Limitations

This review contained extensive inclusion and exclusion criteria which afforded only six studies to be critically appraised. Much more research presently exists on the impact of self-affirmation interventions in contexts where threat is deemed to be higher (Hanselman et al. 2014). The studies included in the present review, did not explicitly manipulate the experience of stereotype threat and rather presumed homogeneous susceptibility to its effects based on ethnic group membership. This may not be the case as Binning et al. (2019) demonstrated variations in effects amongst Black students with different educational expectations. It is likely that a more comprehensive review that encompasses studies with explicit manipulation of stereotype threat such studies, may yield different conclusions.

Secondly, this review began with an attempt to explore whether self-affirmation interventions could be applicable to a UK context. This was not possible as the studies included in the review were all conducted in the United States and largely by persons belonging to the same research teams. Thus, a more comprehensive review could allow for potential variations in effectiveness of the intervention by context to be explored, as well as to gather a wider range of evidence from a range of independent researchers.

Lastly, this review included a range of studies which all utilised various outcomes measures of academic achievement and this limits the ability to draw a consensus on the findings with regards to academic achievement.

Areas of Further Research

Future research may wish to specifically explore self-affirmation interventions that focus on the affirming academic competency. For example, teaching concepts of malleable intelligence has been found to effectively improve standardised test performance for minority students (Good et al. 2003). In line with the outlined limitations that self-affirmations rooted in belonging may not change motivation, effort or learning needs, (Aronson et al. 2009), it is likely that interventions which focus on malleable intelligence, may specifically target the outcome measures of standardised test performance more effectively.

References:

Aronson, J., Cohen, G., McColskey, W., & (ED), R. E. L. S. (2009). Reducing Stereotype Threat in Classrooms: A Review of Social-Psychological Intervention Studies on Improving the Achievement of Black Students. Issues & Answers. REL 2009-076. In *Regional Educational Laboratory Southeast*.

Bacon, D. R., & Bean, B. (2006). GPA in research studies: An invaluable but neglected opportunity. *Journal of Marketing Education*, *28*(1), 35–42.

Bancroft, A., Bratter, J., & Rowley, K. (2017). Affirmation effects on math scores: The importance of high school track. *Social Science Research*, *64*, 319–333.

Baumeister, R. F., & Leary, M. R. (1995). The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation. *Interpersonal Development*, *117*(3), 57–89.

Baumeister, R. F., Twenge, J. M., & Nuss, C. K. (2002). Effects of social exclusion on cognitive processes: Anticipated aloneness reduces intelligent thought. *Journal of Personality and Social Psychology*, *83*(4), 817–827.

Binning, R. K., Wang, M. Te, & Amemiya, J. (2019). Persistence Mindset among Adolescents: Who Benefits from the Message that Academic Struggles are Normal and Temporary? Journal of Youth and Adolescence, 48(2), 269–286.

Bowen, N. K., Wegmann, K. M., & Webber, K. C. (2013). Enhancing a Brief Writing Intervention to Combat Stereotype Threat among Middle-School Students. *Journal of Educational Psychology*, *105*(2), 427–435.

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 L, Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: a social-psychological intervention. *Science (New York, N.Y.)*, *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*(5925), 400–403.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.).Hillside, NJ: Erlbaum

Department for Education. (2019). *Schools, pupils and their characteristics.* (June), 1–15. Retrieved from www.nationalarchives.gov.uk/doc/open-government-licence/version/3

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*(2), 149–164.

Gillborn, D., Demack, S., Rollock, N., & Warmington, P. (2017). Moving the goalposts: Education policy and 25 years of the Black/White achievement gap. *British Educational Research Journal*, *43*(5), 848–874.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, *24*(6), 645–662.

Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*(2), 213–228.

Graham, B., White, C., Edwards, A., Potter, S., & Street, C. (2019). *School exclusion: a literature review on the continued disproportionate exclusion of certain children*. Department for Education.

Hamilton, P & Morgan, G (2018) An exploration of the factors that lead to the successful progression of students in alternative provision. Educational and Child Psychology 35(1), 80-95.

Hanselman, P., Bruch, S. K., Gamoran, A., & Borman, G. D. (2014). Threat in Context: School Moderation of the Impact of Social Identity Threat on Racial/Ethnic Achievement Gaps. *Sociology of Education*, *87*(2), 106–124.

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), 405–424.

Higher Education Statistics Agency (HESA) (2017) Equality in higher education: students statistical report. Equality Challenge Unit.

Human Resources Research Organization. (2016). Independent Evaluation of the Validity and Reliability of STAAR Grades 3-8 Assessment Scores : Part 2 Final Report. Texas Education Agency.

Hutchison, D., & Styles, B. (2010). *A guide to running randomised controlled trials for educational researchers*. Slough: NFER. 1–65.

Jensen, A. R. (1972). Genetics and education. New York: Harper & Row.

Ladson-Billings, G. (2006). From the Achievement Gap to the Education Debt: Understanding Achievement in U.S. Schools. *Educational Researcher*, *35*(7), 3–12.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis. Applied social research methods series*. Sage Publications, Inc.

Maslow, A.H. (1943). "A theory of human motivation". *Psychological Review*. 50 (4): 370–96.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., … Urbina, S. (1996). Intelligence: Knowns and Unknowns. *American Psychologist*, *51*(2), 77–101.

Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PLoS ONE*, *11*(1), 1–25.

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiol Community Health*, *57*(January 2009), 527–529.

Rampton, A. (1981). *West Indian Children in our Schools*. London: HMSO.

Rollock, N., Gillborn, D., Vincent, C. & Ball, S. J. (2015) *The colour of class: The educational strategies of the Black middle classes* (London, Routledge).

Rosenthal, H. E. S., & Crisp, R. J. (2006). Reducing stereotype threat by blurring intergroup boundaries. *Personality and Social Psychology Bulletin*, *32*(4), 501–511.

Selm, K. R., Peterson, M.N, Hess, G. R., Beck, S. M., & McHale, M. R. (2019). Educational attainment predicts negative perceptions women have of their own climate change knowledge. *PLoS ONE*, *14*(1), 1–10.

Shapiro, J. R., Williams, A. M., & Hambarchyan, M. (2013). Are all interventions created equal? A multi-threat approach to tailoring stereotype threat interventions. *Journal of Personality and Social Psychology*, *104*(2), 277–288.

Shnabel, N., Purdie-Vaughns, V., Cook, J. E., Garcia, J., & Cohen, G. L. (2013). Demystifying Values-Affirmation Interventions: Writing About Social Belonging Is a Key to Buffering Against Identity Threat. *Personality and Social Psychology Bulletin*, *39*(5), 663–676.

Siegal, J & Anderson, C.S (1991) Considerations in Calculating High School GPA and Rank-in-Class. A Case Study. NASSP Bulletin. 96-109.

Steele, C. M; Aronson, J. (1995). Stereotype Threat and the Intellectual Task Performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811.

Taylor, V. J., & Walton, G. M. (2011). Stereotype threat undermines academic learning. *Personality and Social Psychology Bulletin*, *37*(8), 1055–1067.

Timpson, E (2019) Timpson Review of School Exclusion. Retreived from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/807862/Timpson_review.pdf

U.S. Department of Education (2002) Office of Elementary and Secondary Education, No Child Left Behind: Washington, D.C.

Appendix A: Tables of studies considered during the selection process

Table 7.

List of Studies Included in the Review

Full reference
Bancroft, A., Bratter, J., & Rowley, K. (2017). Affirmation effects on math scores: The importance of high school track. *Social Science Research*, 64, 319–333.
Binning, K.R., Wang, M.T, & Amemiya, J. (2019). Persistence Mindset among Adolescents: Who Benefits from the Message that Academic Struggles are Normal and Temporary? *Journal of Youth and Adolescence*, 48(2), 269–286.
Bowen, N. K., Wegmann, K. M., & Webber, K. C. (2013). Enhancing a Brief Writing Intervention to Combat Stereotype Threat among Middle-School Students. *Journal of Educational Psychology*, 105(2), 427–435.
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. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: a social-psychological intervention. *Science (*New York, N.Y.), 313(5791), 1307-1310.
Shnabel, N., Purdie-Vaughns, V., Cook, J. E., Garcia, J., & Cohen, G. L. (2013). Demystifying Values-Affirmation Interventions: Writing About Social Belonging Is a Key to Buffering Against Identity Threat. *Personality and Social Psychology Bulletin*, 39(5), 663–676.

Table 8.

Excluded Studies after Title and Abstract Screening

	Reference	Exclusion and Inclusion Criteria Code
1.	Alesi, M., Rappo, G., & Pepi, A. (2016). Investigating the improvement of decoding abilities and working memory in children with incremental or entity personal conceptions of intelligence: Two case reports. *Frontiers in Psychology,* 6, 1939.	3= Outcome measure was non-academic and focused on working memory.
2	Aronson, J., Cohen, G., McColskey, W., & (ED), R. E. L. S. (2009). Reducing Stereotype Threat in Classrooms: A Review of Social-Psychological Intervention Studies on Improving the Achievement of Black Students. *Issues & Answers.* REL 2009-076. In Regional Educational Laboratory Southeast.	4= A review
3	Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology,* 38(2), 113–125.	1= Conducted with college/university students
4	Baker, D. J., Skinner, B. T., & Redding, C. H. (2019). Affirmative Intervention to Reduce Stereotype Threat Bias: Experimental Evidence from a Community College. *Journal of Higher Education, 1-33.*	1= Conducted with college/university students
5	Barber, S. J. (2017). An examination of age-based stereotype threat about cognitive decline: Implications for stereotype-threat research and theory development. *Perspectives on Psychological Science*, 12(1), 62–90.	1= Conducted with a sample who are not of school age.
6	Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education: An International Journal*, 15(4), 427–448.	1= Conduced with college/university students .
7	Beckmann, N., Wood, R. E., Minbashian, A., & Tabernero Carmen (2012). Small group learning: Do group members' implicit theories of ability make a difference? *Learning and Individual Differences,* 22(5), 624–631.	2= Interventions are not self-affirmation exercises.
8	Berjot, S., & Gillet, N. (2011). Stress and coping with discrimination and stigmatization. *Frontiers in Psychology*, 2, 33 1-13.	4= A review
9	Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development*, 78(1), 246–263.	1= Sample does not focus on Black students

10	Brass, J., & Niu, W. (2016). The impact of implicit self-theory on rated creativity in third- and sixth-graders. *Creativity Research Journal*, 28(3), 318–327.	1= Sample does not focus on Black students
11	Burgess, D. J., Taylor, B. C., Phelan, S., Spoont, M., van Ryn, M., Hausmann, L. R, Gordon, H. S. (2014). A brief self-affirmation study to improve the experience of minority patients. *Applied Psychology. Health and Well-Being,* 6(2), 135-150.	1= Sample does not focus on Black students
12	Casad, B. J., & Bryant, W. J. (2016). Addressing stereotype threat is critical to diversity and inclusion in organizational psychology. *Frontiers in Psychology*, 7 (8),1-18	4= Non-empirical publication
13	Chalabaev, A., Major, B., Sarrazin, P., & Cury Philippe(2012). When avoiding failure improves performance: Stereotype threat and the impact of performance goals. *Motivation and Emotion*, 36(2), 130–142.	1= Sample does not focus on Black students
14	Chen, W.-W., & Wong, Y.-L. (2014). What my parents make me believe in learning: The role of filial piety in Hong Kong students' motivation and academic achievement. *International Journal of Psychology*, 49(4), 249–256	1= Sample does not focus on Black students
15	Cohen, G. L., & Garcia, J. (2008). Identity, belonging, and achievement: A model, interventions, implications. *Current Directions in Psychological Science,* 17(6), 365–369.	4= Non-empirical publication.
16	Cortland, C. I., Shapiro, J. R., Guzman, I. Y., & Ray, L. A. (n.d.). The ironic effects of stigmatizing smoking: combining stereotype threat theory with behavioral pharmacology. *Addiction* (Abingdon, England), 14, (10), 1842-1848	1= Sample is not focused on Black students
17	Covarrubias, R., Herrmann, S. D., & Fryberg, S. A. (2016). Affirming the Interdependent Self: Implications for Latino Student Performance. *Basic and Applied Social Psychology*, 38(1), 47–57.	1= Sample is not focused on Black students.
18	de Jong, E. M., Jellesma, F. C., Koomen, H. M. Y., & de Jong, P. F. (2016). A values-affirmation intervention does not benefit negatively stereotyped immigrant students in the Netherlands. *Frontiers in Psychology*, 7, 691, 1-20.	1= Sample does not focus on Black students
19	Dennehy, T. C., Smith, J. S., Moore, C., & Dasgupta, N. (2018). Stereotype threat and stereotype inoculation for underrepresented students in the first year of college. *The First Year of College: Research, Theory, and Practice on Improving the Student Experience and Increasing Retention.* 309–344.	1= Sample includes college/university students
20	Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence.	4= A review

	Journal of Research on Adolescence, 21(1), 225–241.	
21	Feltz, D. L., Schneider, R., Hwang, S., & Skogsberg, N. J. (2013). Predictors of collegiate student-athletes' susceptibility to stereotype threat. *Journal of College Student Development*, 54(2), 184–201.	1= Sample includes college/university students
22	Froehlich, L., Martiny, S. E., Deaux, K., Goetz, T., & Mok, S. Y. (2016). Being smart or getting smarter: Implicit theory of intelligence moderates stereotype threat and stereotype lift effects. *The British Journal of Social Psychology*, 55(3), 564–587.	1= Sample does not focus on Black students
23	Gelbgiser, D., & Alon, S. (2016). Math-oriented fields of study and the race gap in graduation likelihoods at elite colleges. *Social Science Research*, 58, 150–164.	1= Sample does not focus on Black students of school age.
24	Goyer, J. P., Cohen, G. L., Cook, J. E., Master, A., Apfel, N., Lee, W., … Walton, G. M. (2019). Targeted identity-safety interventions cause lasting reductions in discipline citations among negatively stereotyped boys. *Journal of Personality and Social Psychology,* 117(2), 229–259.	3= The outcomes are non-academic
25	Griffin, W. (2017). Who is Whistling Vivaldi? How Black football players engage with stereotype threats in college. *International Journal of Qualitative Studies in Education,* 30(4), 354–369.	4= Non-experimental study
26	Guzman, L., Goto, S. G., & Wei, K. (2016). Self-control depletion in predominantly white institutions: Intra and intergroup variability in the relations among stigma sensitivity, mental health, and academic motivation. *Journal of Social and Clinical Psychology*, 35(9), 754–780.	4= Non-experimental study (correlational design).
27	Hernandez, P. R., Schultz, P. W., Estrada, M., Woodcock, A., & Chance, R. C. (2013). Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. *Journal of Educational Psychology,* 105(1), 89–107.	1= Sample contains college/university students
28	Herrmann, S. D., Adelman, R. M., Bodford, J. E., Graudejus, O., Okun, M. A., & Kwan, V. S. Y. (2016). The effects of a female role model on academic performance and persistence of women in STEM courses. *Basic and Applied Social Psychology,* 38(5), 258–268.	1= Sample does not focus on Black students.
29	Ikizer, E. G., & Blanton, H. (2016). Media Coverage of "Wise" Interventions Can Reduce Concern for the Disadvantaged. *Journal of Experimental Psychology: Applied,* 22(2), 135–147.	2= The intervention is not focused on self-affirmation exercises.
30	Inzlicht, M., Aronson, J., & Mendoza-Denton, R. (2009). On being the target of prejudice: Educational implications. *Coping with Minority*	4= Not an empirical study

	Status: Responses to Exclusion and Inclusion. 13–37.	
31	Isik, U., Tahir, O. El, Meeter, M., Heymans, M. W., Jansma, E. P., Croiset, G., & Kusurkar, R. A. (2018). Factors Influencing Academic Motivation of Ethnic Minority Students: A Review. *SAGE Open*, 8(2), 1-23	4= A review
32	Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, 16(3), 175-179.	1= Sample does not focus on Black students
33	Johnson-Ahorlu, R. N. (2013). "Our Biggest Challenge Is Stereotypes": Understanding Stereotype Threat and the Academic Experiences of African American Undergraduates. *Journal of Negro Education,* 82(4), 382–392.	4= Not an empirical study
34	Jordt, H., Eddy, S. L., Brazil, R., Lau, I., Mann, C., Brownell, S. E. & Freeman, S. (2017). Values affirmation intervention reduces achievement gap between underrepresented minority and white students in introductory biology classes. *CBE Life Sciences Education*, 16(3), 1-10.	1= Sample includes college/university students.
37	Kolb, A. Y., & Kolb, D. A. (2009). The learning way: Meta-cognitive aspects of experiential learning. *Simulation & Gaming*, 40(3), 297–327.	4= Non-experimental study
38	Massey, D. S., & Probasco, L. (2010). Divergent streams: Race-Gender Achievement Gaps at Selective Colleges and Universities. *Du Bois Review,* 7(1), 219–246	1= Sample includes college/university students
39	Matthes, B., & Stoeger, H. (2018). Influence of parents' implicit theories about ability on parents' learning-related behaviours, children's implicit theories, and children's academic achievement. *Contemporary Educational Psychology,* 54, 271–280.	1= Sample does not focus on Black students
40	McGee, E. O., & Pearman, F. A. I. I. (2014). Risk and protective factors in mathematically talented Black male students: Snapshots from kindergarten through eighth grade. *Urban Education,* 49(4), 363–393.	4= Non-experimental study
41	Mendoza-Denton, R., Pietrzak, J., & Downey, G. (2008). Distinguishing Institutional Identification From Academic Goal Pursuit: Interactive Effects of Ethnic Identification and Race-Based Rejection Sensitivity. *Journal of Personality and Social Psychology,* 95(2), 338–351.	1= Sample are not of school age
42	Micari, M., & Pazos, P. (2014). Worrying about what others think: A social-comparison concern intervention in small learning groups. *Active Learning in Higher Education*, 15(3), 249–262.	1= Sample does not focus on Black students

43	Nadler, D. R., & Komarraju, M. (2016). Negating stereotype threat: Autonomy support and academic identification boost performance of African American college students. *Journal of College Student Development,* 57(6), 667–679.	1= Sample includes college/university students
44	Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314–1334.	4= A meta-analysis
45	Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, 67(2), 130–159.	4= A review
46	Owens, J., & Massey, D. S. (2011). Stereotype threat and college academic performance: A latent variables approach. *Social Science Research*, 40(1), 150–166.	1= Sample includes college/university students
47	Petticrew, T. F. (2010). Commentary: South African contributions to the study of intergroup relations. *Journal of Social Issues,* 66(2), 417–430.	4= A commentary
48	Rattan, A., Savani, K., Komarraju, M., Morrison, M. M., Boggs, C., & Ambady, N. (2018). Meta-lay theories of scientific potential drive underrepresented students' sense of belonging to Science, Technology, Engineering, and Mathematics (STEM). *Journal of Personality and Social Psychology,* 115(1), 54–75.	1= Sample does not focus on Black students
49	Renaud-Dube, A., Guay, F., Talbot, D., Taylor, G., & Koestner, R. (2015). The relations between implicit intelligence beliefs, autonomous academic motivation, and school persistence intentions: A mediation model. *Social Psychology of Education: An International Journal*, 18(2), 255–272	1= Sample does not focus on Black students
50	Rydell, R. J., & Boucher, K. L. (2017). Stereotype threat and learning. *Advances in Experimental Social Psychology,* 81–129.	4= Non-experimental publication
51	Salles, A., Mueller, C. M., & Cohen, G. L. (2016). A Values Affirmation Intervention to Improve Female Residents' Surgical Performance. *Journal of Graduate Medical Education,* 8(3), 378-383.	1= Sample is focused on college/university students
52	Schleider, J. L., & Weisz, J. R. (2017). Can less be more? The promise (and perils) of single-	4= Non-experimental publication

	session youth mental health interventions. *The Behavior Therapist*, 40(7), 256–261.	
53	Schmader, T., & Sedikides, C. (2018). State authenticity as fit to environment: The implications of social identity for fit, authenticity, and self-segregation. *Personality and Social Psychology Review*, 22(3), 228–259.	4= Non-experimental publication
54	Shuls, J. V. (2018). Raising the Bar on Teacher Quality: Assessing the Impact of Increasing Licensure Exam Cut-Scores. *Educational Policy,* 32(7), 969–992.	1= Sample does not focus on Black students
55	Smith, L. V, & Cokley Leann V (2016). Stereotype threat vulnerability: A psychometric investigation of the Social Identities and Attitudes Scale. *Measurement and Evaluation in Counseling and Development*, 49(2), 145–162.	4= No intervention pertaining to self-affirmation values
56	Steele, C. M. (1997). A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance. *American Psychologist*, 52(6), 613–629.	4= Non-experimental publication
57	Steele-Johnson, D., & Leas, K. (2013). Importance of race, gender, and personality in predicting academic performance. *Journal of Applied Social Psychology,* 43(8), 1736–1744.	1= Sample does not focus on school aged Black students
58	Stricker, L. J., & Ward, W. C. (2008). Stereotype threat in applied settings re-examined: A reply. *Journal of Applied Social Psychology,* 38(6), 1656–1663.	4= No intervention included.
59	Taylor, E., & Antony, J. S. (2000). Stereotype threat reduction and wise schooling: Towards the successful socialization of African American doctoral students in education. *Journal of Negro Education*, 69(3), 184–198.	1= Sample does not focus on school-aged Black students
60	Taylor, E., Guy-Walls, P., Wilkerson, P., & Addae, R. (2019). The Historical Perspectives of Stereotypes on African-American Males. *Journal of Human Rights and Social Work,* 4(3), 213–225.	4= Non-experimental publication
61	Teherani, A., Hauer, K. E., Fernandez, A., King, T. E., & Lucey, C. (2018). How small differences in assessed clinical performance amplify to large differences in grades and awards: A cascade with serious consequences for students underrepresented in medicine. *Academic Medicine*, Vol. 93, pp. 1286–1292.	1= Samples does not focus on school-aged Black students
62	Torok, L., Szabo, Z. P., & Toth, L. (2018). A critical review of the literature on academic self-handicapping: Theory, manifestations, prevention and measurement. *Social Psychology of*	4= Non-experimental publication

	Education: An International Journal, 21(5), 1175–1202.	
63	Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology,* 92(1), 82–96.	1= Sample does not focus on school-aged Black students
64	Walton, G. M., Spencer, S. J., & Erman, S. (2013). Affirmative meritocracy. *Social Issues and Policy Review,* 7(1), 1–35.	4= Non-experimental publication
65	Wasserberg, M. J. (2014). Stereotype Threat Effects on African American Children in an Urban Elementary School. *Journal of Experimental Education*, 82(4), 502–517.	4= No intervention included
66	Wasserberg, M. J. (2017). Stereotype threat effects on African American and Latina/o Elementary students tested together. *Journal for Multicultural Education,* 11(1), 51–60.	4= No intervention included
67	Whaley, A. L. (2018). Advances in stereotype threat research on African Americans: continuing challenges to the validity of its role in the achievement gap. *Social Psychology of Education*, Vol. 21, pp. 111–137.	4= Non-experimental publication
67	Wilson, Z. S., Holmes, L., deGravelles, K., Sylvain, M. R., Batiste, L., Johnson, M. & Warner, I. M. (2012). Hierarchical Mentoring: A Transformative Strategy for Improving Diversity and Retention in Undergraduate STEM Disciplines. *Journal of Science Education and Technology*, 21(1), 148–156.	4= Non-experimental publication
68	Worrell, F. C. (2014). Theories school psychologists should know: Culture and academic achievement. *Special Issue: Culture Counts*., 51(4), 332–347.	4= Non-experimental publication
69	Zhao, J., & Wang, M. (2014). Mothers' academic involvement and children's achievement: Children's theory of intelligence as a mediator. *Learning and Individual Differences*, 35, 130–136.	1= Sample does not focus on Black students

Table 9.

Studies Excluded at Full Text Review

	Reference	Exclusion and Inclusion Criteria Code
1	Bécares, L., & Priest, N. (2015). Understanding the influence of race/ethnicity, gender, and class on inequalities in academic and non-academic outcomes among Eighth-Grade Students: Findings from an intersectionality approach. *PLoS ONE*, 10(10), 1-17/	4= Non-experimental publication
2	Burnette, J. L., Russell, M. V., Hoyt, C. L., Orvidas, K., & Widman, L. (2018). An online growth mindset intervention in a sample of rural adolescent girls. *British Journal of Educational Psychology*, 88(3), 428–445.	1= Sample does not focus on Black students
3	Cole, B., Matheson, K., & Anisman, H. (2007). The moderating role of ethnic identity and social support on relations between well-being and academic performance. *Journal of Applied Social Psychology*, 37(3), 592–615.	3= Outcomes are not focused on measures of academic achievement.
4	Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, 24(6), 645–662.	2= Intervention is not focused on self-affirmation exercises
5	Goyer, J. P., Garcia, J., Purdie-Vaughns, V., Binning, K. R., Cook, J. E., Reeves, S. L. & Cohen, G. L. (2017). Self-affirmation facilitates minority middle schoolers' progress along college trajectories. *Proceedings of the National Academy of Sciences of the United States of America*, 114(29), 7594–7599.	3= Outcomes are not focused on measures of academic achievement.
6	Kappes, H. B., Stephens, E. J., & Oettingen, G. (2011). Implicit theories moderate the relation of positive future fantasies to academic outcomes. *Journal of Research in Personality*, 45(3), 269–278.	1= Sample does not focus on Black students
7	Powers, J. T., Cook, J. E., Purdie-Vaughns, V., Garcia, J., Apfel, N., & Cohen, G. L. (2016). Changing Environments by Changing Individuals: The Emergent Effects of Psychological Intervention. *Psychological Science*, 27(2), 150–160.	3= Outcomes are not focused on measures of academic achievement

8	Rosenthal, H. E. S., Crisp, R. J., & Suen, M. W. (2007). Improving performance expectancies in Stereotypic domains: Task relevance and the reduction of stereotype threat. *European Journal of Social Psychology,* 37(3), 586–597.	1= Sample does not focus on Black students
9	Rozek, C. S., Hanselman, P., Feldman, R. C., Quast, E. A., Crawford, E. P., Borman, G. D., & (SREE), S. for R. on E. E. (2015). Inside the Black Box of Self-Affirmation: Which Parts of Affirmation Exercises Are Critical for Treatment Efficacy? *Society for Research on Educational Effectiveness.* 1-4.	4= The intervention focuses on compliance to the intervention rather than the intervention itself.
10	Schmidt, A., Canela, C., & (SREE), S. for R. on E. E. (2015). The Behavioral Outcomes of a Self-Affirmation Intervention for Middle School Students. *Society for Research on Educational Effectiveness.*1-4	3= Outcome measures are not focused on academic achievement
11	Shapiro, J. R., Williams, A. M., & Hambarchyan, M. (2013). Are all interventions created equal? A multi-threat approach to tailoring stereotype threat interventions. *Journal of Personality and Social Psychology,* 104(2), 277-288.	1= Sample does not focus on school-aged Black students
12	Shelvin, K. H., Rivadeneyra, R., & Zimmerman, C. (2014). Stereotype threat in African American children: The role of Black identity and stereotype awareness. *Revue Internationale de Psychologie Sociale*, 27(3–4), 175–204.	2= No intervention included
13	Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S. & 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.	1= Sample does not focus on Black students.
14	Talbert, E., Hofkens, T., & Wang, M. Te. (2019). Does student-centered instruction engage students differently? The moderation effect of student ethnicity. *Journal of Educational Research*, 112(3), 327–341.	2= The intervention is not focused on self-affirmation exercises
15	Tarbetsky, A. L., Collie, R. J., & Martin A.J (2016). The role of implicit theories of intelligence and ability in predicting achievement for indigenous (Aboriginal) Australian students. *Contemporary Educational Psychology*, 47, 61–71.	2= The intervention is not focused on self-affirmation exercises

16	Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, 331(6023), 1447–1451.	1= Sample does not focus on school-aged Black students
17	Weber, S., Kronberger, N., & Appel, M. (2018). Immigrant students' educational trajectories: The influence of cultural identity and stereotype threat. *Self and Identity*, 17(2), 211–235.	1= Sample does not focus on school-aged Black students
18	Yeager, D. S., Purdie-Vaughns, V., Garcia, J., Apfel, N., Brzustoski, P., Master, A., … Cohen, G. L. (2014). Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide. *Journal of Experimental Psychology: General*, 143(2), 804–824.	2= Intervention do not focus on self-affirmation values exercises

Table 10.

Studies Excluded After Examination of Method and Data Analysis

Reference	Exclusion and Inclusion Criteria Code
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.	4= The design has combined Black students with pupils of other ethnicities and analyses does not permit examination of individual effects.
Borman, G. D., Grigg, J., Rozek, C & Hanselman, P (2015). The Sustained Effects of a Brief Self-Affirmation Intervention on Students' Academic Outcomes across Middle and High School. In *Society for Research on Educational Effectiveness.* 1-11.	4= The design has combined Black students with pupils of other ethnicities and analyses does not permit examination of individual effects.
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(5925), 400–403.	4= Follow up study combining samples from another study included in the review
Hanselman, P., Bruch, S. K., Gamoran, A., & Borman, G. D. (2014). Threat in Context: School Moderation of the Impact of Social Identity Threat on Racial/Ethnic Achievement Gaps. *Sociology of Education*, 87(2), 106–124.	4= The design has combined Black students with pupils of other ethnicities and analyses does not permit examination of individual effects.
Protzko, J., & Aronson, J. (2016). Context Moderates Affirmation Effects on the Ethnic Achievement Gap. *Social Psychological and Personality Science*, 7(6), 500–507.	4= The design has combined Black students with pupils of other ethnicities and analyses does not permit examination of individual effects.

Appendix B: Weight of Evidence A

Coding Protocol:

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71(2), 149–164.

This protocol was adapted to also include the descriptive criteria from the published paper. In addition, I have tailored the questions by providing relevant examples of criteria that should be considered when exploring the concept of stereotype threat. Explanations of the adaptations are italicised.

Reference of the study:

Bancroft, A., Bratter, J., & Rowley, K. (2017). Affirmation effects on math scores: The importance of high school track. *Social Science Research*, 64, 319–333.

ESSENTIAL QUALITY INDICATORS		
A. Quality indicators for describing participants:		**Criteria satisfied?**
Criteria 1.	Was sufficient information provided to determine/confirm whether the participants demonstrated the disability (ies) or difficulties presented? *(For the purpose of this systematic literature review, the focus of this question will be refined to focus on the concept of belonging to an ethnic group susceptible to stereotype threat, as a presenting difficulty).*	***Overall Quality Indicator Met for criteria 1?*** ☒ Yes ☐ No ☐Partially met
Sub questions considered within the rating	Individuals included in the study met the requirements of the definition of susceptibility of stereotype threat presented?	☒ Yes ☐ No ☐ N/A ☐ Unknown/Unable to Code
	Additional demographic and socio-economic information provided? *(e.g. Ethnicity, Gender, School Demographics/Student Body, Free School Meals etc)*	☒ Yes ☐ No ☐ N/A ☐ Unknown/Unable to Code
	Group differences on salient variables and characteristics are presented?	☒Yes ☐ N/A ☐ Unknown/Unable to Code

Criteria 2.	Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?	***Overall Quality Indicator Met for Criteria 2?*** ☐Yes ☐ No ☒Partially met
Sub questions considered within the rating	Participants were randomly assigned to the two conditions?	☒Yes - *within each school* ☐ No ☐ N/A ☐ Unknown/Unable to Code
	Administrators of the intervention are randomly assigned to the two conditions?	☐Yes ☐No ☐N/A ☒Unknown/Unable to Code
	Participants are matched on salient variables or a stratified assignment procedure is employed?	☒Yes - *stratified by ethnicity due to unequal groups across schools.* ☐ No ☐ N/A ☐Unknown/Unable to Code
	Attrition rates between the intervention and the comparison group are not substantially different?	☐Yes ☐ No ☐ N/A ☒ Unknown/Unable to Code-*although the groups are similar in sample size*
Criteria 3.	Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions? *(e.g. all administrators are teachers or researchers, teachers taught the same subjects and whether expertise requirements for delivery of the intervention are met)*?	***Overall Quality Indicator Met for Criteria 3?*** ☒Yes - *all teachers were trained by the researchers* ☐ No ☐Partially met
B. Quality indicators for implementation of the Intervention and description of Comparison Conditions		
Criteria 4.	Was the intervention clearly described and specified?	***Overall Quality Indicator Met for Criteria 4?*** ☒Yes ☐ No

		☐Partially met
Sub questions considered within the rating	Enough information is provided about the intervention to allow for replication *(e.g. conceptual underpinnings, detailed instructional procedures, teachers actions and language, use of instructional material and what students are required to do and day)*?	☒Yes ☐ No ☐ N/A ☐ Unknown/Unable to Code
Criteria 5.	Was the fidelity of implementation described and assessed?	***Quality Indicator Met for Criteria 5?*** ☒Yes - *teachers were required to report on the success and failure, student enjoyment and suggestions for future implementation* ☐No ☐Partially met
Criteria 6.	Was the nature of services provided in comparison conditions described?	***Quality Indicator Met for Criteria 6?*** ☒Yes ☐ No ☐Partially met
C. Quality indicators for outcome measures		
Criteria 7.	Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance?	***Quality Indicator Met for Criteria 7?*** ☒ Yes - *two measures of maths skills were used – A PSAT maths exam and STAAR algebra exam.* ☐No ☐Partially met
Criteria 8.	Were outcomes for capturing the intervention's effect measured at the appropriate times?	***Quality Indicator Met for Criteria 8?*** ☒ Yes ☐No ☐ Partially met
D. Quality indicators for data analysis		
Criteria 9.	Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study?	***Quality Indicator Met for Criteria 9?*** ☒Yes ☐ No ☐Partially met

Criteria 10.	Did the research report include not only inferential statistics but also effect size calculations?	*Quality Indicator Met for Criteria 10?* ☐ Yes ☒No - *effect sizes not reported, but statistics to compute them are available* ☐Partially met
DESIRABLE QUALITY INDICATORS	
Criteria 11.	Was data available on attrition rates among intervention samples? Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?	*Overall Quality Indicator Met for Criteria 11?* ☐Yes ☐No ☐ N/A ☒Unknown/Unable to Code
Criteria 12.	Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures?	*Overall Quality Indicator Met for Criteria 12?* ☐ Yes ☒No ☐Partially met
Sub questions considered within the rating	Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?	☐Yes ☐ No ☐ N/A ☒Unknown/Unable to Code
Criteria 13.	Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?	*Overall Quality Indicator Met for Criteria 13?* ☒ Yes ☐ No ☐ N/A ☐ Unknown/Unable to Code
Criteria 14.	Was evidence of the criterion-related validity and construct validity of the measures provided?	*Overall Quality Indicator Met for Criteria 14?* ☐Yes ☒ No ☐ N/A ☐ Unknown/Unable to Code
Criteria 15.	Did the research team assess not only surface features of fidelity implementation (e.g. number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?	*Overall Quality Indicator Met for Criteria 15?* ☐Yes ☒ No ☐ N/A ☐ Unknown/Unable to Code
Criteria 16.	Was any documentation of the nature of instruction or series provided in comparison conditions?	*Overall Quality Indicator Met for Criteria 16?* ☒ Yes ☐No

		☐ N/A ☐ Unknown/Unable to Code
Criteria 17.	Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?	***Overall Quality Indicator Met for Criteria 17?*** ☐Yes ☒No ☐ N/A ☐ Unknown/Unable to Code
Criteria 18.	Were results presented in a clear, coherent fashion?	***Overall Quality Indicator Met for Criteria 18?*** ☒Yes ☐No ☐ N/A ☐ Unknown/Unable to Code

		1=Low Quality *(Less than 9 of the Essential Quality Indicators are met. Even if more than one Desirable Indicator is met, the study is still ranked as Low Quality)*	**2=Acceptable Quality** *(9 or more of the Essential Quality indicators are met and at least one, but not more than three of the Desirable Indicators)*	**3=High Quality** *(9 or more of the Essential Quality indicators met and at least 4 of the Desirable Indicators)*	**Overall Evidence Rating (1-3)**
Number of Essential Quality Indicators Met	**(8/10)**	X			1
Number of Desirable Quality Indicators Met	**(3/8)**				

Appendix C: Weight of Evidence B

Appropriateness of the study design to answering the review question on the effectiveness of values-affirmation interventions in raising achievement outcomes of Black students based on Petticrew and Roberts (2003) typology of evidence.

WOE B Rating and Criteria		

HIGH = 3
Randomized Control Trials (at student level) with:
    - An appropriate control group (which also contains Black students for direct comparison), undertaking a similar non self- affirmation writing task.
    - Measures collected for both pre and post values affirmation intervention and the control condition.
	MEDIUM = 2
	Quasi Experimental Study (Non-random assignment) with:
	- An appropriate control group
	- Pre and post measures of the self-affirmation values intervention and the control condition
	- Indication of assessing difference between groups before intervention occurs
		LOW = 1
		Qualitative, Single Case and Non-Experimental Designs as these do not allow for effectiveness of an intervention to be measured appropriately. In addition studies which have:
		– No pre-post measures
		- No control groups
		- Small sample sizes such as single case experimental designs.

Study	WOE B Rating given
Bancroft et al. (2017)	3 = HIGH
Binning et al. (2019)	3 = HIGH
Bowen et al. (2013)	3 = HIGH
Bratter et al. (2016)	3 = HIGH
Cohen et al. (2006)	3 = HIGH
Shnabel et al. (2013)	2.5 = SOMEWHAT HIGH *due to randomisation at the class level

Appendix D: Weight of Evidence C

Relevance of the focus of the study to the review question.

Criteria	WOE C Rating and Descriptor	Rationale
A) **Implementation of the intervention**	**3= HIGH** One clear self-affirmation values intervention writing task and a control writing task. The intervention occurs on more than one occasion throughout the school year **2= MEDIUM** One self-affirmation intervention implemented at one time point in the school year **1= LOW** More than one intervention implemented alongside self-affirmation interventions which are not individually reported upon.	In order to be able to evaluate the specific effectiveness of the self-affirmation values intervention, it cannot have been combined with other interventions. Whilst the literature has not developed a conclusive specification on how many sessions the self-affirmation intervention should occur for, some researchers have critiqued the confidence in a brief one-time intervention having dramatic effects upon academic achievement (Bowen et al. 2013). Thus, studies which employ the intervention at more than one time point will receive higher ratings.
B) **Sample and context of the school**	**3= HIGH** The sample contains students from schools where the composition of ethnic groups in the overall student body is heterogeneous. **2= MEDIUM** The sample contains students from a combination of schools which contribute to the overall sample of ethnic groups being heterogeneous **1= LOW** The sample is comprised of students from schools that are comprised of predominantly one main ethnic group and this is NOT combined with other samples for heterogeneity and/or the details on the student body of the school	Evidence in the analyses carried out by Bratter et al. (2016) has found that self-affirmation interventions can have reverse effects; demonstrating a trend of decreasing academic achievement for Black students in schools with a student body that is predominantly comprised of a majority of Black students. In reference to stereotype threat, Hanselman et al. (2014) evidence that experiences of stereotype threat most often occur in contexts where there exists a salience of a marginalised group. This may be most likely in school contexts whereby the student body is not predominantly comprised of Black students. Therefore, studies that report heterogeneity in

	are not documented in the research publication.	the sample of students will receive higher ratings.
C) Outcome measures	**3 = HIGH** A standardised global measure of academic achievement. **2 = MEDIUM** GPA or another measure of global academic achievement which are non-standardised. **1 = LOW** Only one specific measure of attainment (e.g. Maths outcomes/ English) - even if this is standardised.	Achievement is a global measure of a range of academic skills and this would specifically answer the review question. Specific measures are also useful, but limit the generalisability of the intervention to raise achievement as a whole. Measures such as GPA are non-standardised teacher-ratings. Whilst they comprise a range of academic skills they can fluctuate in reliability of measuring academic outcomes (Bacon & Bean, 2006). Thus, studies utilising global standardised measure of academic achievement will receive higher ratings.

Study	WOE C Rating	Averaged score to give WOE C final rating
Bancroft et al. (2017).	A=3 B=2 C=1	1.67
Binning et al. (2019).	A=1 B=2 C=2	1.67
Bowen et al. (2013).	A=2 B=1 C=1	1.33
Bratter et al. (2016).	A=3 B=2 C=3	2.67
Cohen et al. (2006)	A=2 B=1 C=2	1.67
Shnabel et al. (2013)	A=3 * *although they didn't use the data from all three interventions in their analyses* B=1 C=2	2