Self-affirmation theory in educational contexts

Matthew J. Easterbrook1 | Peter R. Harris1 | David K. Sherman2

1 School of Psychology, University of Sussex, Brighton, East Sussex, UK
2 Psychological and Brain Sciences, University of California, Santa Barbara, California, USA

Correspondence
Matthew J. Easterbrook, School of Psychology, University of Sussex, Brighton, East Sussex, UK.
Email: m.j.easterbrook@sussex.ac.uk

Abstract
Self-affirmation, operationalized as value-affirmation interventions, can have long-term beneficial effects on the academic performance and trajectories of members of negatively stereotyped groups, thus reducing achievement gaps. Yet, there is significant heterogeneity in the effectiveness of value affirmations, and we do not yet have a clear understanding of why. In this introduction to the special issue, we review the literature on self-affirmation theory in educational contexts, providing overviews of the heterogeneity in the effectiveness of affirmation interventions, the methods of implementation, potential moderators, and underlying processes. We identify several questions that are important for researchers to address, the answers to which would progress the field towards being able to more confidently implement value-affirmations in contexts in which, and/or for groups for whom, they are most likely to produce benefits. We then introduce the articles included in this special issue, which showcase several of the latest theoretical and empirical advances to self-affirmation theory in educational contexts.

Some groups of students—typically members of groups who have low status in wider society, such as some ethnic or racial groups, those from lower social classes, or those with immigrant backgrounds—have, on average, worse educational outcomes than others. Across societies, such inequalities in educational outcomes (henceforth educational inequalities) exist from the earliest years through to tertiary education, and their effects continue throughout the lifespan (Department for Education in England, 2020; Nation’s Report Card, 2019; OECD, 2015). Indeed, an individual’s level of education is a key predictor of a range of life outcomes (Hout & DiPrete, 2006),...
including mental and physical health and wellbeing (Blanchflower & Oswald, 2000, von dem Kenebeek et al., 2011; Schütte et al., 2013), life expectancy (Meara et al., 2008), trust (Easterbrook et al., 2015), and income (Britton et al., 2020). Those with higher qualifications are also more engaged in politics and are more likely to vote, giving them a louder political voice (Bynner & Ashford, 1994; Helliwell & Putnam, 2007; Persson, 2013; Stubager, 2008). There are thus important economic, social, and moral reasons for attempting to remove or dampen any forces that damage educational outcomes for certain groups of students.

A multitude of factors contribute to educational inequalities, including economic and structural inequalities that disadvantage some groups. In the US, for example, African Americans and Latinos tend to have lower academic achievement and progression than European Americans (Nation’s Report Card, 2019). Undoubtedly, economic and structural inequalities that are rooted in the history of race relations in the US—such as lower incomes and less access to good quality schools—disadvantage these groups in the educational domain and thus contribute to educational inequalities. In England, the largest achievement gap between the major pupil groups is between pupils eligible for free school meals—often taken as an indicator of economic disadvantage—and their peers (Department for Education, 2020). Given that English society has historically been stratified by class, this gap has been fueled by historical, structural, and economic inequalities deeply rooted in history and society. Educational inequalities therefore tend to reflect deeper, ingrained inequalities in society that need to be addressed if educational inequalities are to be significantly reduced.

Not only are some groups disadvantaged educationally by these broader structural and economic factors, but are also more likely to experience psychological states that stem from these wider inequalities and that further impede these groups’ educational success and progression. Reflecting the historical and structural inequalities that have existed in their societies, members of groups that have historically been minoritized, alienated, stigmatized, and/or excluded from education tend to perceive educational environments as psychologically threatening; as places in which they and their group are not valued or wanted, and are expected to fail. This can make members of those groups feel as if they do not belong in education, that education is not for ‘people like them’, and may induce experiences of stereotype threat—a fear of confirming a negative stereotype about their group (Easterbrook et al., 2019; Spencer et al., 2016; Steele & Aronson, 1995; Veldman et al., 2019; Walton & Cohen, 2007). These psychological factors are associated with reduced confidence and wellbeing (Iyer et al., 2009; Spencer & Castano, 2007), increased stress (Hadden et al., 2020), and can impede learning (Taylor & Walton, 2011), performance (Flore & Wicherts, 2015; Nguyen & Ryan, 2008; Shewach et al., 2019; Walton & Spencer, 2009), and motivation within educational contexts (Manstead, 2018). They have been found to account for a sizable proportion of educational inequalities, even after academic ability has been accounted for (Nieuwenhuis et al., 2019; Walton & Spencer, 2009).

**SELF-AFFIRMATION THEORY**

Psychological factors therefore contribute to educational inequalities, which makes interventions that target psychological states and processes potentially effective techniques to reduce educational inequalities (Easterbrook & Hadden, 2021). Indeed, self-affirmation theory has proved to be useful for understanding and addressing inequalities in educational outcomes, particularly academic performance (Cohen & Sherman, 2014; Sherman & Cohen, 2006; Steele, 1988). The key tenet of the theory posits that people have a need to see themselves as a good and adequate person; as having “self-integrity” (Steele, 1988, p 262). A challenge to one’s self-integrity is often
experienced as a sense of psychological threat; an aversive state in which important aspects of the self are challenged. Individuals experiencing threat can often become motivated to attempt to alleviate the threat and bolster their sense of self-integrity. Often, however, these attempts are defensive in character and thus detrimental to productivity, performance, perseverance, and/or progress. These detrimental consequences can be reduced or even eliminated if the person affirms a highly valued aspect of the self that is unrelated to the threat. Affirming an important self-aspect bolsters the individual’s sense of global self-integrity by drawing on important self-aspects that are not threatened, which reduces the need to engage in detrimental attempts to restore the sense of integrity derived from the threatened aspect of the self.

Interventions based on the theory, known as value affirmation exercises, tend to be brief writing exercises that encourage participants to reflect on their most important values. Cohen and colleagues’ (2006) pioneering article in *Science* was the first demonstration of the efficacy of value-affirmation interventions in reducing educational achievement gaps. These researchers showed that brief value affirmation exercises, administered by teachers to all students in their classrooms and implemented at several points throughout the academic year, increased the academic performance of 12-13-year-old African American US middle-school students, but had no effect on European American students’ performance, thus, reducing the racial achievement gap by 40%. In fact, the exercises stopped the downward trajectory in performance that African American students tended to show across the academic year and were particularly beneficial for low performing African American students. The finding that the exercises only benefited the academic performance of African American students supported the theoretical prediction that self-affirmation ameliorates the detrimental effects of stereotype threat by bolstering threatened students’ sense of self-integrity. Self-affirmation thus removes a psychological barrier that impedes performance among negatively stereotyped students and enables the positive forces within the educational environment to take effect.

Since that first demonstration, research into self-affirmation interventions in educational contexts has flourished. Because, value affirmation exercises reduce the detrimental consequences of threat, they tend to only exert influence on threatened groups of students, and have thus been documented to reduce achievement gaps between threatened students and their peers (Wu et al., 2021). Value affirmation exercises have been shown to benefit the educational performance of a range of different stereotyped groups, including Latino American students in US middle schools (Sherman et al., 2013), first generation scholars in US colleges (Harackiewicz et al., 2014), students in further education colleges in the UK (Schwalbe et al., 2019), pupils on free-school meals in English schools (Hadden et al., 2020), students from developing nations in Massive Open Online Courses (Kizilcec et al., 2017), female students in graduate business schools (Kinia & Sim, 2016), and immigrant students in Germany (Lokhande & Müller, 2019).

Even more striking are the long-term effects that have been documented. The beneficial effects of the intervention on grades have been shown to persist for at least 2 years (Cohen et al., 2009; Sherman et al., 2013). Remarkably, a more recent paper has shown downstream beneficial effects of the intervention on college enrollment among the students in Cohen et al.’s first study 7-9 years later (Goyer et al., 2017).

Furthermore, it is not only students’ grades that can be improved by self-affirmation interventions in educational settings, but other important outcomes, also. Affirmations have also been found to reduce academic stress (Hadden et al., 2020), increase trust (Sherman & Cohen, 2006), reduce defensiveness (Sherman & Cohen, 2006), and to benefit those who feel like they do not belong to college (Layous et al., 2017). Binning and colleagues (2019) also found that
self-affirmation interventions promoted better behavior among US middle-school students, decreasing disciplinary infractions over students’ three years of middle school.

How are these profound effects possible with such a relatively brief intervention? Developments to self-affirmation theory suggest that the intervention can cause an initial boost in performance by reducing threat and igniting a cycle of “adaptive potential” in which the effects become reinforced through recursive processes embedded within students’ social environments (Cohen et al., 2017; Sherman et al., 2021). For example, an initial boost to performance can produce an increase in confidence, which can have beneficial effects on subsequent performance, which can further increase confidence, and so on. There may also be non-recursive, domino-like effects that the initial intervention causes, which can bring about long-term positive outcomes (Hecht et al., 2019). For instance, an initial boost to performance may cause a pupil to be recognized differently by a teacher and to be placed into a higher-achieving and more academically-oriented class, which then facilitates their progression to college (Goyer et al., 2017). Finally, repeated interventions may lead to intrapersonal changes in students’ latent habits (Hecht et al., 2019) by, for example, increasing the likelihood that they will spontaneously think about their values and thus self-affirm when they feel threatened (Brady et al., 2016).

Given such striking findings it is tempting to think of self-affirmation interventions as a panacea for group-based educational inequalities. Such interventions have been shown to have profound, long-term beneficial effects, are brief and cheap to administer, appear to benefit only (or at least primarily) those students who are suffering from threat (thus reducing achievement gaps), and can have effects that are stronger than lengthier, more traditional, and expensive educational interventions (Walton, 2014; Yeager & Walton, 2011). Indeed, there is increasing interest in value affirmation interventions from educational policy makers and practitioners and several large-scale trials have tested (or are testing) the potential of self-affirmation interventions to be scaled up (e.g., Borman, Choi et al., 2020; Borman, Griggs et al., 2018).

However, there are reasons to be cautious about the unqualified promotion of self-affirmation interventions. We believe that there are several crucial questions, researchers in the field need to consider before rolling out self-affirmation interventions across schools and potentially integrating them into curricula. Most crucially, there is a significant heterogeneity in the effectiveness of the interventions across contexts and groups, and we do not yet have a very good understanding of why. Indeed, there have been several replication studies that have not found benefits of the intervention (e.g., Bratter et al., 2016; de Jong et al., 2016; Hanselman et al., 2017; Protzko & Aronson, 2016; Serra-Garcia et al., 2020) and even some in which the interventions have led to negative effects in some students (see Binning & Browman, 2020). This may, in part, be due to variations in the implementation of the interventions (Borman, 2017), but it may also result from hitherto unidentified moderators of the effectiveness of the intervention. It is essential that the field understands more about the key moderators of the effects before the intervention is implemented on a large-scale and/or integrated into curricula. Indeed, psychological interventions in general could be characterized as suffering from a “context sensitivity” problem (Goroff et al., 2018), in which currently unspecified differences in a situation or sample can change the effectiveness of the intervention in ways we cannot yet predict. Yet, theories (e.g., Easterbrook & Hadden, 2021; Walton & Yeager, 2019) and empirical studies (Borman et al., 2018; Yeager et al., 2019) have started to investigate and indeed find contextual moderators. We hope this special issue will further these endeavors.

Heterogeneity in effectiveness is one reason we should be cautious about the extent to which self-affirmation interventions are broadly promoted, but there are other reasons. Self-affirmation interventions tackle a symptom of stereotype threat, but do little to address the causes, such as
salient negative stereotypes and underrepresentation of some groups in high status positions (Easterbrook & Hadden, 2021). In particular, self-affirmations offer a potentially effective individual approach to a collective and societal issue, an approach that can help students if there are positive pedagogical forces within a school environment that are stymied by identity threat (Cohen et al., 2006). Sustained change that reduces inequality will require collective solutions and societal change, and creating an affirming social psychological context can be part of this change, but is not intended by affirmation researchers to be the sole or even primary intervention that is needed to address educational inequalities. The social environment available to the student— including the cognitive, emotional, and social support provided—will determine whether individual approaches, such as self-affirmation, can lead to lasting change (Sherman et al., 2021); efforts to address inequalities in the social environments of students are paramount.

Relatedly, there can be unintended consequences of promoting wise interventions. Because of their brief and ostensibly simple nature, media coverage of wise interventions such as self-affirmation has been shown in some experimental studies to dampen the public’s concern about inequality and their motivation to address it (Blanton & Ikizer, 2019; Ikizer & Blanton, 2016). This is especially likely as the media often focus on the intervention as the sole cause of any observed improvement, and are less likely to focus on the contextual factors that must exist in order for effects to be perpetuated (Walton & Yeager, 2020). Media accounts of wise interventions are also less likely to focus on the economic and historical factors accounting for achievement gaps and other educational inequalities and thus can have the unintended consequence of focusing the energy on changing the student as opposed to changing the broader context (Ikizer & Blanton, 2016). That a psychological intervention can attenuate educational inequalities does not imply that the educational inequalities have exclusively psychological origins. It implies, as suggested in a review of the affirmation literature that “under some circumstances, psychology can be a key valve through which the influence of cultural, systemic, historical, and institutional forces flow (Sherman et al., 2021, p. 79).”

We thus echo Lewis and Wai’s (2021) call for a cautious approach to science communication that highlights what we know and what we do not know (as well as how we know it), and considers the findings of psychological researchers within the broader societal context. We hope this special issue will advance both the state of the field’s knowledge as well as the most pressing unanswered questions. Furthermore, although some of the mechanisms through which the affirmations can bring about an increase in academic performance have been identified (which we describe below), there are undoubtedly more that are, as yet, not well understood. It is important to understand the processes involved when self-affirmation influences educational outcomes, not least to be able to anticipate how these may be affected by large scale application.

To begin to address these important gaps in our knowledge, in 2019 we organized a two-day meeting at the University of Sussex that brought together many of the world’s leading researchers on the topic (<https://www.easp.eu/news/itm/report_on_joint_easp_spssi_meeting__self_affirmation_in_education-967.html>). Building on that, we have edited this special issue documenting the current literature and understanding of self-affirmation theory in educational contexts and showcasing several of the latest theoretical and empirical advances. In this introduction to the special issue, we give an overview of the current literature on self-affirmation in educational contexts and raise some issues and questions that we believe are useful for researchers to address. We then outline the contributions in this special issue.
HETEROGENEITY OF EFFECTS

If researchers and practitioners are to realize the potential of self-affirmation on a large scale, this heterogeneity must be better understood. Building on the work of others (Binning & Browman, 2020; Borman, 2017; Sherman et al., 2021), we suggest several possible reasons for this heterogeneity and discuss each of them below.

Implementation: Timeliness and the message the intervention portrays

One possible reason for the heterogeneity of self-affirmation effects in educational contexts may be that researchers have deviated from some of the established procedures for implementing the intervention, perhaps because of practical constraints imposed by the schools involved (Borman, 2017). The way in which the intervention is implemented is not solely a matter of procedure but may be crucial to the message that the intervention portrays to students and how this influences their perceptions of their social environment.

One key aspect of the implementation appears to be that the intervention is perceived by pupils as part of their normal classroom activities delivered by teachers who care and want to know more about their most important values (Borman, 2017). In practical terms, this means that pupils must not be aware that the exercises are designed to be beneficial (Sherman, Cohen, et al., 2009) or are an externally imposed task, that the exercises should be introduced and implemented by teachers, and that the exercises should encourage pupils to write about something that they find particularly important in their lives. Smith et al.’s (2021) article in this special issue confirms the importance of the intervention being delivered by teachers rather than researchers. It should be noted, however, that some researchers have successfully implemented affirmation activities that were described as coming from researchers—but they creatively adapted the affirmation materials to do so (Lokhande & Müller, 2019; for a discussion, see Sherman et al., 2021; Smith et al., 2021; Wu et al., 2021).

In theoretical terms, intervening in the way described above can change students’ perceptions of their social reality (Cohen et al., 2017). If a student experiences their school as a threatening place and believes that their teachers do not value members of their social group, then an exercise that portrays the message “we do care about you, about who you are, and about what you value; tell us about yourself” can be impactful and may revolutionize their perception of their environment, reducing threat and thus removing a barrier faced by stereotyped students and enabling them to take advantage of the opportunities offered within the educational context. Furthermore, the forces operating in the social web are dynamic and have bidirectional paths of influence. For example, the intervention has been found to be more effective if the teachers actually read the students’ exercises (Bowen et al., 2013), implying that there are benefits for the students if their teachers do, in fact, know more about them (see Smith et al., 2021).

The intervention should also be implemented at times of heightened stress, such as the beginning of an academic year or before a high-stakes test, so that threat is likely to be salient and the intervention can have a noticeable impact on these negative states. A delay of as little as 2 weeks from the stressful, and presumably threatening, first week of the academic year can be enough to significantly dampen the beneficial effects of the intervention (Cook et al., 2012; Wu et al., 2021). Ideally, the intervention should also be implemented at a time of transition, when a small initial boost can have long term consequences if it channels students into a different “trajectory” and
social environment. For example, students in England are often placed into ability sets at some point in secondary school (from age 11 onwards) and an intervention that boosts their performance and confidence at the crucial time before “setting” could result in the student being placed in a higher-ability and more academically-orientated set, which could lead to long-term benefits (see Easterbrook & Hadden, 2021; Goyer et al., 2017). Therein lies the essence of the “trigger and channel” approach to understanding the effects of self-affirmation (Ferrer & Cohen, 2019; Sherman et al., 2021), which suggests that a well-timed, psychologically precise intervention can act as a trigger that channels students into a more supportive or nourishing social environment, leading to long-term change.

These different aspects of implementation—the message the intervention portrays, timing at periods of high-threat or key transition points—should be understood not as independent, but as intricately intertwined. If a student considers school a threatening environment, but then comes to believe that their school and teacher care and want to know more about their values, their whole educational environment can become a less threatening place, at least for a while. Optimally, even during a time of reduced threat, some testing situation will occur, and the student’s performance may—without the “chains of threat”—be improved. If the intervention is timed at a point of transition, then this improvement could produce a ripple of positive effects through their intricate social web, igniting a cycle of adaptive potential and supportive recursive effects.

**Moderators**

Another reason for the heterogeneity of effects is that there are undoubtedly moderators of the effectiveness of the intervention that we are currently unaware of. To complicate matters further, these moderators could come at any or all of a range of levels of analysis, accounting for variation between intervention time points, individuals, classes, teachers, schools, neighborhoods, districts, or even nations and/or cultures. Indeed, some moderators have already been identified at different levels of analysis. Individual-level moderators have been identified, including proxies for social identity threat—such as group membership, prior performance (Cohen et al., 2006), and sense of belonging (Layous et al., 2017)—and those based on coding of the written exercises, such as level of student engagement (Borman et al., 2018) and whether the students reflected on feelings of belonging (Shnabel et al., 2013).

At the level of the classroom, variables that seem to capture the quality and sensitivity of the learning environment have been found to moderate the effectiveness of the intervention. For example, in one study, self-affirmation interventions were found to improve the attainment of minority students only in classrooms in which all pupils’ average attainment levels (for reading) improved at an above average rate (Dee, 2015). This is in line with theoretical suggestions that, for long-term effects to occur, the interventions must be implemented in social environments that are sensitive enough to react to small initial boosts in performance so that recursive effects can be initiated (Walton & Yeager, 2020). There is also evidence that the exercises are more beneficial for student attainment when the teachers read what the students have written in the exercises, even when the students are not aware of this (Bowen et al., 2013). If teachers read the exercises, which tend to be self-disclosive in nature, they are likely to feel more connected to their students and may notice and react to small initial boosts in performance and confidence brought about by the interventions, igniting recursive effects. The interventions have also been found to be more effective in classes in which there is a greater density of self-affirmed students (Powers et al., 2015). Self-affirmation exercises have been found to reduce prejudice (Badea & Sherman, 2019), increase
prosocial behaviors (Lindsay & Creswell, 2014), increase students’ trust in teachers, and decrease disruptive behavior (Binning et al., 2019), so that classes in which a higher proportion of students are affirmed may well be more supportive and reactive to the initial boosts in performance and engagement brought about by the reduction in threat produced by the intervention.

At the school-level, moderators have been found that seem to be objective indicators of the degree of stereotype threat that minority students are likely to be experiencing. For example, Borman and colleagues (2018) created a composite measure to indicate how threatening the school context is likely to be for minority students, which consisted of an indicator of the group’s historical performance and the proportion of threatened students in the school (Borman et al., 2018). They found that self-affirmation interventions were more effective in more threatening school contexts. Theoretical arguments have been made that also take account of the prevalent stereotypes in the wider national context within which the intervention is being implemented, which are suggested to moderate the extent that different groups experience threat (Easterbrook et al., 2019; Easterbrook & Hadden, 2021; Sherman et al., 2021). For example, as mentioned above, self-affirmation research in the US tends to focus on improving the performance of African Americans and Latino Americans (Cohen et al., 2006; Sherman et al., 2013), whereas in the UK the focus has been on improving the performance of economically disadvantaged students (Hadden et al., 2020); in Germany, self-affirmations have been found to improve the performance of immigrants (Lokhande & Müller, 2019). The groups that experience threat seem to depend, therefore, on the meaning and value of different social identities in the local context (Easterbrook et al., 2019; Easterbrook & Hadden, 2021).

These moderators demonstrate the importance of understanding the wider context in which self-affirmations are implemented and, in particular, recognizing how the context can determine which groups are likely to be experiencing threat and are thus likely to benefit from self-affirmation interventions (Binning & Browman, 2020; Easterbrook & Hadden, 2021; Manstead et al., 2020). Critically, however, there has yet to be an empirical demonstration that threat moderates (or mediates) the effectiveness of self-affirmation interventions in education, as proxies for social identity threat are often operationalized as objective indicators of group membership (e.g., ethnicity or eligibility for free school meals, but see Celeste et al., 2021; Layous et al., 2017). This is, in part, because there are not yet valid, reliable, and established measures of threat. Developing measures of threat is not an easy task because people may not be aware of the threat, defense mechanisms can distort people’s self-reports of threat (Vignoles, 2014), and there is great heterogeneity in threats across time and place.

A useful alternative, therefore, is to conduct extensive preliminary investigations of the conditions in the local educational context (Binning & Browman, 2020; Easterbrook & Hadden, 2021). Gaining a deep understanding of the existing inequalities and sociocultural factors that are salient in the local educational context may indicate whether the relationship between group membership and educational outcomes is mediated by threat, and hence which groups (if any) may benefit from an intervention that reduces threat (Easterbrook & Hadden, 2021, Easterbrook et al., 2019; Sherman et al., 2021). It may also enable more careful pre-registration analytic plans, which would enable stronger confidence in the moderation results that do occur (Gehlbach & Robinson, 2018). This approach also reflects recent calls for self-affirmation interventions to be tailored to the inequalities in the local context and the specific needs of the groups involved (Binning & Browman, 2020; Sherman et al., 2021).

Within large, multi-school studies, these contextual factors should be measured so that this contextual variation can be demonstrated empirically. Indeed, we suggest that identifying and measuring variables that objectively characterize the local or national context may be a particu-
larly fruitful avenue for researchers to follow (e.g., Celeste et al., 2021; Easterbrook et al., 2019; Manstead et al., 2020), as may using alternative sources to quantify the context, such as aggregating questionnaire responses (Yeager et al., 2019), interviews with teachers (Gehlbach et al., 2016), or computing measures of student discrepancies from their classmates (Benish-Weisman et al., 2020).

So far, then, moderators have seemed to be in one of three categories: indicators of the degree of threat that groups of students are likely to be experiencing, indicators that capture the quality of the learning environment, and indicators of student engagement. To progress the search for moderators that help to explain the heterogeneity of effects, we suggest that an important yet achievable initial objective is to document where the variation in the effectiveness of the intervention lies. Large, multi-school trials enable multilevel analyses to be conducted that take account of variation between individuals and schools (and could be extended to account for clustering of, or within, timepoints, classrooms, and/or teachers; Hox, 2010). Such analyses can allow the relationship between, say, an intervention (versus control) and academic grades to vary across higher-level units (be they classes, teachers, or schools), and then enable attempts to explain this variation using predictor variables measured at the higher level of analysis (such as the proportion of threatened students in the class, ethnicity of the teacher, or rating of the school). Indeed, such approaches have already proved useful (Borman et al., 2018). Large scale trials would enable analyses to be conducted that would estimate the proportion of the variation in relationships between variables that is at each level, and so would provide empirical guidance about which levels of analysis researchers should focus on in the search for potential moderators (see Yeager et al., 2019).

MEDIATORS

While implementation procedures and moderators are likely to account for much of the heterogeneity in the effectiveness of self-affirmation interventions in different contexts, the processes through which the interventions work are of course key. Practically, knowledge of processes would allow the interventions to become ever more refined and targeted; theoretically, they may shed light on yet more reasons for the heterogeneity of effects. Below we describe the processes that have been identified in the education literature.

Self-affirmation interventions have been found to reduce stress (Creswell et al., 2005; Hadden et al., 2020; Sherman et al., 2009), although stress was not found to statistically mediate the effect of the intervention on academic outcomes (Hadden et al., 2020). Perhaps this is not surprising, given the many dynamic and interactive processes that are thought to be ignited by the interventions and that ripple through the social environment. Nevertheless, a reduction in stress does seem likely to benefit pupils’ educational experience, as long as the resultant level of arousal or threat is not too low or too high (Binning & Browman, 2020; Yerkes & Dodson, 1908).

Two other, related processes through which self-affirmation interventions have been found to affect academic achievement are psychological un tethering and a broadening of perspective. The predictions that gave rise to these findings are based on the assumption that threats tend to absorb people’s attention: people become fixated on the threat and whatever may confirm or compound it (Critcher & Dunning, 2015). This results in a heightened sense of vigilance that causes people to interpret innocuous events as threatening and confirming of their fears of not being valued or not fitting in within education. Self-affirmation broadens people’s perspective beyond the immediate threat, which effectively dilutes the impact of the threat on attention and cognition and brings other sources of self-integrity into awareness (Critcher & Dunning, 2015). This was first
demonstrated in the context of educational settings by Sherman and colleagues (Sherman et al., 2013; see also Schmeichel & Vohs, 2009; Wakslen & Trope, 2009), who found that threatened students (using daily survey assessments) in the affirmation condition used higher-level construals than threatened students in the control condition, and that the correlations between measures of daily adversity and measures of identity threat and belonging within school were only significant among Latino American students in the control condition. These findings suggest that threatened students interpret commonplace daily adversity in terms of threat—indicative of heightened vigilance—and that affirmation can reduce this tendency, partly by increasing the construal level. Thus, self-affirmation can “untether” the self from the threat. This process of untethering is demonstrated in two of the papers in this special issue (Binning et al., 2021; Manke et al., 2021).

Trust in teachers also seems to be an important process through which self-affirmation interventions may operate. In one study, trust in teachers was found to increase for all students after multiple affirmation interventions were administered over two school years (Binning et al., 2019; see also Sherman & Cohen, 2006). The message that the intervention portrays to students—that their teachers care about who they are—can instill a sense of trust that their teachers and school have their best interests at heart. However, while trust is undoubtedly an important outcome in its own right, there is, as yet, no evidence that trust mediates the effect of the affirmation on academic outcomes. Nevertheless, trust does seem to be a theoretically plausible mediator, particularly given the importance of trust to the teacher-pupil relationship (Yeager et al., 2013). Furthermore, these findings were found after eight self-affirmation interventions over two full academic years. It may be that such a density of self-affirmation interventions may ignite processes additional to those sparked by fewer interventions, such as encouraging students to spontaneously self-affirm when they experience threat. Individual differences in tendency to report engaging in spontaneous self-affirmation have been found to produce similar effects to self-affirmation interventions in the health domain (Lannin et al., 2021; Harris et al., 2019), so this may be a process that plays a role in education. Indeed, one study reported that the improvement in GPA for Latino American students in the US was partially explained by an increased tendency among those students to spontaneously self-affirm when threatened (Brady et al., 2016).

There are, no doubt, many processes through which self-affirmation interventions operate that have yet to be identified. These may include processes that affect the teachers themselves indirectly through observing and teaching classes of students who have been self-affirmed. For instance, Morgan and Atkin (2016) found that teachers who had been self-affirmed reported reduced anxiety and increased wellbeing; there may be positive effects on student outcomes of affirmed teachers and vice versa. In order to test such ideas, methods additional to those that have been used to date may be helpful, such as interviews, social network analyses, and diary studies. There is also the possibility that the interventions operate through different processes for different groups experiencing different psychological states (Çetinkaya et al., 2020), but there have not been any investigations of moderated mediation or mediated moderation in the self-affirmation literature that we are aware of.

OVERVIEW OF THE SPECIAL ISSUE

Given the gaps in our understanding about the heterogeneity of effects and the processes through which self-affirmation interventions operate, as well as the increased interest in implementing self-affirmation interventions at scale, we considered a special issue of the Journal of Social Issues on Self-affirmation Theory in Educational Contexts to be both timely and important. Below, we
Self-affirmation theory in educational contexts

outline the contributions included within this special issue and the contribution they make to the field.

The issue is divided into two sections: (1) heterogeneity and moderators of the effectiveness of affirmations, and (2) processes, adaptations and integrations.

The first section contains three papers that illuminate heterogeneity in the effectiveness of self-affirmation interventions in education and some moderators of its effectiveness. Wu et al. (2021) report the first meta-analysis of self-affirmation interventions in education. They find that the interventions are effective at increasing the attainment for identity-threatened groups (but not non-threatened groups), that the effect size is heterogeneous, and, importantly, test several moderators of its effectiveness. Specifically, they report that the effect size is larger when there is a larger attainment gap between threatened and non-threatened groups in the control condition, in schools that are likely to have more financial resources, when the performance outcomes are more distal from the intervention, and when the affirmation was delivered as a normal classroom assignment from a teacher. This work quantifies many of the theoretical arguments about why the effects of affirmations vary.

Smith, Rozek et al. (2021) provide a targeted empirical investigation of one important aspect of implementation: whether teachers or researchers distribute the exercises. In a field experiment, they demonstrate that affirmations delivered by teachers—but not those delivered by researchers—increased students’ perceptions that teachers at their school provide care and support to students and are interested in students’ lives beyond their academic work. Students who were given the intervention by their teachers also wrote more than those who were given the intervention by researchers, suggesting that they were more engaged with the exercises. Thus, Smith, Rozek, and colleagues experimentally demonstrate the importance of implementing self-affirmation interventions in ways that portray a caring message to students.

In the final paper in the first section, Perry et al. (2021) present a study in which they implement an online affirmation to students in medical schools in the US, aiming to increase Black students’ wellbeing, sense of belonging and career self-efficacy. They found racial inequalities in sense of belonging, suggesting stereotype threat may contribute to these inequalities. They also found that, while the affirmations tended to weaken Black students’ own perceived competitiveness, it also reduced the likelihood that they opted to switch their residency over time. Perry and colleagues provide a detailed discussion of why this might be so and offer several important lessons for other researchers working in the area.

The second section presents research that demonstrates the processes through which self-affirmation interventions are effective, as well as work that adapts self-affirmation interventions for different target groups and/or contexts. Binning and colleagues (2021) show that repeated self-affirmation exercises lessened a downward trajectory in academic performance for students in the affirmation condition, regardless of whether they were thought to be experiencing threat. Furthermore, they found that, although the affirmation did not lessen the experience of threat, it did attenuate the association between self-reported threat and academic performance. Binning et al.’s findings suggest some intriguing possibilities regarding the relevance of self-affirmation to different students, and we hope it will motivate future studies that investigate how the intensity and frequency of the exercises may moderate its effects.

Celeste et al. (2021) report the results of a study in which they implement both a self-affirmation and a new dual-identity intervention in a highly diverse school in England, presenting both as a new school policy that the school wanted students’ feedback on. These interesting adaptations led to some intriguing results. They found that Black students in the dual-identity condition performed better than non-Black students, but that Black students performed worse than non-Black
students in the self-affirmation condition. Furthermore, a direct measure of stereotype threat partially mediated the effects. These findings generate a range of important questions that we hope future researchers will investigate, and provide a much-needed demonstration of the role of threat.

Hecht and colleagues (2021) report the results from a study of an adaptation of self-affirmation interventions in a different context and for different groups. They adapt the exercises to encourage students in US colleges to write about both independent and interdependent values and find that this reduced self-reported cultural mismatch and boosted achievement for first generation college students. This demonstrates a key mechanism through which affirmations may operate in certain settings for some groups—by reducing cultural mismatch—and provide an important demonstration of the importance of being sensitive to the context and the cultural values of those involved in the intervention.

In the final paper in the second section, Manke and colleagues (2021) present a novel adaptation to self-affirmation exercises by implementing a brief self-affirmation intervention via text message before a stressful event. They found that those who received the affirmation text messages before the stressful event reported greater belonging compared to those in the control condition. Thus, Manke and colleagues provide an initial demonstration of a novel and efficient way of implementing and potentially upscaling self-affirmation interventions.
Finally, the pioneer of both stereotype threat and self-affirmation theory, Claude Steele, provides the concluding piece to the special issue, reflecting on the interplay between theoretical and applied questions raised by this body of research and the contributions—past, present, and potential—of self-affirmation theory to understanding the societal issue of educational disparities.

Table 1 presents a list of the authors of the articles, the samples they employed in their research, and the key question of their research. We would like to close by thanking these authors for these contributions, along with the reviewers whose thoughtful feedback helped shape these articles. We would also like to note the range of educational contexts employed in these papers—middle schools, high schools, universities, and graduate medical schools. With each research endeavor, the authors tested and extended self-affirmation theory, and in so doing, raised new questions for the field and for society. Collectively, this body of work demonstrates how intractable and difficult social issues can serve as a testbed for scientific inquiry. This research speaks to the hope that such scientific inquiry can make a positive difference in the lives of students of all ages as they pursue their educational goals.

ACKNOWLEDGEMENTS
The authors received joint funding from the European Association of Social Psychology and the Society for the Psychological Study of Social Issues to host a meeting at the University of Sussex, UK in June 2019 entitled, “Self-affirmation in Education: Why, when, and how are self-affirmation interventions effective at reducing educational inequalities?” It was this meeting that formed the basis for this special issue.

ORCID
Matthew J. Easterbrook https://orcid.org/0000-0002-9353-5957
Peter R. Harris https://orcid.org/0000-0003-4599-4929

REFERENCES


OECD. (2015) Helping immigrant students to succeed at school - and beyond. doi.org/15189667


Yerkes, R.M. & Dodson, J.D. (1908) The relation of strength of stimulus to rapidity of habit-formation in the kitten. *Journal of Comparative Neurology and Psychology*, 18(5), 459–482. [https://doi.org/10.1037/h0073415](https://doi.org/10.1037/h0073415)

**AUTHOR BIOGRAPHIES**

**Matthew J. Easterbrook** is a Senior Lecturer in Psychology at the University of Sussex. He researches self, identity, and group processes in a variety of settings, and designs, implements, and evaluates social psychological interventions that aim to benefit disadvantaged groups and reduce inequalities. He has taught classes on social psychological interventions in education at national and international level, is an author of the first investigation into the efficacy of self-affirmation interventions to reduce achievement gaps in England (https://doi.org/10.1111/bjep.12291), and has led the largest evaluation to date of whether self-affirmations can reduce achievement gaps (https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/writing-about-values/). He has been an associate editor of European Journal of Social Psychology and is currently co-chair of Sussex Psychology’s Equality, Diversity, and Inclusion committee.

**Peter R. Harris** is Emeritus Professor of Psychology at the University of Sussex, where he led the Social and Applied Psychology Subject Group. He received his first degree from UCL and his PhD from the LSE. He has a long-standing research interest in self-affirmation, particularly as applied to the promotion of more open-minded processing (e.g., of information about health-risks or pro-environmental behaviour) and the understanding of message resistance. He also undertakes research into individual differences in self-affirmation. He is one of the founding members of the self-affirmation research group at Sussex, which brings together
researchers interested in self-affirmation as applied to health, education, and the environment (https://www.sussex.ac.uk/psychology/sarg/index). He is a Fellow of the Association for Psychological Science and Associate Fellow of the British Psychological Society.

**David K. Sherman** is Professor of Psychological & Brain Sciences at the University of California, Santa Barbara. He received his PhD from Stanford University and was a post-doctoral fellow in health psychology at UCLA. His research centers on how people cope with information and events that threaten the self. Along with his colleagues, he received the 2014 Cialdini Award for the publication that best explicates social psychological phenomena through field research for their 2013 article on self-affirmation, academic performance, and motivation under identity threat. He served as President of the International Society for Self and Identity (2016–2019), co-taught the class “Interventions” at the Summer Institute for Social and Personality Psychology (2017), and is Editor of Personality & Social Psychology Review.

**How to cite this article:** Easterbrook M. J., Harris P. R. & Sherman D. K. Self-affirmation theory in educational contexts. *Journal of Social Issues* 2021;1–19.