Abstract

This meta-analysis assessed the impact of values affirmation on the academic achievement of students under social identity threats in actual classrooms. After a systematic search yielded 58 relevant studies, multilevel analyses identified an overall affirmation effect for identity-threatened students (Hedges' $g = 0.15$), not for identity-nonthreatened students (Hedges' $g = 0.01$). Heterogeneity in the affirmation effect was moderate to high for identity-threatened students, with effect sizes associated with (1) a larger covariate-controlled achievement gap between nonthreatened and threatened students in the control condition, suggestive of psychological underperformance, (2) the availability of financial resources in school, (3) more distal performance outcomes, and (4) the presentation of values affirmation as a normal classroom activity rather than a research study or a non-normal classroom activity. Affirmation appears to work best when it is delivered as a normal classroom activity and where identity threat co-occurs with resources for improvement and time to await cumulative benefits.

Keywords: achievement gap, social identity threat, stereotype threat, values affirmation, self-affirmation, intervention

Introduction

Well-documented achievement gaps persist between members of disadvantaged social groups and their more advantaged peers. These gaps occur for students at virtually all educational levels. For example, fourth-grade White students outperformed both African American and Latino American students by 0.74 and 0.61 standard deviations in the 2017 National Assessment of Educational Progress test (NAEP), respectively (Musu-Gillette et al., 2017), and such racial achievement gaps persist through middle school, high school, and college (Barton & Coley, 2010; Lee, 2002). The gender achievement gap favoring males in Science, Technology, Engineering and Mathematics (STEM) is widely documented on math tests such as the math SAT (Fryer Jr. & Levitt, 2010), national tests such as the NAEP (McGraw et al., 2006), and international science tests such as the Programme for International Student Assessment (PISA; Stoet & Geary, 2013). Large achievement gaps have also been found between students eligible for free or reduced school meals – an indicator of economic disadvantage – and their more advantaged peers both in the US and in the UK (Domina et al., 2018; Easterbrook et al., 2021; Hobbs & Vignoles, 2007). Likewise, students who are the first in their families to attend college, another indicator of economic disadvantage, tend to perform worse in college and drop out in greater numbers (Harackiewicz et al., 2014; Sirin, 2005).

Certainly, systemic and structural factors explain most of these gaps, including school composition, socioeconomic disadvantage, systemic racism and sexism, and implicit forms of bias (Bohrnstedt et al., 2015; Duncan & Murnane, 2014; Reardon, 2018; Steele, 2005). Yet, there is another contributor "down on the ground" that plays out every day in the classroom: social identity threat (Steele, 1997; Steele et al., 2002). It provides a partial explanation for these gaps and refers to the psychological adaptation people make to the social fact of bias against their group. They become aware and apprehensive of the possibility that they could be judged negatively in the light of their social group or identity. Stereotype threat, a more specific version of social identity threat, occurs when people are specifically concerned about being seen in the light of a stereotype against their group. Both social identity threat and stereotype threat can impair students' performance due to the distraction and stress they cause in high-stakes, evaluative environments (Cohen & Sherman, 2014). In real-world educational settings, social identity threat is likely to be chronic rather than acute, with effects that accumulate over time (Cohen et al., 2006; Cohen & Sherman, 2014). Dealing with these threats repeatedly over time can be Sisyphean, with the resulting stress and cognitive depletion leading to a downward spiral in performance (Aronson et al., 2002; Cohen & Garcia, 2008).

Because social identity threat takes a psychological form, it is possible to alleviate it with strategies that address how students think and feel in the classroom. Specifically, strategies that assure students that the stereotype is not in play, or that they are valued here and now despite the widespread discrimination against their group, can prove effective. One such strategy is known as "self-affirmation." While many acts could be self-affirming (e.g., spending time with friends or attending religious services; Cohen & Sherman, 2014), the most studied experimental manipulation of self-affirmation is "values affirmation," a brief writing activity in which students write about cherished personal values that transcend the evaluative situation (Cohen & Sherman, 2014; Sherman & Cohen, 2006; Steele, 1988). Claude Steele and his colleagues pioneered this technique, showing the potential of brief moments of values reflection to alleviate ego-protective behavior (Steele, 1988). In a typical values affirmation activity, students write for about 10 minutes on their cherished values, such as relationships with friends and family or religion, and why they are important to them. Ideally, the activities are timed to key evaluative situations where students are likely to be under threat, such as before tests or at the start of the academic year, when they do not know if they will be seen as fully belonging. In this way, values affirmation brings to mind a broadened self-concept, against which a specific threat looms less large.

Building on more than a decade of laboratory research on self-affirmation (see Liu et al., 2020; McQueen & Klein, 2006), a group of researchers worked with teachers to deliver a series of values affirmation writing activities to middle schoolers in the US (Cohen et al. 2006, 2009). In two sequential randomized controlled trials, they found that the intervention lifted African American students' grades, closing the gap between them and their White peers. Later studies tested the effect of values affirmation in boosting the academic achievement of identity-threatened students from various disadvantaged social groups. For instance, Harackiewicz et al. (2014) found that values affirmation reduced the social-class achievement gap by raising the grades of first-generation US students taking an undergraduate biology course. Hadden et al. (2019) found that values affirmation raised the academic achievement of young adolescents low in socioeconomic status. Miyake et al. (2010) found that values affirmation raised the exam scores of women who worried relatively more about gender stereotypes. Another key discovery from past studies is that the effects of values affirmation can propagate over surprisingly long periods. Students receiving values affirmation interventions have been found to earn higher grades and enroll in more advanced courses over periods ranging from one to seven years (Cohen & Sherman, 2014; Goyer et al., 2017). How? Two mechanisms address how social psychological interventions such as values affirmation can persist and even grow over time (Cohen et al., 2009; Walton & Wilson, 2018). First, their effects can be recursive in nature (Cohen et al., 2009). Students may perform better, and performing better, feel more affirmed, and thus continue to perform better, in a repeating cycle. Second, the effects of the interventions can interact with other processes in the school. Performing better, affirmed students may be held to higher expectations by their teachers (Cohen & Sherman, 2014; Rosenthal & Jacobson, 1968). As more is demanded of them, they then continue to perform well because the affirmation's initial effects connect them with powerful currents in the social environment. For example, in one study, as a result of the affirmation, not only did African American middle schoolers earn higher grades, but they were also more likely to be assigned to a higher-level classroom track, which could in turn increase their likelihood of college admission (Goyer et al., 2017).

Despite the breadth of empirical evidence, no systematic review or meta-analysis has assessed the overall effectiveness of values affirmation in improving the real-world academic achievement of identity-threatened students in educational settings. Moreover, along with positive findings, mixed and null effects of values affirmation have been reported (Bratter et al., 2016; Dee, 2015; Protzko & Aronson, 2016).

The primary aim of the present meta-analytical review is to estimate the average treatment effect of values affirmation on students' academic achievement in actual educational settings by consolidating all the studies conducted to date. We also assess whether, as predicted, the effect of values affirmation is apparent for members of identity-threatened groups and not for members of identity-nonthreatened groups.

The second research aim is to explore whether values affirmation is heterogeneous or homogeneous in its effects, and, if heterogeneous, the variables that predict when, where, and for whom it yields the most benefit. In the first empirical paper (Cohen et al., 2006) and subsequent reviews (Cohen & Sherman, 2014; Sherman & Cohen, 2006), the effect of values affirmation was asserted to be not invariant but dependent on key conditions. The most straightforward condition is the identity-threatened group status of students. Thus, affirmation effects should be consistently evident among identity-threatened students and not evident among identity-nonthreatened students.

Beyond identity-threatening group status, other key variables may moderate the effect of values affirmation and other social-psychological interventions (Cohen et al., 2017). The moderators fall into four broad categories (see Borman, 2017; Easterbrook et al., 2021). The first category concerns the student-level moderators. Some students may experience more identity threats than others and thus benefit more from values affirmation (Aronson & Good, 2002). Group status is thus one key moderator to test. So is the degree of underperformance, as roughly reflected by the degree to which identity-threatened students perform more poorly than nonthreatened students in the control conditions, and continue to do so even when prior indicators of achievement are statistically controlled. Steele (1997) argued that underperformance is the telltale sign of the effect of disruptive psychological states, such as stereotype threat, in a setting. The second category of moderators concerns the implementation of values affirmation (Bradley et al., 2016) because studies may implement the writing activities in many different ways. For instance, the following implementation variables have been suggested to be especially important in past studies: whether the intervention is presented as a classroom activity from the teacher or as a research activity from outsiders (Smith et al., 2021), whether it is framed as a normal assignment or as something to help the students (Sherman et al., 2013), and whether the intervention has been adapted or added to as a means to ensure student engagement and age-appropriateness (de Jong et al., 2016). The third category of moderators concerns aspects of the social context. These include whether the school or classroom contains resources for learning and growth, such as financial endowments that permit the purchasing of strong teaching materials, books, and other scaffolds of learning (Cohen et al., 2006; Ferrer & Cohen, 2018; Sherman et al., 2021). In a meta-analysis of affirmation interventions in a health context, Ferrer and Cohen (2018) found evidence that affirmation's benefits were most pronounced in situations where people were under psychological threat (due to a health-risk) and had access to resources to improve their health behavior. The dual role of threat and opportunity may also moderate the effects of affirmation in an education context, complementing recent analogous findings from Yeager et al. (2019) showing similar conditions to moderate the effectiveness of a mindset intervention. The fourth and final category of moderators concerns study-specific procedural characteristics such as sample size, study location, and study year. Although there is little theoretical basis to expect this latter class of moderators to matter, they are worth testing on an exploratory basis.

Method

Criteria for Inclusion of Studies

To be eligible for inclusion, a study had to randomly allocate students or groups of students either to a values affirmation treatment group or to a control group. The student participants could be drawn from any grade level ranging from kindergarten to twelfth grade (primary, pre-secondary and secondary schools) to colleges, universities, and graduate schools (tertiary education) and other educational institutions. Empirically, the youngest students to have been studied were 6th graders (11-12 years old). To be included, studies had to meet several additional requirements beyond the random assignment to conditions:

First, the study had to test a values affirmation intervention, which required *three key elements*. First, students had to be presented with a list of personal values (*value menu*). Next, they had to be given a chance to write a brief essay (*writing task*) about their most important value, typically a value in a domain unrelated to school (*value domain*) (Harris & Napper, 2005; McQueen & Klein, 2006). Students might be encouraged to write about their own values or values relevant to their social groups (see Tibbetts et al., 2018). This requirement led to the exclusion of studies that did not use the standard values affirmation activity but an alternative self-enhancement intervention (e.g., relating the self to a role model; Hoffman & Kurtz-Costes, 2019). Also excluded were studies that combined values affirmation with other interventions because it was impossible to disentangle the unique effect of values affirmation in these studies (e.g., Goyer et al., 2019, Experiment 1).

Second, the study had to have a *control condition*, specifically one that was neutral, typically accomplished through a non-affirming writing activity. Most control conditions had students select unimportant values from a menu and then asked them to write about why those values might be important to someone else. Others had students write about another neutral topic, such as "everything you had eaten or drank in the past 48 hours" (Cohen et al., 2000) or "possible uses for a knife" (Harvey & Oswald, 2000). For a review of standard control and affirmation activities, see McQueen & Klein (2006).

The *primary outcome* was grade point average, or GPA. *Secondary outcomes* encompassed other official measures of academic achievement, such as standardized test scores, exam performance, classroom grade, and percentile rank in class. They also included less standardized measures of course persistence and attainment, such as school retention, course completion rates, teachers' evaluation of students' performance, and performance on assignments and quizzes. Disciplinary outcomes such as behavioral conduct (e.g., Binning et al., 2019) were not included because our focus was on academic achievement outcomes. If the same outcome was measured at multiple time points (e.g., quarterly GPA) and cumulatively (e.g., cumulative GPA), we extracted only the outcome that was most distal from the commencement of the intervention and thus most cumulative. This prevented using highly-correlated outcomes from the same study.

We placed virtually no restrictions on the *setting* where the intervention was implemented, though we did restrict our analysis to studies of actual classrooms or schools. Students could complete the writing activities in classrooms or through online links outside of the classroom as long as the activity was associated with a course. Because our focus was on the effect of values affirmation in actual learning settings, we excluded studies that brought participants to a psychology laboratory and studies in completely virtual classrooms (e.g., Kizilcec, Davis, et al., 2017; Kizilcec, Saltarelli, et al., 2017). All the studies included in the meta-analysis were thus true field experiments.

Search Strategies

Studies were identified through a database search of published articles in PsycArticles & PsycINFO, ERIC, Medline, Scopus, Web of Science, and ProQuest Dissertations & Theses Global with the following search terms: *(edu* or learn* or adolescent* or stud* or teach* or pupil* or child* or undergrad* or grad* or youth* or young* or class* or school* or college* or universit*) AND (self-affirm* or self affirm* or values-affirm* or values affirm* or interdependent self-affirm* or relational self-affirm* or familial self-affirm* or close others self-affirm*) AND (learn* or develop* or perform* or achiev* or abilit* or outcome* or improve** *not health)*[1]. The search was conducted in December 2018 and again in February 2020 to check for recent publications.

Unpublished reports and literature were captured using Google Scholar, the Grey Literature Report, and OpenGrey and by reaching out to experts in relevant research areas. Conference papers from the American Psychological Association (APA), the American Educational Research Association (AERA), and the Society for Personality and Social Psychology (SPSP) were searched by scraping their *listservs*. The comprehensive review of "wise interventions" provided by Walton and Wilson (2018), which included an extensive reference list for studies on values affirmation interventions, was also searched, as was the wise intervention website curated by Gregory Walton at https://www.wiseinterventions.org/.

Meta-Analytic Procedures

Effect Size Calculation

All effect sizes were calculated as Hedges' *g* using the Campbell Collaboration effect size calculator (Wilson, 2011), R's compute.es package (Del Re, 2013) or esc package (Lüdecke, 2017). Hedges' *g* is an effect size measure that corrects for biases resulting from the inclusion of studies with a small sample size. It is a similar measure to Cohen's *d*, a conventional effect size measure that tends to slightly overestimate the effect size when studies have small sample sizes (Hedges, 1981).

For most studies, we could calculate two types of effect sizes. The first was the *raw effect size* unadjusted for students' baseline performance or other pre-intervention variables. This was

[1] Wildcard characters (*) were used to search for words that could be spelled differently but had the same meaning. For example, *achiev** finds *achievement*, *achieving*, *achieve*, etc. We searched a wide range of keywords related to the different types of values affirmation and self-affirmation interventions. We specified "*not health*" to maximize the efficiency of our search: we wanted to exclude the large body of affirmation research focused on behavioral health change. We avoided "*not health**" because the wildcard term *health** might capture words like "healthy," which could plausibly be included in the education studies we wanted to feature in our meta-analysis.

calculated as the difference in post-intervention mean scores between the treatment and control group and dividing it by the raw standard deviation (SD). Where raw means were unavailable, the raw mean difference between the two conditions was estimated using the relevant unstandardized regression coefficient from the analytic model. Second, the *covariate-adjusted effect size* for each study was also calculated where possible. This effect size constituted a more precise estimate of the treatment effect than the raw one because baseline confounders were controlled and uncertainty reduced. Specifically, the covariate-adjusted effect size controlled for students' baseline performance and other pre-intervention variables. It was calculated the same way as the raw effect size, except that covariate-adjusted means or unstandardized regression coefficients were used to compute the numerator. We retained the raw SD as the denominator, rather than the covariate-adjusted SD, so that all effect sizes would be estimated relative to the actual variability in the sample (Wilson, 2011). In some studies, the covariate-adjusted effect size was reported and did not need to be computed (Borman et al., 2016; Bowen et al., 2013; Lokhande & Muller, 2019; Sherman et al., 2013; Silverman & Cohen, 2014). In the few cases where necessary statistics were not reported, we were able to reverse-engineer either the SD or the covariate-adjusted effect size from *t* or *F* statistics, binary proportion tables, or standardized regression coefficients (Lipsey & Wilson, 2001; Peterson & Brown, 2005; Rosenthal & DiMatteo, 2001).

Importantly, when later studies used the same dataset to repeat earlier analyses, we used only the analysis from the original studies to avoid "double-dipping" from the same datasets. Meanwhile, effect sizes for different outcomes assessed among the same participants were extracted and treated as interdependent. For instance, affirmation effects on both the difficulty level of course load and GPA were reported in Goyer et al. (2017, Study 1), but only the former was extracted because the latter was already reported in the original study and extracted from it (Sherman et al. 2013); additionally, the two outcomes were treated as interdependent in the meta-analysis.

Where more data were needed to calculate an effect size, we reached out to the relevant authors, who shared either analytic outputs or raw data. Also, if specific data points were reported in figures or plots rather than the text, we used WebPlotDigitizer (Rohatgi, 2020) to extract the relevant data.

The two types of affirmation effect sizes were calculated for both identity-threatened students and identity-nonthreatened students. Identity-threatened students were defined as those potentially experiencing social identity threat because of their group identity, often because of a negative stereotype targeted at them, which was almost always specified in the original report. Identity-nonthreatened students were defined as those unlikely to experience social identity threat, typically because they were not negatively stereotyped, which was again almost always specified in the original report. There were two methods to test these two effect sizes, and different studies used different methods. For some reports, the researchers provided an estimate of the affirmation effect by testing it separately for threatened students and nonthreatened students. For other reports, the researchers provided data that could be used to estimate the two effects from a regression model conducted on the full sample (see Appendix Table A1).

Five studies did not specify a targeted identity-threatened group and reported only the main affirmation effect for all sampled students and did not provide extractable information on whether the effect differed by students' group status (Churchill et al., 2018; De Clercq et al., 2019; Lauer et al., 2013; E. Peters et al., 2017; Rapa, 2016). Because we could not combine these effects with the pooled estimated for either threatened or nonthreatened students, we excluded them in the calculations of the average affirmation effect (see Appendix for the calculation of average affirmation effect among this small subset of studies). Alternative methods of specifying effect sizes did not alter the direction, magnitude, or statistical significance of the results reported in the main text (see Appendix).

Analytical Strategy

Project-Cluster-Level Models. First, we used the meta package in R (Schwarzer, 2007) to estimate the weighted average affirmation effect separately for threatened students and nonthreatened students at the level of project cluster, what we refer to as the "project-cluster-level" estimate. Effect sizes were weighted using the inverse of the variance to place higher weights on project clusters with greater precision (Borenstein et al., 2011). This project-cluster-level estimate differed from some meta-analytical reviews that estimate effect sizes at the study level because we thought it safer to assume that effect sizes from different experiments at the same school site (e.g., sequential experiments on different cohorts) were interdependent and they should be consolidated at the project cluster level rather than the study level (e.g., Cohen et al., 2006, 2009). Pooling effect sizes at the study level yielded similar results (see Appendix).

This project-cluster-level model addressed dependency among studies (Cheung & Chan, 2004) by estimating one aggregate effect size for each project cluster (Van den Noortgate et al., 2013). This "sample-wise" aggregating procedure we use is a common method to distill multiple effect sizes into a higher-level group (i.e., project cluster), especially when the outcomes generally reflect the same construct (i.e., academic achievement). However, because this procedure could lead to an underestimation of the degree of heterogeneity across groups (Cheung & Chan, 2004, 2008) and possibly biased estimates of the average affirmation effect (Pastor & Lazowski, 2018), we supplemented it with multilevel models.

Multilevel Models (MLM). We also conducted multilevel meta-analyses to model the nested organization of effect sizes by project clusters. Specifically, we built multilevel models using the metafor package in R (Viechtbauer & Viechtbauer, 2020) and estimated the average affirmation effect both for threatened and for nonthreatened students (Assink & Wibbelink, 2016) with effect sizes (Level 1) nested in project clusters (Level 2) (see Appendix for mathematical notations of these models). Random effects models were fitted because the assumption that all project clusters came from the same population was unlikely (Schwarzer et al., 2015).

Quantifying Uncertainty in Treatment Effect Estimates. For both the project-cluster-level and multilevel models, we calculated the confidence interval (*CI*) and the prediction interval (*PI*) of the average affirmation effect. The 95% *CI* estimated the possible range of the true average affirmation effect, accounting for uncertainty in its estimation. The 95% *PI* accounted both for the uncertainty of the estimates *and* for the between-cluster heterogeneity. It can be interpreted as the plausible range in the treatment effect we would expect to see for a single future project (IntHout et al., 2016).

Test for Heterogeneity. We tested for statistical heterogeneity among the effect sizes by calculating Cochrane's Q to quantify the "difference between the observed effect sizes and the fixed-effect model estimate" (Higgins & Thompson, 2002). Because the Q statistic increases with the number of project clusters, we supplemented it with the I^2 statistic, which is insensitive to the number of project clusters. Specifically, I^2 measures "the percentage of variability in the effect sizes which is not caused by sampling error" (Higgins & Thompson, 2002).

Analysis of Moderators. Theoretical reviews and empirical papers suggested several moderators of affirmation effectiveness. As discussed in the Introduction, these moderators pertain to characteristics of the students, implementation fidelity, conditions of the social context, and procedural characteristics of the study (Borman, 2017; Bradley et al., 2016; Easterbrook et al., 2021; Ferrer & Cohen, 2018; Sherman et al., 2021). We attempted to provide operational definitions for each of these moderators and code them in all studies. Our efforts are summarized in Table 1, which enumerates the coding criteria for each moderator we attempted to measure. Importantly, moderators related to student characteristics fall beyond the scope of this meta-analysis because they typically vary across students rather than studies or project clusters. Therefore, the only student-level moderators we were able to test were the type of threatened social identity and achievement gaps in the control group (see Table 1). All moderators were coded by two independent raters, and disagreements were reconciled in consultation with the first author.

Moderators could vary at three levels. First, they could vary between effect sizes *within* the same study, for example, as when a single experiment tested the affirmation effect on two groups with different threatening identity, thus providing variation in student characteristics. Second, they could vary between studies within the same project cluster, for example, as when two experiments from the same project used different sample sizes and thus provided variation in a procedural characteristic. Third, moderators could vary between project clusters, for example, as when different projects took place at schools with different resources and implemented the activities differently, thus providing variation in the condition of social context and implementation fidelity. Which level a moderator varied at was determined by examining its variation at different levels and treating the level with the most variation as the primary one (see Table 1).

The effect-size-level and project-cluster-level moderators fitted well with the main two-level MLM because each moderator could be tested as a predictor at its appropriate level. However, testing the effect of study-level moderators proved challenging. We chose to supplement our main analyses with what we thought the best solution: a three-level MLM that treated effect sizes as nested within studies and studies as nested within project clusters. These models should be viewed tentatively as they ended up being complex and introduced a risk of over-fitting the data (see Appendix for mathematical notations of these models).

Each moderator could be coded as either categorical (binary or more than two levels) or continuous. For the categorical variables, we report results only where at least five data points were available for each level (see Table 1 for data related to the number of studies or project clusters available at each level of these moderators). With fewer than five data points at a given level, the moderator analysis was likely to be underpowered (see Higgins & Thompson, 2004). We then built MLMs using the metafor package in R (Viechtbauer & Viechtbauer, 2020) to separately test the association between each moderator and the affirmation effects in studies or project clusters that offered sufficient data, and report below the unstandardized regression coefficients (*b*).

Assessment of Risk of Bias in Included Studies

The Cochrane Collaboration's tool (Higgins et al., 2011) was used to assess the presence of various sources of bias related to selection, performance, detection, attrition, reporting, and other factors. *Selection bias* addresses whether 1) random assignment to conditions was used to allow for comparable groups, and 2) allocation sequence was concealed from personnel so that intervention assignment could not have been foreseen before or during enrollment. *Performance bias* addresses whether students and personnel were blinded to the experimental condition that each student was in. *Detection bias* addresses whether people who assessed the outcome (e.g., the teachers) were blinded to the experimental condition that each student was in. *Attrition bias* addresses whether outcome assessment was incomplete for some participants as a function of condition and whether the reasons for such differential attrition were reported and strategies to deal with it were undertaken. *Reporting bias* addresses whether there was evidence for selective reporting of results, and *other bias* addresses any sources of bias that emerged in a study but was uncaptured by the categories noted above. Each type of bias was rated as "low," "high," or "unclear" by two raters, with any discrepancies reconciled through consultation with the first author. Any variable that had at least 5 cases per level was treated as a study-level moderator in MLM.

Publication bias was assessed in three ways. First, we created a funnel plot (Light & Pillemer, 1984) to detect if project clusters with less precision and smaller effect sizes were underrepresented in the dataset. If so, the published articles would not constitute a representative sample of the available evidence, leading to a bias towards a larger treatment effect. We also performed Egger's test of the intercept (Egger et al., 1997) to statistically quantify the asymmetry of the distribution in the observed effect sizes around the pooled effect sizes. Second, we calculated the Fail-Safe N (Rosenthal File Drawer Analysis; Rosenthal, 1979), the number of project clusters with null results that would need to be added to the pool of project clusters to reduce the overall significance level to null, i.e., to a target level of $\alpha = 0.05$. Because this method is criticized as overly conservative since typically a very large number of studies (project clusters in our case) are needed to render an effect nil, we also calculated the Orwin Fail-Safe N (Orwin, 1983). This specifies the number of project clusters with null effects that would need to be added to the pool to reduce the treatment effect to a specified smaller effect size, which we set to half of the observed effect size. Third, we plotted *p*-curves to detect potential over-representation of *p* values close to 0.05, a red flag for *p*-hacking (J. P. Simmons et al., 2011). We also performed right-skewness and flatness test to assess if the pool of available studies seemed to be missing evidential "true *p*-values" based on the *p*-curve (J. P. Simmons et al., 2011; J. P. Simmons & Simonsohn, 2017).

Results

The systematic search of the databases yielded 7647 articles. One additional journal article (Tibbetts et al., 2018) was recommended by experts; one institutional report (Borman, 2012) and three additional thesis articles (Gutmann, 2019; Schwalbe, 2018; C. M. Simmons, 2011) were identified through the search for unpublished reports; and four articles currently under review were included with the permission of the authors (Binning et al., under review; Purdie-Greenaway et al., under review; Serra-Garcia et al., 2020; Turetsky, 2020[2]). Full-text screening yielded 58 studies that met the inclusion criteria, and we were able to retrieve adequate statistics to calculate effect sizes for all studies. Figure 1 provides a graphical representation of the study selection process (Moher et al., 2009). An a priori power analysis for random-effects models showed that the statistical power was 0.99 for detecting an average effect size of 0.2 for 58 studies, with an alpha level of 0.05. These calculations assume 30 students in both treatment and control groups in each study and modest heterogeneity in the estimated effect sizes (Hedges & Pigott, 2001).

The study sample sizes ranged from 35 to 4482, and all but 9 studies were conducted in the US (see Tables 1 and 2 for descriptive information related to the included studies). Study duration ranged from 2 weeks to roughly 3 years; the number of interventions administered

[2] Serra-Garcia et al. (2020) and Turetsky (2020) were under review while the current meta-analysis was also under review.

ranged from 1 to 8, and the number of value domains offered to students ranged from 5 to 14. There were 30 studies that used GPA as the primary outcome, 37 studies that included a variety of secondary achievement outcomes such as class grade, test scores, and course enrollment, and 9 studies that reported both primary and secondary outcomes (see Table 2). Several clusters of effect sizes were specified so that MLMs would treat them as nested (see Table 3).

The final analytical sample contained 39 project clusters and 49 studies that had non-duplicative findings. Of these, 34 project clusters and 44 studies provided effects sizes for an identity-threatened group (see Appendix for analysis of the 5 studies that reported only the condition main effect for all students).

Overall Treatment Effect

Figure 2 displays the average affirmation effect among threatened students for each project cluster, using the average of the raw and covariate-adjusted affirmation effect where both were available (N = 18), or only one of the former or latter was (N = 21). The overall weighted average effect size was g = 0.15, 95% *CI* [0.06, 0.23], p = 0.001. This effect could be interpreted in standard deviation units, suggesting that treatment and control groups differed on average by 0.15 standard deviations. Although the effect size was small according to the conventional "rules of thumb" offered by J. Cohen (2013)[3], it should be considered a medium to large size in the context of the benchmarks for educational interventions (Bakker et al., 2019; Kraft, 2020)[4]. The Discussion section further addresses the statistical and pragmatic significance of this effect.

Heterogeneity in the affirmation effect among threatened students was moderate to high ($I^2 = 68.4\%$) according to the "rules of thumb" offered by Higgins et al. (2003)[5], such that

[3] Decontextualized benchmarks for general effect sizes: Small = 0.2, Medium = 0.5, Large = 0.8
[4] Empirical benchmarks for effect sizes of educational interventions: Small = 0.05, Medium = 0.15, Large = 0.2
[5] Low heterogeneity: $I^2 = 25\%$; Moderate heterogeneity: $I^2 = 50\%$; Substantial heterogeneity: $I^2 = 75\%$

68.4% of the variability between project clusters could not be ascribed to sampling error. Cochrane's Q test also suggested that affirmation effects were heterogeneous across project clusters ($Q(33) = 104.38$, $p < 0.001$). Therefore, a multilevel meta-analysis was necessary to account for the heterogeneity across project clusters.

The prediction interval suggested that we would expect a specific effect size from a single future study to fall within an interval containing zero (95% *PI* [-0.29, 0.58]). Thus, while the overall effect for threatened students was positive and significant, there was a level of heterogeneity in the observed effect sizes such that any *specific* future study could plausibly yield a null effect. Prediction intervals encompassing zero are not uncommon in psychological research (Harrer et al., 2019). This can occur when a treatment effect is present but conditional on other variables and thus heterogeneous in its manifestation.

The effect size was significant for threatened student both on the primary outcome of GPA ($g = 0.13$, 95% *CI* [0.01, 0.25], $p = 0.035$) and other secondary achievement outcomes ($g = 0.17$, 95% *CI* [0.07, 0.27], $p = 0.003$), again with moderate to high between-cluster heterogeneity (GPA: $I^2 = 40.6\%$; $Q(13) = 21.89$, $p = 0.057$; secondary outcomes: $I^2 = 72.8\%$; $Q(24) = 88.36$, $p < 0.001$).

Figure 3 shows the comparable affirmation effect sizes for nonthreatened students. The overall weighted mean effect size was small and not significantly different from zero: $g = 0.02$, 95% *CI* [-0.03, 0.06], 95% *PI* [-0.18, 0.21], $p = 0.456$. Tellingly, in contrast to the pattern for threatened students, the between-cluster heterogeneity was small ($I^2 = 17.6\%$; $Q(21) = 25.50$, $p = 0.226$).

MLMs with effect sizes treated as nested in project clusters showed similar results on overall achievement: *threatened student group*: $g = 0.15$, 95% *CI* [0.08, 0.23], $p < 0.001$;

nonthreatened student group: $g = 0.01$, 95% *CI* [-0.04, 0.06], $p = 0.635$. Furthermore, the average effect size was significant for threatened student both on the primary outcome of GPA ($g = 0.15$, 95% *CI* [0.03, 0.26], $p = 0.013$) and on secondary achievement outcomes ($g = 0.17$, 95% *CI* [0.08, 0.26], $p < 0.001$).

Multilevel meta-analysis specifically of the *covariate-adjusted* effect size—a more precise estimate—showed similar results with slightly larger estimates: *threatened student group*: $g = 0.18$, 95% *CI* [0.10, 0.27], $p < 0.001$; *nonthreatened student group*: $g = -0.01$, 95% *CI* [-0.06, 0.04], $p = 0.624$. Once again, the average adjusted effect size was significant for threatened student both on the primary outcome of GPA ($g = 0.17$, 95% *CI* [0.04, 0.30], $p = 0.012$) and on secondary achievement outcomes ($g = 0.21$, 95% *CI* [0.10, 0.32], $p < 0.001$).

Overall, these results suggest that values affirmation improved the academic achievement of threatened students in general, yielding a small effect by statistical standards and a medium to large effect by educational standards, and that this effect showed marked heterogeneity across project clusters. By contrast, the affirmation effect for nonthreatened students was nil and centered relatively tightly around zero.

Exploratory Moderator Analysis

MLMs suggested five variables that moderated the affirmation effect among threatened students.

First, the affirmation effect for threatened students was larger in project clusters characterized by a larger raw and the residual achievement gap between the nonthreatened and the threatened students in the control condition (*raw*: $b = 0.22$, $SE = 0.06$, 95% *CI* [0.10, 0.34], $p < 0.001$; *residual*: $b = 0.28$, $SE = 0.07$, 95% *CI* [0.13, 0.43], $p < 0.001$). Because the raw gap was available for only 18 project clusters and the residual gap for 17 project clusters, we also repeated the analysis using whichever datapoint was available for a given study and the residual gap where both were available. This increased our sample size to 23 project clusters and the same result was obtained ($b = 0.17$, $SE = 0.04$, 95% CI [0.10, 0.25], $p < 0.001$).

Second, the affirmation effect was larger for studies that had longer durations. Studies with longer durations, as measured in months, yielded larger affirmation effects ($b = 0.01$, $SE = 0.002$, 95% CI [0.003, 0.01], $p = 0.001$), even when controlling for the number of intervention doses given ($b = 0.01$, $SE = 0.003$, 95% CI [0.002, 0.01], $p = 0.006$).

Third, the affirmation effect was larger for project clusters that presented the affirmation as a normal classroom activity rather than as part of a research study or an activity separate from normal classroom activities ($b = 0.16$, $SE = 0.08$, 95% CI [0.01, 0.32], $p = 0.038$).

Fourth, the affirmation effect was marginally larger in studies conducted at schools where financial resources were more abundant. This moderator was computed as the inverse of the percentage of students eligible for free or reduced lunch at the primary or secondary school where the study was conducted (the one study undertaken in post-secondary settings that reported on this measure was excluded in this analysis). This variable served as a proxy for the degree of available financial resources. It marginally predicted larger effect sizes ($b = 0.59$, $SE = 0.34$, 95% CI [-0.08, 1.27], $p = 0.083$).

Fifth, larger effect sizes were found in studies that followed the original affirmation materials relatively closely rather than made changes to them (original materials reported in Cohen et al., 2006) ($b = 0.23$, $SE = 0.07$, 95% CI [0.09, 0.38], $p = 0.002$). Specifically, Bayly (2017) added a committed action component that asked students to write about behaviors that are consistent with their values, de Jong et al. (2016; Study 2) added an affirmation component in which the teaching assistants and the students had a conversation about an additional reading comprehension assignment before the values affirmation writing activity, Kim (2019) asked students to affirm values that they "shared with the group of importance," Rapa (2016) added a section that requested students to reflect on critical actions they could take to address inequality, Schwalbe (2018) asked students to reflect on positive, meaningful memories in one session following the standard values affirmation; he also tested a "nudged" values affirmation in the following session where the instruction in the value prompt used present participle rather than present tense (e.g. "write about values that have been most important" rather than "write about values that are most important"), Tibbetts et al. (2018; Study 1b) limited one affirmation condition to choose from a value menu with mostly interdependent values such as "belonging to a social group." This moderator, in particular, should be viewed tentatively. We do not know if some of these modifications may have diminished the effectiveness of the affirmation or if other confounders associated with the need for adaptation were at play.

Because Ferrer & Cohen (2018) had found an interaction in their meta-analysis between the existence of threat and the presence of resources, we decided, on an exploratory basis, to test the same interaction here, using the size of the residual achievement gap as a gauge for the degree of threat operant in the setting. Consistent with that earlier meta-analysis, we found a significant interaction effect between the two moderators (b = 1.58, SE = 0.75, 95% CI [0.04, 3.12], p = 0.044). As shown in Figure 4, the affirmation effect was strongest in those settings where the achievement gap was large and relatively more financial resources were available. Simple effects tests found that the average affirmation effect was large and significant when the context was one with a large residual gap and plentiful resources (g = 0.36, 95% CI [0.16, 0.56], p = 0.001), but not significantly different from zero when context was low on these variables (g = -0.17, 95% CI [-0.58, 0.24], p = 0.401).

Assessment of Risk of Bias

Figure 5 presents the distribution of different risks of bias. Overall, most studies had either low or unknown risk on all coded dimensions. Most studies had a low risk of selection bias (*Random sequence generation* and *Allocation concealment*). Most studies also had low performance bias and detection bias, as both participants and personnel were blind to the experimental condition (*Blinding of participants and personnel*) in most cases, as were outcome assessors (*Blinding of outcome assessment*). Seven studies (Baker et al., 2019; Bancroft et al., 2017; Bratter et al., 2016; Churchill et al., 2018; Kost-Smith et al., 2012; E. Peters et al., 2017; Rapa, 2016) had a high risk of attrition bias due to incomplete outcome data either overall or in one condition relative to another (*Incomplete outcome data*). Interestingly, these high-risk studies yielded a small and non-significant average affirmation effect (g = 0.05, 95% CI [-0.14, 0.24], p = 0.626), while the low-risk and the unclear-risk studies had medium and significant average affirmation effect (*low*: g = 0.11, 95% CI [0.03, 0.20], p = 0.012; *unclear*: g = 0.25, 95% CI [0.15, 0.34], p < 0.001). *Selective reporting* and *other sources of bias* were not detected or could not be clearly determined.

Publication Bias

Publication bias was first examined by assessing whether project clusters with less precision and smaller effects seemed underrepresented in the pool of project clusters (Sutton, 2009). Figure 6 displays a standard method for assessing this possibility: a funnel plot of the average affirmation effect among threatened student groups in each project cluster. Effect sizes were plotted on the x-axis, and the standard error of each estimated effect size (a proxy for imprecision and possibly small sample size) plotted on the y-axis on a *reversed* scale. We would expect project clusters to lie symmetrically around the pooled effect size (the vertical dotted line) in the absence of bias. Figure 6 shows that the distribution of project clusters was indeed symmetrical around the vertical line. Among project clusters with large uncertainty (i.e., higher standard errors), small effects (bottom left corner) did not seem to be underrepresented relative to large ones (bottom right corner). Results from the Egger's test (Egger et al., 1997) also confirmed a lack of asymmetry at the significance level $\alpha = 0.05$ ($b = 0.742$, 95% *CI* [-0.35, 1.83], $p = 0.193$), suggesting little evidence of publication bias. The contour lines were also added to this figure to provide a further visual aid to detect possible missing areas of statistical insignificance (J. L. Peters et al., 2008). However, no such area could be found on this plot, providing further evidence for lack of publication bias.

Next, publication bias was examined by calculating the Fail-Safe N using both the Rosenthal method and the Orwin method. The Rosenthal Fail-Safe N was 464 ($p < 0.001$), indicating that an implausibly large number of project clusters with null effects would be needed to nullify the average affirmation effect. The Orwin Fail-Safe N was lower, at 34, meaning that we would need to double the number of project clusters used in the analysis to halve the affirmation effect observed for threatened students.

Publication bias was also examined in terms of the distribution of significant *p*-values (*p-curve*) to detect evidence of "*p*-hacking" (see Figure 7). The right-skewness and flatness test suggested that the *p*-curve was significantly right-skewed ($p < 0.001$) and not flat ($p = 0.978$), each of which indicated the existence of a "true effect." Hence, the significant meta-analytic results could not plausibly be attributed to various forms of publication bias.

Discussion

This meta-analysis examined the effect of values affirmation on students' real-world academic achievement in educational institutions and showcased the quantity and range of field experiments on values affirmation that have been conducted since the first one (Cohen et al., 2006). Values affirmation has been tested as an intervention to help students under social identity threat in a wide variety of contexts, including middle schools and high schools, regular public schools and charter schools, college STEM classes, MBA classes, community colleges, music courses, and learning centers for the disabled.

The meta-analysis identified a significant average affirmation effect of 0.15 standard deviations for threatened students and, as expected, no effect for nonthreatened students. Importantly, this affirmation effect should be interpreted with respect to the empirical benchmarks for other educational interventions (Baird & Pane, 2019; Bakker et al., 2019; Kraft, 2020; Lipsey et al., 2012). The effect of values affirmation on threatened students' academic achievement should be considered medium to large according to the empirical guidelines for interpreting effect sizes for "PreK-12 education interventions evaluating effects on student achievement" (Kraft, 2020) and for brief psychological interventions on real-world educational outcomes (Hill et al., 2008; Yeager et al., 2019). These empirical benchmarks are recommended because the effect of other effective, larger-scale education interventions on standardized achievement tests have shown similar or smaller effects (d = 0.08 – 0.15; Lipsey et al., 2012).

The affirmation effect among threatened students was robust to whether covariate-adjusted or raw effect sizes were used and to whether GPA or other measures of institutional achievement such as enrollment in difficult courses was the focal outcome (gs = 0.15 – 0.21). However, the covariate-adjusted effect on the secondary outcomes was slightly larger, a large effect based on empirical benchmarks (g = 0.21). A positive effect size of almost any magnitude is arguably important given the fact that the intervention's cost in money and classroom time is very small. On the whole, the results suggest that values affirmation is a cost-effective strategy for closing achievement gaps between nonthreatened and threatened students.

Beyond discerning a signal for the affirmation effect, the meta-analysis identified heterogeneity in its manifestation among threatened students. Exploratory analyses suggested that the affirmation effect was not invariant but dependent on how it was implemented and where.

First, affirmation's benefit was greater when there was a larger achievement gap between nonthreatened and threatened students in the control condition. It seems plausible that the larger the achievement gap, the more likely that psychological factors depress the performance of threatened students. In particular, a larger residual achievement gap, the portion of the gap that remains when prior performance is controlled, is a classic symptom of underperformance due to stereotype threat (Steele, 1997).

Second, affirmation's benefit was marginally greater in schools with a higher concentration of financial resources, as roughly measured by the inverse of the proportion of students receiving free or reduced school meal or lunch. With more opportunities, students' aspirations are more likely to translate into success. Social-psychological interventions that support student psychology have long been posited to act not alone but in a catalytic relationship with the social context, activating unrealized resources (Cohen et al., 2017; Cohen & Sherman, 2014; Walton & Wilson, 2018; Walton & Yeager, 2020; Yeager et al., 2019).

Third, affirmation's benefit was greater in the studies that assessed outcomes more distal from the commencement of the intervention. When affirmed, students may benefit from a recursive cycle, in which success begets success, thus leading to benefits that are more discernible with potentially more distal and accumulative outcomes. Of course, at some point, perhaps with the transition to a new school, we would expect this relationship between temporal distance from the treatment and treatment impact to break down. But the positive recursive processes underlying values affirmation and other social-psychological interventions (Cohen et al., 2009) suggest that when the affirmation triggers initial benefits in a context supportive of student growth, its benefits may propagate and even grow through time.

Fourth, the benefit of affirmation was larger in studies that presented values affirmation as a regular classroom activity rather than as part of a research project or an activity separate from normal classroom activities. It is possible that when the intervention is delivered by the teacher as a regular classroom activity, it conveys to students that their teacher is interested in their values and regards them as a “whole person” (see also Smith et al., 2021) or that they can safely reflect on their values without stress (Bradley et al., 2016).

Fifth, affirmation effects were greater when the materials used were closer to the original materials. While this might be due to the fact that the studies where the materials were adapted featured more novel settings for testing affirmation, it also serves as a reminder that what practitioners see as improvements may not necessarily be so.

Finally, we also found evidence consistent with the “perfect anti-storm” of conditions posited to underlie the effects of values affirmation and other social-psychological interventions (Ferrer & Cohen, 2018; Goyer et al., 2017; Yeager et al., under review). In particular, values affirmation had synergistic effects when given in a context where a larger achievement gap was present and financial resources were high. It was the combination of threats and opportunities that predicted the most positive effects, a finding that meshes with results from Yeager et al. (under review). When students are under threat but in a resource-rich environment, they are especially helped by the seemingly small but potent act of reflecting on core values.

Disencumbered of threat, they can seize the resources available. Under these conditions, the affirmation effect was relatively large (g = 0.35), whereas it was closely clustered around zero when these conditions were absent. These results dovetail with the “3 T’s” of the Trigger and Channel Model, which predicts that social-psychological interventions will have stronger effects when they are targeted to people in need, tailored to their needs, and timed to occasions when resources for growth are available (Cohen et al., 2017; Cohen & Sherman, 2014; Ferrer & Cohen, 2018; Walton & Wilson, 2018).

By highlighting these moderators, the present meta-analysis provides insight into the mechanisms behind the effect of values affirmation through time, building on previous research (Cohen & Sherman, 2014; Sherman & Cohen, 2006). It is important to note, however, that these moderator analyses were exploratory, and the moderators were not allocated randomly to students or studies. Hence, any strong causal claims regarding their role in affirmation effects must be averted. For instance, it could be the case that our measure of financial resources moderated the affirmation effect because it was a proxy for whether students came from families that provided stronger academic preparation and thus whether they found the writing activity more intelligible or interesting (Manstead, 2018; Stephens et al., 2014). Also, the fact that many studies took pains to make their materials intelligible to low literacy students militated against this possibility (see also Sherman et al., 2013). Still, there is no avoiding the fundamental limitation in any exploratory and correlational analysis. We simply view these analyses as promising, consistent with the previous theorizing, and suggestive of future avenues of research.

Notably, several moderators did not have enough variation in our sample of studies to test, but they are still potentially important. For instance, most studies followed the guidelines to implement the intervention in a timely way (either early in the school year or before stressful events; Cook et al., 2012), ensured students could complete it in private (Cohen et al., 2006), presented the writing activity as nonevaluative, and minimized any mention of its being "beneficial" to students. The latter has been found to sometimes undermine the effect of affirmation by conveying that students are in need of help and thus potentially deepening social identity threat (Sherman et al., 2009). Also notable are the moderators that displayed variation in our sample of studies but showed no effect. For instance, it did not appear to matter how many times the intervention was given, its dosage (see also Cook et al., 2012). We suspect that timeliness is more important than dosage, consistent with previous work by Cook et al. (2012) because what is critical is that the affirmation interrupts a recursive cycle before its costs compound and become irreversible. Additional doses should be beneficial insofar as they are timed to occur with key stressors that might otherwise worsen or re-instantiate the downward spiral of performance that social identity threat can sometimes cause (Cohen & Sherman, 2014).

On the whole, the meta-analysis suggests that there is a medium to large affirmation benefit on the academic achievement of students experiencing social identity threat in classroom settings. However, this effect is not unitary and is dependent on context. Powerful but conditional seems an apt way to describe its effects (Ferrer & Cohen, 2018).

Some other considerations merit comment. First, few studies reported a power analysis, a limitation that future replications should avoid by referring to the average affirmation effect found in our meta-analysis (g = 0.15 overall, g = 0.25 in middle schools, and g = 0.13 in universities; see the Open Science Framework project page). Second, future studies should consistently report both covariate-adjusted and raw effect sizes. Covariate adjustment has some benefits and costs. In terms of benefits, including covariates, especially baseline performance, reduces error, and increases precision. It also increases statistical power, comparable to the effect of increasing the sample size. On the other hand, covariate adjustment can lead to the misleading interpretation that achievement gaps are eliminated when in fact they are only reduced relative to baseline (but see Sherman et al., 2013 for an analytic solution). The reporting of both estimates is conducive to a thorough understanding of the effects of the intervention, as overly simplified communications could lead to inaccuracies in understanding among the public (Blanton & Ikizer, 2019).

Finally, the studies excluded in our meta-analysis are worth highlighting because they raised theoretically-relevant issues beyond the scope of this meta-analysis. Although we did not include studies that were conducted in the lab, we did identify several experiments that tested the effects of a values affirmation given in the lab on later achievement outcomes outside the lab. Specifically, Layous et al. (2017) and Brady et al. (2016) both conducted a conventional lab-administered values affirmation among college students and collected official GPA data. Brady et al. (2016) found a positive effect on the GPA for racial minority students (g = 0.50, 95% *CI* [0.19, 0.80]), and Layous et al. (2017) found a positive effect on students overall, and especially for students experiencing a low sense of belonging at a large public university (g = 0.54, 95% *CI* [0.10, 0.97]). This suggests that, at least sometimes, affirmation’s benefits might spillover from the laboratory and into the field, perhaps because of self-perceptual changes the experience induces. We also excluded three studies that were conducted with students in online courses or other online evaluative settings (Kizilcec, Davis, et al., 2017; Kizilcec, Saltarelli, et al., 2017; Linos et al., 2017) because experiences in these virtual settings can be very different from those in an institutional classroom. The first study found that values affirmation significantly increased the completion rate among the lowest-performing group, lower-class men; the second study found that values affirmation increased the persistence and completion rate of online courses for students in less developed countries, who, the authors suggest, are stereotyped as less capable in a global learning context; and the third study found that values affirmation improved the test performance of minority applicants for the job of the police constable. Finally, Cook et al., 2012 (Study 2) was excluded because they did not have a control group. But they specifically tested the treatment effect of values affirmation implemented early in the school year against one that was implemented four weeks later and found that students receiving early implementation of values affirmation earned higher grades.

Limitations

First, although both the primary and secondary outcomes reflected students' achievement, it might be problematic to claim that they all measured the same underlying construct (Dorans & Holland, 2000). The primary outcome GPA might be based on different criteria in different schools. Likewise, secondary outcomes also varied. Nevertheless, all the outcome measures were the de facto indicators of academic achievement at the sites where the studies were conducted.

Second, we assessed only the moderators that displayed sufficient variability across studies. As discussed above, there were many theoretically important moderators that did not vary sufficiently across studies to test.

Third, we could not determine the causal role of moderators, as noted. We hope our results inspire researchers to experimentally manipulate them in future studies.

Fourth, we were unable to assess longitudinal trends of the affirmation effect, as Borman et al. (2018) and Cohen et al. (2009) did in their studies. As a consequence, we could not determine whether the long-range effects of the intervention took one of three forms: 1) a single jump in performance after the intervention, 2) gradual changes at a constant rate over time, or 3) gradual changes at a different rate (Miller et al., 2017).

Fifth, we could not find a perfect solution to disentangle the affirmation effect for one identity group from another if they were both recorded in the same study or the participants embodied more than one threatening identities. Mostly, we were only able to extract effect sizes for one identity subgroup controlling for the other identity (e.g., in a regression, if reported), rather than effects solely for students with intersectional identities as most studies assumed a main threatening identity. Therefore, it remained a question whether multiple identities ever interacted in producing stereotype threats within students with multiple salient identities in each study (see also Gonzales et al., 2002, Harackiewicz et al., 2014, and Lokhande & Muller, 2019 for further discussions). We encourage future researchers to undertake analyses that specifically address this issue of intersectionality.

Finally, although we found a significant affirmation effect and identified key moderators, we are limited in our ability to address the “financial and logistic” as well as “political” challenges in scaling up values affirmation (Ross et al. 2010). Borman et al. (2016) noted that values affirmation could serve as one useful tool among a larger arsenal of approaches to close achievement gaps, as it imposes few costs, is easy to administer, and, under the right conditions, yields effect sizes comparable to other curricular interventions and school reforms (Borman et al., 2003). Borman et al. (2019) further provided an estimate of $1.35 per student per year to deliver two brief social-psychological interventions, “a small fraction” of the $86,000 typically spent to implement an average school reform program (Borman et al. 2016, p. 38). The costs of the intervention are negligible, making the potential harms of administering it relatively small. That said, we suspect that any cost of the intervention can be averted by “aiming well”: administering it in contexts where it is likely to be beneficial, and to students for whom it is likely to be beneficial (see also Easterbrook & Hadden, 2020). The model of this intervention is analogous to medical treatment: first a set of diagnostic criteria is applied, then a treatment prescribed (Garcia & Cohen, 2011). As our moderator analysis suggests, scaling up values affirmation is not so much a matter of disseminating it to every student as it is allocating it judiciously to those students, and in those contexts, where it is beneficial (Binning & Browman, 2020; Cohen et al., 2017). The finding that values affirmation benefited students more in well-resourced schools should remind practitioners and researchers that affirmation is far from a psychological cure-all. It is a catalyst, and like all psychological interventions, has an appropriate time and place for their use.

Conclusion

Overall, our meta-analytic review found that values affirmation improved the academic achievement of identity-threatened students but not identity-nonthreatened students, and that intervention effectiveness was predicted by key moderators. We hope that these findings provide guidance to practitioners in their efforts to understand how to create more equitable classrooms for students from diverse backgrounds. We also hope they help researchers plan future studies that identify the boundary conditions of affirmation, the ideal conditions for its benefits, and methods for increasing its efficacy. Our results suggest that values affirmation intervention is one useful tool to help close achievement gaps in some contexts. However, values affirmation, like other social-psychological interventions, should not replace larger structural and systemic efforts to close achievement gaps. Rather, they should complement them (Cohen & Sherman, 2014). Aspiration, hope, a sense of belonging, and other precious psychological states do not bear fruit by themselves. They require structural pathways and material resources for their benefits to come to fruition.

Reference

References marked with an asterisk indicate studies included in the meta-analysis.

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. https://doi.org/10.1006/jesp.2001.1491

Aronson, J., & Good, C. (2002). The development and consequences of stereotype vulnerability in adolescents. *F. Pajares & T. Urdan (Hg.), Academic Motivation of Adolescents*, 299–330.

Assink, M., & Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. https://doi.org/10.1016/j.cpr.2015.08.002

Baird, M. D., & Pane, J. F. (2019). Translating standardized effects of education programs into more interpretable metrics. *Educational Researcher*, *48*(4), 217–228. https://doi.org/10.3102/0013189X19848729

*Baker, D. J., Skinner, B. T., & Redding, C. H. (2019). Affirmative intervention to reduce stereotype threat bias: Experimental evidence from a community college. *The Journal of Higher Education*, 1–33. https://doi.org/10.1080/00221546.2019.1650582

Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V., & Van Dooren, W. (2019). Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educational Studies in Mathematics*, *102*(1), 1–8. https://doi.org/10.1007/s10649-019-09908-4

*Bancroft, A., Bratter, J., & Rowley, K. (2017). Affirmation effects on math scores: The importance of high school track. *Social Science Research, 64*, 319–333. https://doi.org/10.1016/j.ssresearch.2016.10.001

Barton, P. E., & Coley, R. J. (2010). *The Black-White achievement gap: When progress stopped. Policy Information Report.* Educational Testing Service. https://files.eric.ed.gov/fulltext/ED511548.pdf

*Bayly, B. L. (2017). *Evaluating the effectiveness of an online values affirmation among first-year college students* [Doctoral dissertation, Washington State University]. http://hdl.handle.net/2376/13015

Binning, K. R., & Browman, A. S. (2020). Theoretical, ethical, and policy considerations for conducting social-psychological interventions to close educational achievement gaps. *Social Issues and Policy Review*, *14*(1), 182–216. https://doi.org/10.1111/sipr.12066

Binning, K. R., Cook, J. E., Purdie-Greenaway, V., Garcia, J., Chen, S., Apfel, N., Sherman, D. K., & Cohen, G. L. (2019). Bolstering trust and reducing discipline incidents at a diverse middle school: How self-affirmation affects behavioral conduct during the transition to adolescence. *Journal of School Psychology*, *75*, 74–88. https://doi.org/10.1016/j.jsp.2019.07.007

*Binning, K. R., Cook, J. E., Purdie-Vaughns, V., Garcia, J., Apfel, N., Sherman, D. K., & Cohen, G. L. (under review). *Understanding psychological threat in middle school: Emergent benefits of self-affirmation writing exercises*.

Blanton, H., & Ikizer, E. G. (2019). Elegant science narratives and unintended influences: An agenda for the science of science communication. *Social Issues and Policy Review*, *13*(1), 154–181. https://doi.org/10.1111/sipr.12055

Bohrnstedt, G., Kitmitto, S., Ogut, B., Sherman, D., & Chan, D. (2015). *School composition and the Black-White achievement gap. NCES 2015-018.* National Center for Education Statistics.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

*Borman, G. D. (2012). Examination of a self-affirmation intervention in St. Paul Public Schools. *The Senior Urban Education Research Fellowship Series*, *9*, 1–44.

Borman, G. D. (2017). Advancing values affirmation as a scalable strategy for mitigating identity threats and narrowing national achievement gaps. *Proceedings of the National Academy of Sciences*, *114*(29), 7486–7488. https://doi.org/10.1073/pnas.1708813114

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. https://doi.org/10.3102/0162373715581709

*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*. Society for Research on Educational Effectiveness. https://eric.ed.gov/?id=ED562534

*Borman, G. D., Grigg, J., Rozek, C. S., Hanselman, P., & Dewey, N. A. (2018). Self-affirmation effects are produced by school context, student engagement with the intervention, and time: Lessons from a district-wide implementation. *Psychological Science*, *29*(11), 1773–1784. https://doi.org/10.1177/0956797618784016

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *73*(2), 125–230. https://doi.org/10.3102/00346543073002125

Borman, G. D., Rozek, C. S., Pyne, J., & Hanselman, P. (2019). Reappraising academic and social adversity improves middle school students' academic achievement, behavior, and well-being. *Proceedings of the National Academy of Sciences*, *116*(33), 16286–16291. https://doi.org/10.1073/pnas.1820317116

*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. https://doi.org/10.1037/a0031177

Bradley, D. N., Crawford, E. P., & Dahill-Brown, S. E. (2016). Defining and assessing FoI in a large-scale randomized trial: Core components of values affirmation. *Studies in Educational Evaluation*, *49*, 51–65. https://doi.org/10.1016/j.stueduc.2016.04.002

Brady, S. T., Reeves, S. L., Garcia, J., Purdie-Vaughns, V., Cook, J. E., Taborsky-Barba, S., Tomasetti, S., Davis, E. M., & Cohen, G. L. (2016). The psychology of the affirmed learner: Spontaneous self-affirmation in the face of stress. *Journal of Educational Psychology*, *108*(3), 353–373. https://doi.org/10.1037/edu0000091

*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. https://doi.org/10.1007/s12552-016-9187-4

Cheung, S. F., & Chan, D. K.-S. (2004). Dependent effect sizes in meta-analysis: Incorporating the degree of interdependence. *Journal of Applied Psychology*, *89*(5), 780–791. https://doi.org/10.1037/0021-9010.89.5.780

Cheung, S. F., & Chan, D. K.-S. (2008). Dependent correlations in meta-analysis: The case of heterogeneous dependence. *Educational and Psychological Measurement*, *68*(5), 760–777. https://doi.org/10.1177/0013164408315263

*Churchill, S., Jessop, D. C., Goodwin, S., Ritchie, L., & Harris, P. R. (2018). Self-affirmation improves music performance among performers high on the impulsivity dimension of sensation seeking. *Psychology of Music*, *46*(2), 292–302. https://doi.org/10.1177/0305735617705007

Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, *26*(9), 1151–1164. https://doi.org/10.1177/01461672002611011

Cohen, G. L., & Garcia, J. (2008). Identity, belonging, and achievement: A model, interventions, implications. *Current Directions in Psychological Science*, *17*(6), 365–369. https://doi.org/10.1111/j.1467-8721.2008.00607.x

*Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*(5791), 1307–1310. https://doi.org/10.1126/science.1128317

Cohen, G. L., Garcia, J., & Goyer, J. P. (2017). Turning point: Targeted, tailored, and timely psychological intervention. In *Handbook of competence and motivation: Theory and application, 2nd ed* (pp. 657–686). The Guilford Press.

*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. https://doi.org/10.1126/science.1170769

Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*, *65*(1), 333–371. https://doi.org/10.1146/annurev-psych-010213-115137

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic Press.

*Cook, J. E., Purdie-Vaughns, V., Garcia, J., & Cohen, G. L. (2012). Chronic threat and contingent belonging: Protective benefits of values affirmation on identity development.

Journal of Personality and Social Psychology, *102*(3), 479. https://doi.org/10.1037/a0026312

*De Clercq, M., Michel, C., Remy, S., & Galand, B. (2019). Providing freshmen with a good "starting-block": Two brief social-psychological interventions to promote early adjustment to the first year at university. *Swiss Journal of Psychology*, *78*(1–2), 69–75. https://doi.org/10.1024/1421-0185/a000217

*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*. https://doi.org/10.3389/fpsyg.2016.00691

*Dee, T. S. (2015). Social identity and achievement gaps: Evidence from an affirmation intervention. *Journal of Research on Educational Effectiveness*, *8*(2), 149–168. https://doi.org/10.1080/19345747.2014.906009

Del Re, A. C. (2013). *compute.es: Compute Effect Sizes. R package* (Version 0.2-2) [Computer software]. https://cran.r-project.org/web/packages/compute.es/index.html

Domina, T., Pharris-Ciurej, N., Penner, A. M., Penner, E. K., Brummet, Q., Porter, S. R., & Sanabria, T. (2018). Is free and reduced-price lunch a valid measure of educational disadvantage? *Educational Researcher*, *47*(9), 539–555. https://doi.org/10.3102/0013189X18797609

Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equatability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, *37*(4), 281–306. https://doi.org/10.1111/j.1745-3984.2000.tb01088.x

Duncan, G. J., & Murnane, R. J. (2014). *Restoring opportunity: The crisis of inequality and the challenge for American education*. Harvard Education Press.

Easterbrook, M. J., & Hadden, I. R. (2020). Tackling educational inequalities with social psychology: Identities, contexts, and interventions. *Social Issues and Policy Review*, *0*(0), 1–57. https://doi.org/10.1111/sipr.12070

Easterbrook, M. J., Harris, P. R., & Sherman, D. K. (2021). Self-affirmation theory in educational contexts [Special issue]. *Journal of Social Issues*.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Ferrer, R., & Cohen, G. (2018). Reconceptualizing self-affirmation with the trigger and channel framework: Lessons from the health domain. *Personality and Social Psychology Review*, *23*, 108886831879703. https://doi.org/10.1177/1088868318797036

Fryer Jr., R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, *2*(2), 210–240. https://doi.org/10.1257/app.2.2.210

Garcia, J., & Cohen, G. L. (2011). *A social psychological perspective on educational intervention* (SSRN Scholarly Paper ID 1910232). Social Science Research Network. https://papers.ssrn.com/abstract=1910232

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, *28*(5), 659–670. https://doi.org/10.1177/0146167202288010

Goyer, J. P., Cohen, G. L., Cook, J. E., Master, A., Apfel, N., Lee, W., Henderson, A. G., Reeves, S. L., Okonofua, J. A., & 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. https://doi.org/10.1037/pspa0000152

*Goyer, J. P., Garcia, J., Purdie-Vaughns, V., Binning, K. R., Cook, J. E., Reeves, S. L., Apfel, N., Taborsky-Barba, S., Sherman, D. K., & Cohen, G. L. (2017). Self-affirmation facilitates minority middle schoolers' progress along college trajectories. *Proceedings of the National Academy of Sciences*, *114*(29), 7594–7599. https://doi.org/10.1073/pnas.1617923114

*Gutmann, B. (2019). *Tools for underprepared students in engineering physics with a focus on online mastery learning exercises* [Doctoral dissertation, University of Illinois at Urbana-Champaign]. http://hdl.handle.net/2142/105623

*Hadden, I. R., Easterbrook, M. J., Nieuwenhuis, M., Fox, K. J., & Dolan, P. (2019). Self-affirmation reduces the socioeconomic attainment gap in schools in England. *British Journal of Educational Psychology*, *90*(2), 517–536. https://doi.org/10.1111/bjep.12291

*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. https://doi.org/10.1177/0038040714525970

*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. https://doi.org/10.1037/edu0000141

*Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Giffen, C. J., Blair, S. S., Rouse, D. I., & Hyde, J. S. (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. *Journal of Educational Psychology*, *106*(2), 375–389. https://doi.org/10.1037/a0034679

*Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., & Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of Personality and Social Psychology*, *111*(5), 745. https://doi.org/10.1037/pspp0000075

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2019). Doing meta-analysis in R: A hands-on guide. *Https://Bookdown.Org/MathiasHarrer/Doing_Meta_Analysis_in_R*. https://doi.org/10.5281/zenodo.2551803

Harris, P. R., & Napper, L. (2005). Self-affirmation and the biased processing of threatening health-risk information. *Personality and Social Psychology Bulletin*, *31*(9), 1250–1263. https://doi.org/10.1177/0146167205274694

Harvey, R. D., & Oswald, D. L. (2000). Collective guilt and shame as motivation for White support of Black programs. *Journal of Applied Social Psychology*, *30*(9), 1790–1811. https://doi.org/10.1111/j.1559-1816.2000.tb02468.x

*Hayes, L., Zinner, L., Wise, J., & Carton, J. (2019). Effects of a self-Affirmation intervention on grades in middle school and first-year college students. *Journal of Articles in Support of the Null Hypothesis*, *16*(1), 57–79.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, *6*(3), 203–217. https://doi.org/10.1037/1082-989x.6.3.203

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., & Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, *343*. https://doi.org/10.1136/bmj.d5928

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, J. P. T., & Thompson, S. G. (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine*, *23*(11), 1663–1682. https://doi.org/10.1002/sim.1752

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

Hobbs, G., & Vignoles, A. (2007). *Is free school meal status a valid proxy for socio-economic status (in schools research)?* Centre for the Economics of Education, London School of Economics and Political Science.

Hoffman, A. J., & Kurtz-Costes, B. (2019). Promoting science motivation in American Indian middle school students: An intervention. *Educational Psychology*, *39*(4), 448–469. https://doi.org/10.1080/01443410.2018.1527019

IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, *6*(7), e010247. https://doi.org/10.1136/bmjopen-2015-010247

*Jordt, H., Eddy, S. L., Brazil, R., Lau, I., Mann, C., Brownell, S. E., King, K., & 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), ar41. https://doi.org/10.1187/cbe.16-12-0351

*Kim, J. (2019). *The role of self-affirmation and self-construal levels in attenuating the gender performance gap* [Doctoral dissertation, Columbia University]. https://doi.org/10.7916/d8-83yx-b711

*Kinias, Z., & Sim, J. (2016). Facilitating women's success in business: Interrupting the process of stereotype threat through affirmation of personal values. *Journal of Applied Psychology*, *101*(11), 1585. https://doi.org/10.1037/apl0000139

Kizilcec, R. F., Davis, G. M., & Cohen, G. L. (2017). Towards equal opportunities in MOOCs: Affirmation reduces gender & social-class achievement gaps in China. *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, 121–130. https://doi.org/10.1145/3051457.3051460

Kizilcec, R. F., Saltarelli, A. J., Reich, J., & Cohen, G. L. (2017). Closing global achievement gaps in MOOCs. *Science*, *355*(6322), 251–252. https://doi.org/10.1126/science.aag2063

*Kost-Smith, L. E., Pollock, S. J., Finkelstein, N. D., Cohen, G. L., Ito, T. A., & Miyake, A. (2010). Gender differences in physics 1: The impact of a self-affirmation intervention. *AIP Conference Proceedings*, *1289*(1), 197–200. https://doi.org/10.1063/1.3515197

*Kost-Smith, L. E., Pollock, S. J., Finkelstein, N. D., Cohen, G. L., Ito, T. A., & Miyake, A. (2012). Replicating a self-affirmation intervention to address gender differences: Successes and challenges. *AIP Conference Proceedings*, *1413*(1), 231–234. https://doi.org/10.1063/1.3680037

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

*Lauer, S., Momsen, J., Offerdahl, E., Kryjevskaia, M., Christensen, W., & Montplaisir, L. (2013). Stereotyped: Investigating gender in introductory science courses. *CBE—Life Sciences Education*, *12*(1), 30–38. https://doi.org/10.1187/cbe.12-08-0133

Layous, K., Davis, E. M., Garcia, J., Purdie-Vaughns, V., Cook, J. E., & Cohen, G. L. (2017). Feeling left out, but affirmed: Protecting against the negative effects of low belonging in college. *Journal of Experimental Social Psychology*, *69*, 227–231. https://doi.org/10.1016/j.jesp.2016.09.008

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher*, *31*(1), 3–12. https://doi.org/10.3102/0013189X031001003

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Harvard University Press. https://agris.fao.org/agris-search/search.do?recordID=XF2015046841

Linos, E., Reinhard, J., & Ruda, S. (2017). Levelling the playing field in police recruitment: Evidence from a field experiment on test performance. *Public Administration*, *95*(4), 943–956. https://doi.org/10.1111/padm.12344

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. National Center for Special Education Research.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE publications, Inc.

Liu, S., Liu, P., Wang, M., & Zhang, B. (2020). Effectiveness of stereotype threat interventions: A meta-analytic review. *The Journal of Applied Psychology*. https://doi.org/10.1037/apl0000770

*Lokhande, M., & Muller, T. (2019). Double jeopardy—Double remedy? The effectiveness of self-affirmation for improving doubly disadvantaged students' mathematical performance. *Journal of School Psychology*, *75*, 58–73. https://doi.org/10.1016/j.jsp.2019.07.006

Lüdecke, D. (2017). *esc. R package* (Version 0.4.0) [Computer software]. https://cran.r-project.org/web/packages/esc/index.html

Manstead, A. S. R. (2018). The psychology of social class: How socioeconomic status impacts thought, feelings, and behaviour. *British Journal of Social Psychology*, *57*(2), 267–291. https://doi.org/10.1111/bjso.12251

McGraw, R., Lubienski, S. T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. *Journal for Research in Mathematics Education*, *37*(2), 129–150. JSTOR. https://doi.org/10.2307/30034845

McQueen, A., & Klein, W. M. P. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity*, *5*(4), 289–354. https://doi.org/10.1080/15298860600805325

Miller, D. T., Dannals, J. E., & Zlatev, J. J. (2017). Behavioral processes in long-lag intervention studies. *Perspectives on Psychological Science*, *12*(3), 454–467. https://doi.org/10.1177/1745691616681645

*Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, *330*(6008), 1234–1237. https://doi.org/10.1126/science.1195996

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), 6. https://doi.org/10.1136/bmj.b2535

Musu-Gillette, L., De Brey, C., McFarland, J., Hussar, W., Sonnenberg, W., & Wilkinson-Flicker, S. (2017). *Status and trends in the education of racial and ethnic groups 2017. NCES 2017-051.* National Center for Education Statistics. https://nces.ed.gov/pubs2017/2017051.pdf

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*(2), 157–159. https://doi.org/10.2307/1164923

Pastor, D. A., & Lazowski, R. A. (2018). On the multilevel nature of meta-analysis: A tutorial, comparison of software programs, and discussion of analytic choices. *Multivariate Behavioral Research*, *53*(1), 74–89. https://doi.org/10.1080/00273171.2017.1365684

*Peters, E., Shoots-Reinhard, B., Tompkins, M. K., Schley, D., Meilleur, L., Sinayev, A., Tusler, M., Wagner, L., & Crocker, J. (2017). Improving numeracy through values affirmation enhances decision and STEM outcomes. *PLoS One*, *12*(7). https://doi.org/10.1371/journal.pone.0180674

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991–996. https://doi.org/10.1016/j.jclinepi.2007.11.010

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology*, *90*(1), 175–181. https://doi.org/10.1037/0021-9010.90.1.175

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. https://doi.org/10.1177/0956797615614591

*Protzko, J., & Aronson, J. (2016). Context moderates affirmation effects on the ethnic achievement gap. *Social Psychological and Personality Science*, *7*(6), 500–507. https://doi.org/10.1177/1948550616646426

*Purdie-Greenaway, V., Pasek, M., Cook, J. E., Sumner, R., Apfel, N., Garcia, J., & Cohen, G. L. (under review). *A bend in history: 2008 Obama election intervention reduces threat and improves performance for minority youth*.

*Rapa, L. J. (2016). *Fostering marginalized youths' academic achievement and critical consciousness through a values-affirmation intervention* [Doctoral dissertation, Michigan State University]. https://d.lib.msu.edu/etd/3881

Reardon, S. F. (2018). The widening academic achievement gap between the rich and the poor. In D. B. Grusky & J. Hill (Eds.), *Inequality in the 21st Century* (1st ed., pp. 177–189). Routledge. https://doi.org/10.4324/9780429499821-33

Rohatgi, A. (2020). *WebPlotDigitizer* (Version 4.3) [Computer software]. Pacifica, CA, USA. https://automeris.io/WebPlotDigitizer/

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*(1), 16–20. https://doi.org/10.1007/BF02322211

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, *52*(1), 59–82. https://doi.org/10.1146/annurev.psych.52.1.59

Ross, L., Lepper, M., & Ward, A. (2010). History of social psychology: Insights, challenges, and contributions to theory and application. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of Social Psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy001001

*Rozek, C. S., Hanselman, P., Feldman, R. C., Quast, E. A., Crawford, E. P., & Borman, G. D. (2015). Inside the black box of self-affirmation: Which parts of affirmation exercises are critical for treatment efficacy? *Society for Research on Educational Effectiveness*.

*Schwalbe, M. (2018). *Economic disadvantage as a threat to adaptive adequacy: The role of self-affirmation in improving cognitive and academic performance of the poor* [Doctoral dissertation, Stanford University]. ProQuest Dissertations & Theses Global. http://proxy.library.nyu.edu/login?url=https://www-proquest-com.proxy.library.nyu.edu/docview/2436423459?accountid=12768

Schwarzer, G. (2007). meta: An R package for meta-analysis. *R News*, *7*(3), 40–45.

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R* (Vol. 4784). Springer.

*Serra-Garcia, M., Hansen, K. T., & Gneezy, U. (2020). Can short psychological interventions affect educational performance? Revisiting the effect of self-affirmation interventions. *Psychological Science, 31*(7), 865–872. https://doi.org/10.1177/0956797620923587

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. In *Advances in Experimental Social Psychology* (Vol. 38, pp. 183–242). Elsevier. https://doi.org/10.1016/S0065-2601(06)38004-5

Sherman, D. K., Cohen, G. L., Nelson, L. D., Nussbaum, A. D., Bunyan, D. P., & Garcia, J. (2009). Affirmed yet unaware: Exploring the role of awareness in the process of self-affirmation. *Journal of Personality and Social Psychology*, *97*(5), 745.

*Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., Tomassetti, S., Nussbaum, A. D., & Cohen, G. L. (2013). Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology*, *104*(4), 591. https://doi.org/10.1037/a0031495

Sherman, D. K., Lokhande, M., Muller, T., & Cohen, G. L. (2021). Self-affirmation interventions. In *G. M. Walton and A. J. Crum (Eds.). Handbook of Wise Interventions: How Social Psychology Can Help People Change* (pp. 63–99). New York: Guilford Press.

*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. https://doi.org/10.1177/0146167213480816

*Silverman, A. M., & Cohen, G. L. (2014). Stereotypes as stumbling-blocks: How coping with stereotype threat affects life outcomes for people with physical disabilities. *Personality and Social Psychology Bulletin*, *40*(10), 1330–1340. https://doi.org/10.1177/0146167214542800

*Simmons, C. M. (2011). *Reducing stereotype threat in academically at-risk African-Americans students: A self-affirmation intervention* [Doctoral dissertation]. University of California, Berkeley.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simmons, J. P., & Simonsohn, U. (2017). Power posing: P-curving the evidence. *Psychological Science*. https://doi.org/10.1177/0956797616658563

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Smith, E. N., Rozek, C. S., Manke, K. J., Schmidt, A. J., Crawford, E. P., Dweck, C. S., & Walton, G. M. (2021). Why am I doing this? Affirmations that convey teachers' broad regard for students' lives increase engagement and benefits [Special issue]. *Journal of Social Issues*.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In *Advances in Experimental Social Psychology* (Vol. 21, pp. 261–302). Elsevier. https://doi.org/10.1016/S0065-2601(08)60229-4

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613. https://doi.org/10.1037//0003-066x.52.6.613

Steele, C. M. (2005). Stereotypes and the fragility of academic competence, motivation, and self-concept. In *Handbook of competence and motivation* (pp. 436–456). Guilford Publications.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In *Advances in Experimental Social Psychology* (Vol. 34, pp. 379–440). Elsevier. https://doi.org/10.1016/S0065-2601(02)80009-0

Stephens, N. M., Markus, H. R., & Phillips, L. T. (2014). Social class culture cycles: How three gateway contexts shape selves and fuel inequality. *Annual Review of Psychology*, *65*(1), 611–634. https://doi.org/10.1146/annurev-psych-010213-115143

Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PLoS ONE*, *8*(3), e57988. https://doi.org/10.1371/journal.pone.0057988

Sutton, A. J. (2009). Publication bias. In *The handbook of research synthesis and meta-analysis* (pp. 435–452). Russell Sage Foundation.

*Tibbetts, Y., Harackiewicz, J. M., Canning, E. A., Boston, J. S., Priniski, S. J., & Hyde, J. S. (2016). Affirming independence: Exploring mechanisms underlying a values affirmation intervention for first-generation students. *Journal of Personality and Social Psychology*, *110*(5), 635. https://doi.org/10.1037/pspa0000049

*Tibbetts, Y., Priniski, S. J., Hecht, C. A., Borman, G. D., & Harackiewicz, J. M. (2018). Different institutions and different values: Exploring first-generation student fit at 2-year colleges. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00502

*Turetsky, K. M., Purdie-Greenaway, V., Cook, J. E., Curley, J. P., & Cohen, G. L. (2020). A psychological intervention strengthens students' peer social networks and promotes persistence in STEM. *Science Advances*, *6*(45), eaba9221. https://doi.org/10.1126/sciadv.aba9221

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6

Viechtbauer, W., & Viechtbauer, M. W. (2020). *metafor. R pacakge* (2.4-0) [Computer software]. https://cran.r-project.org/web/packages/metafor/index.html

Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review*, *125*(5), 617–655. https://doi.org/10.1037/rev0000115

Walton, G. M., & Yeager, D. S. (2020). Seed and soil: Psychological affordances in contexts help to explain where wise interventions succeed or fail. *Current Directions in Psychological Science*, *29*(3), 219–226. https://doi.org/10.1177/0963721420904453

Wilson, D. B. (2011). *Practical meta-analysis effect size calculator. The Campbell Collaboration*. The Campbell Collaboration.

*Woolf, K., McManus, I. C., Gill, D., & Dacre, J. (2009). The effect of a brief social intervention on the examination results of UK medical students: A cluster randomised controlled trial. *BMC Medical Education*, *9*(1), 35. https://doi.org/10.1186/1472-6920-9-35

*Wynne, C. B. (2011). *The implementation of a promising self-affirmation intervention among negatively stereotyped students in a randomized field experiment* [Doctoral dissertation, Harvard University]. ProQuest Dissertations and Theses Global

Yeager, D. S., Carroll, J. M., Buontempo, J., Cimpian, A., Woody, S., Crosnoe, R., Muller, C., Murray, J. S., Mhatre, P., Kersting, N., Hulleman, C., Kudym, M., Murphy, M., Duckworth, A. L., Walton, G. M., & Dweck, C. S. (under review). *Teacher mindsets help explain where a growth mindset intervention does and doesn't work*.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho, C. M., … Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364–369. https://doi.org/10.1038/s41586-019-1466-y