

A Replicable Identity-Based Intervention Reduces the Black-White Suspension Gap at Scale

Geoffrey D. Borman 

Arizona State University

Jaymes Pyne 

Stanford University

Christopher S. Rozek 

Washington University in St. Louis

Alex Schmidt

Christensen Associates

Nationally, educators suspend Black students at greater rates than any other group. This disproportionality is fueled by stereotypes casting Black students as “troublemakers”—a label students too often internalize as part of their identities. Across two independent double-blind randomized field trials involving over 2,000 seventh graders in 11 middle schools, we tested the efficacy of a brief intervention to buffer students from stereotypes and mitigate the racial suspension gap. The self-affirmation intervention helps students access positive aspects of their identities less associated with troublemaking in school.

GEOFFREY D. BORMAN is the Alice Wiley Snell Endowed Professor at Mary Lou Fulton Teachers College, Arizona State University, H.B. Farmer Education Building, 1050 South Forest Mall, Tempe, AZ 85281; email: gborman@asu.edu. His research interests include the design and analysis of large-scale randomized trials, racial/ethnic stereotypes, and how students’ social-psychological mindsets can foster or hinder school success.

JAYMES PYNE is a sociologist and a research associate in the Graduate School of Education at Stanford University, using experimental, quasi-experimental, and correlational designs to understand social policy and theory. He writes on how the U.S. education and criminal justice systems produce punishment, social engagement, and inequality.

CHRISTOPHER S. ROZEK is an assistant professor in the Department of Education at Washington University in St. Louis. Using laboratory and field experimental studies, along with longitudinal designs, he examines how societal and social contexts shape the psychological factors associated with students’ motivation and success.

ALEX SCHMIDT is a researcher at Christensen Associates and completed his PhD in sociology at the University of Wisconsin-Madison. His academic research explores the effects of educational and social policies on academic outcomes.

Confirmed in both trials, treatment effects cut Black-White suspension and office disciplinary referral gaps during seventh and eighth grade by approximately two thirds, with even greater impacts for Black students with prior infractions.

KEYWORDS: school discipline, social-psychological interventions, self-affirmation, identity threat

Suspension, or the temporary removal of students from the regular educational program for real or perceived violations of school rules or policies, is a widely used and contentious disciplinary practice in U.S. public schools. The use of exclusionary discipline practices has become an issue of social justice, both with respect to historically discriminated groups and universally under the belief that no student should be denied the opportunity to access a free public education. In particular, Black¹ students are at greater risk than any other group for being suspended (Gregory et al., 2010; U.S. Department of Education Office for Civil Rights, 2016) and this racial *disciplinary gap* can interfere with Black students' opportunities to learn, thus explaining a sizeable portion of national achievement gaps (Morris & Perry, 2016; Pearman et al., 2019).

Suspended students are subsequently at greater risk of high school non-completion (Gregory et al., 2010; Mizel et al., 2016; Noltemeyer et al., 2015), with Black students in particular citing having been expelled or suspended too often as a reason that they left high school without graduating (Jordan et al., 1996). The notion of the “school-to-prison pipeline” has been fueled by the strong relationship between school discipline problems and subsequent criminal activity and incarceration during adolescence and adulthood (Fabelo et al., 2011; Shollenberger, 2014; Wolf & Kupchik, 2017). Yet teachers, parents, and the public generally favor the use of strict and exclusionary policies like suspensions to correct misbehavior and increase safety in schools (Bushaw & Lopez, 2010; Robbins, 2008).

In this article, we report on an intervention intended to mitigate the racial suspension gap by helping students cope with negative labels placed on them by their school and society, so they can instead focus on the many positive qualities of their identities. Across two independent yearly cohorts totaling over 2,000 seventh graders in all 11 middle schools from a Midwestern school district, we conducted two double-blind randomized field trials of an intervention that builds on decades of social-psychological theory and research concerning *self-affirmation* (Cohen & Sherman, 2014; Steele, 1988). This intervention involves a series of expressive writing exercises asking students to reflect on positive aspects of their identities less associated with the “troublemaker” label in school. Self-affirmation theory posits that individuals are motivated to maintain a positive overall self-view. When one's self-competence is threatened, it helps to have opportunities to reflect on sources of self-worth (e.g., being a family member, enjoying sports, being creative, or

having a sense of humor) beyond the threatened aspect of the self (in this case, being stereotyped as a troublemaker in school). Those opportunities to reflect on positive aspects of one's identity can buffer students from the threat and help restore an overall sense of "moral and adaptive adequacy," or an outlook that one is good, successful, and able to control important life outcomes (Cohen & Sherman, 2014; Steele, 1988).

We posit that the self-affirmation intervention in question helps provide Black students with psychological resources to manage social identity threats and contend with internalized racial oppression—or the psychological appropriation of prevailing biases and stereotypes in society (David et al., 2019). Self-affirmation may buffer students from these threats by allowing them to concentrate on their positive identities (e.g., artist, athlete, friend, daughter) rather than internalizing racial oppression and seeing themselves as troublemakers. This shift in identity focus may help break one link in the vicious cycle of apprehension, mistrust, and punishment Black students endure in our nation's schools (Okonofua, Walton, & Eberhardt, 2016).

Understanding the Suspension Gap

Approaches to school discipline are generally based on deterrence theory, which assumes that individuals respond rationally and positively to rules and expectations because they are fearful of the consequences that might result if they defy them (Gibbs, 1975). However, Black students receive disproportionate (Losen, 2011) and unusually harsh discipline from school authority figures (Fisher et al., 2000). As a result, Black students perceive limited support from educators in their schools and consequently view the dispensation of disciplinary sanctions as unfair. This mistrust can lead to defiance rather than deterrence (Kupchik, 2010; Way, 2011). Yeager et al. (2017) recount such a process empirically, finding that Black middle-school students develop a growing awareness of racial bias in school discipline and subsequently lose trust in their teachers. In turn, that loss of trust predicts increases in discipline incidents.

National data from the U.S. Department of Education Office for Civil Rights (2016) indicate that 18% of Black boys and 10% of Black girls receive one or more out-of-school suspensions each year, compared with only 5% of White boys, 2% of White girls, 7% of Hispanic boys, and 3% of Hispanic girls. Why are Black students punished more often and more severely than all other students, including those from historically marginalized ethnic minority groups (Gopalan & Nelson, 2019; Gregory et al., 2010)? Some explanations have focused on deficit-based hypotheses that cast Black students as the problem, implying that Black students misbehave in school because they may lack critical social-emotional skills, such as empathy, self-control, or compassion (see Okonofua, Walton, & Eberhardt, 2016, for a discussion of these explanations).

Other explanations point to explicit and implicit biases held by teachers and other school staff. Disproportionality is clearly shaped by long-documented biases within U.S. society, which tend to stereotype Black people as criminal and violent (Gabbidon et al., 2002; Levy et al., 1998). For example, in a recent laboratory study (Goff et al., 2014), raters perceived Black children older than 10 years as significantly less innocent than their same-aged non-Black peers. Raters in the study considered Black children aged 13 to 14 years—the ages of the majority of students in the current study—to have the same innocence of non-Black children aged 17 to 18 years. These outcomes are highly relevant in school-based disciplinary and criminal-justice contexts, where biased perceptions of innocence have severe consequences, including chance and length of incarceration. Field research confirms this in middle schools. Teachers perceive Black students as more defiant, disrespectful, and rule-breaking than their peers (Skiba et al., 2002; Wentzel, 2002) and are more apt to label a misbehaving Black middle-school student as a troublemaker who deserves to be disciplined (Goff et al., 2014; Okonofua & Eberhardt, 2015).

Rather than placing blame on students or teachers, Okonofua, Walton, and Eberhardt (2016) point to a “vicious cycle” of breakdowns in teacher-student relationships. In this cycle, teachers are motivated by the basic goals of teaching and inspiring students. At the same time, teachers stereotype Black students as potential troublemakers and worry that these students could prevent them from achieving their teaching goals. These stereotypes lead teachers to attribute misbehavior among Black students as enduring features of their identities, and thus especially problematic. In turn, teachers sanction Black students more frequently and more harshly than other students in the hope that doing so will preserve order in their classrooms.

Black students, like their peers, come to school valuing education and aspiring to learn and develop (Mickelson, 1990). However, they also worry that teachers are biased against them and will treat them unfairly (Steele, 1997). When Black students are punished, it provides confirmation that their teachers really are unfair and are undermining their educational goals and values. This confirmation of their fears of bias eventually means that they lose trust in their teachers, feel like they do not fit in at school, and experience threats to their sense of identity as a good student (Okonofua, Walton, & Eberhardt, 2016; Pyne, 2019).

Such identity threats lead students to further misbehave (Belmi et al., 2015; Gregory & Weinstein, 2008), which in turn, reinforces teachers’ beliefs that the students are troublemakers and results in additional harsh disciplinary tactics. The cycle repeats and escalates with additional teacher-student encounters, and students’ identities as “troublemakers” become more deeply entrenched over time in the minds of both teachers and the students themselves. Because Black students face the “double jeopardy” of being both stereotyped as troublemakers and being more frequently disciplined than other

student groups, they are uniquely at risk to experience this vicious cycle during middle school (Okonofua, Walton, & Eberhardt, 2016).

Intervening to Close the Suspension Gap

Recent events, such as the deaths of George Floyd, Breonna Taylor, and others at the hands of the police, have heightened national awareness of structural racism in U.S. society and its institutions. Recognizing that schools and teachers are not immune to bias and racism, many school districts have engaged in efforts to train staff to avoid racial biases that could affect their relationships with students (Sparks, 2020). However, few examples exist of research-based and scalable interventions that can sustainably reduce individuals' prejudices and implicit biases. For example, a review of the literature by Cameron and Turner (2010) reveals some laboratory-based interventions having short-term benefits, but a paucity of evidence supporting the "real-world" efficacy of scalable field-based interventions. A more recent meta-analysis of 492 studies suggests that interventions generally produce weak short-term reductions in implicit bias, even weaker reductions in explicit bias, and trivial changes in behavior (Forscher et al., 2019). In some cases, initiatives to prevent prejudice, reduce implicit bias or embrace diversity backfire and actually increase stereotyping and prejudice (Duguid & Thomas-Hunt, 2015; Legault et al., 2011) or cause the nonstigmatized majority group to feel threatened and defensive (Dover et al., 2016). Though efforts must continue to root out and end the long legacy of structural and individual racism, considerable research evidence suggests that this is a highly complex and difficult endeavor.

In addition to intervening with teachers and school staff to prevent prejudice or promote diversity, one recent and promising example for reducing the suspension gap instead focuses on teachers' mindsets about discipline and their relationships with students. Okonofua, Walton, and Eberhardt (2016) theorize that when teachers engage in punitive responses to student misbehavior, they alienate students and incite them to respond with the oppositional behaviors that the punishment is intended to prevent. Brief online intervention modules developed by Okonofua, Paunesku, and Walton (2016, p. 5223) suggest to participants that "a teacher who makes his or her students feel heard, valued, and respected shows them that school is fair and they can grow and succeed there." These ideas are reinforced through stories from students, helping encourage teachers to use *empathic discipline*, which involves attempting to understand the root causes for misbehavior while maintaining trusting student-teacher relationships. This intervention, consisting of two online modules totaling 70 minutes, halved the number of suspensions for Black students in intervention teachers' classrooms and also improved relationships between teachers and students (Okonofua, Paunesku, & Walton, 2016). This teacher-focused intervention appears to

be one promising approach for breaking the vicious cycle of punishment by helping teachers adopt a different mindset about classroom discipline.

Student-focused approaches may also help students of color contend with the burdens of society's pervasive biases that play out through school relationships. A focus on students does not relieve teachers of the responsibility to end the vicious cycle of punishment in schools, but instead empowers students to navigate the threatening student-teacher relationship, which fuels greater interpersonal conflict that distracts from their educational progress. It is reasonable to acknowledge that when teachers and other school staff treat students poorly or unfairly, students respond negatively to that person and environment. But this reality does not suggest we should blame vulnerable students for their reaction to harsh and unfair treatment. Rather, student-focused interventions can help alleviate historically underserved and stereotyped students' anxieties about school and provide an avenue for them to feel empowered to focus on positive aspects of their identities and lives. That shift in focus may help temper the psychological toll of mistreatment by educators and, ultimately, improve their academic outcomes.

Some applied social-psychological research emphasizes the power that differing attitudes or mindsets have on how students view themselves, interact with others, and succeed in education and beyond (Dweck, 2006; Easterbrook & Hadden, 2021; Walton & Wilson, 2018; Yeager & Walton, 2011). Relatively simple *mindset interventions*, which deftly apply powerful social-psychological theories in real-world situations, can change how students interpret and respond to challenges in school and put in motion positive recursive cycles that increase academic success over time (Cohen & Garcia, 2014; Paunesku et al., 2015). Because these interventions have produced relatively large recursive academic impacts that do not appear to line up with their perceived simplicity and brevity, these social-psychological interventions can have seemingly "magical" properties (Yeager & Walton, 2011).

One such student-based social-psychological approach used to stem educational disparities is *self-affirmation*. Self-affirmation interventions seek to improve school outcomes among students potentially threatened due to stereotypes about their academic abilities by asking those students to focus on and affirm positive aspects of their identities, such as being a good brother, daughter, athlete, artist, and so on. A seminal field trial for seventh-grade students in a traditional sixth- to eighth-grade middle school setting revealed that self-affirmation reduced the Black-White grade point average (GPA) gap by 40% through middle school (Cohen et al., 2009). More recent longitudinal studies have shown that these promising initial impacts of self-affirmation can endure beyond middle school on Black students' GPA trajectories through high school (Borman et al., 2018; Borman et al., 2021), on-time high school graduation rates (Borman et al., 2021), and probability of attending college (Goyer et al., 2017).

Though there is promising evidence of reductions to achievement gaps, relatively little research exists concerning how self-affirmation interventions

might help mitigate the disciplinary gap. There are several reasons to believe that self-affirmation interventions can affect students' disciplinary outcomes in addition to their academic outcomes. First, the work of Shnabel et al. (2013) suggests that the direct effects of self-affirmation are explained by an increased sense of social belonging, which may help temper Black students' feelings of marginalization in school. A second mechanism is "affirmation as perspective," in which self-affirmations "expand the contents of the working concept—thus narrowing the scope of any threat" (Critcher & Dunning, 2015, p. 4). The working concept is represented by the salient identities that help form one's overall beliefs about "who I am" in consciousness at any point in time. When aspects of identity are threatened, the working self-concept constricts and amplifies the negative experiences of the relevant threat at hand. When Black students are threatened by the troublemaker identity in school, self-affirmation can expand the scope of self-concept, reduce attention to the troublemaker identity threat, and mute students' stress responses.

Relatedly, self-affirmation also helps prevent disciplinary involvement by broadening the perspective or "level of construal" through which students assess salient threats (Cohen & Sherman, 2014). When students are threatened, they construe negative events in narrow, concrete terms. Self-affirmed students instead report higher levels of construal and report less adversity in school related to their social identity than do threatened students who are not affirmed (Sherman, 2013). Finally, some work suggests that the intervention also boosts the self's regulatory function. A threatened self-identity depletes cognitive resources required for adaptive coping, self-regulation, and performance under pressure (Logel & Cohen, 2012). Self-affirmation restores depleted cognitive resources and supports greater cognitive performance in demanding circumstances (Schmeichel & Vols, 2009).

Despite these theoretical justifications for why self-affirmation helps address disparities in school disciplinary involvement, more empirical support for them is needed. Though Goyer et al. (2019) include self-affirmation as one of three interventions delivered to students that successfully address disciplinary outcomes among negatively stereotyped middle-school boys, we know of only two studies reporting the independent effects of self-affirmation interventions on disciplinary referrals in secondary schools. One study by Dee (2015) examined the effects of self-affirmation for students of color on a range of academic and behavioral outcomes, including absences and disciplinary referrals, but found no impacts.² More recently, Binning et al. (2019) conducted a 3-year field trial on self-affirmation and student behavioral conduct, across Grades 6, 7, and 8, at an ethnically diverse middle school. The authors found that the eighth-grade students in the affirmation condition were disciplined at a 69% lower rate than control students. However, the intervention effect did not differ across racial/ethnic groups and the study did not examine whether students with a history of disciplinary involvement prior to

intervention—and who may have been more likely to be labeled as trouble-makers—benefited more than their peers with no prior disciplinary history.

In sum, though the evidence is limited, self-affirmation offers a student-focused way to intervene on the bias and discrimination that Black students too often face by helping empower them to focus on the positive aspects of their identities. Helping Black students see past the negative attributions placed on them and instead focus on positive aspects of their identities may help short-circuit the degrading teacher-student relationship and the vicious cycle that feeds disciplinary and suspension gaps.

The Current Study

We investigate how a replicable self-affirmation intervention, focused on empowering Black students to celebrate positive aspects of their identities, might stem the Black-White suspension gap. To examine the potential of a self-affirmation intervention to break the vicious cycle of punishment, we test it at scale with two successive yearly cohorts of seventh graders across all 11 middle schools in a midsize urban school district, the Madison (Wisconsin) Metropolitan School District (MMSD). The scale of this study is important because it demonstrates the feasibility of implementing social-psychological interventions widely in “real-world” educational settings and without heavy researcher oversight, as would happen if fielded by school districts interested in offering the writing exercises as routine practice in their schools.

Further, our replication across two independent cohorts of seventh-grade students provides greater confidence than a single study that the intervention can be fielded with consistent and policy-relevant benefits. This within-study replication is particularly important in this case because a field trial within the same Madison schools that are the focus of the study reported here revealed strong and enduring impacts on GPA outcomes within one cohort (Borman et al., 2018; Borman et al., 2021), but no such evidence in a second independent cohort (Hanselman et al., 2017). Thus, providing evidence of replicated self-affirmation impacts on disciplinary outcomes across the two independent Madison-based cohorts reported here would be of particular consequence.

In both cohorts in the current study, students are randomly assigned to complete intervention or control writing exercises, which are virtually identical in appearance and demand equal amounts of reading and writing, three to four times during their seventh-grade school year. The materials and administration, described in greater detail in the next section, are designed to replicate prior studies of this intervention that focus on academic achievement (e.g., Cohen et al., 2009). Across the two cohorts in 11 middle schools, a total of more than 2,000 students—over 400 of whom were Black—are included in this study. Teachers, rather than researchers, implemented the interventions during each school year.

Based on prior studies of this intervention's academic impacts in middle-school contexts, we expect that the positive effects of self-affirmation are concentrated among students most likely to suffer from stereotypes that threaten their identity. Therefore, we hypothesize that self-affirmation reduces suspensions for Black students in particular since they are both disproportionately disciplined and, independent of their discipline history, are stereotyped by teachers as troublemakers (Okonofua, Walton, & Eberhardt, 2016). We expect the intervention to be less beneficial for students who are less often disciplined and less often stereotyped as troublemakers in this context (i.e., White, Latinx, and Asian students).

We further hypothesize that Black students who were suspended during their first year in middle school (prior to the seventh-grade intervention) are particularly at risk of being stereotyped as troublemakers. Prior experimental work by Okonofua and Eberhardt (2015) suggests that after only two prior disciplinary incidents, teachers are more apt to see subsequent incidents by a Black student as indicative of a problematic pattern and choose to discipline the Black student more severely than a similar White student. As a result, Black students with prior infractions not only must contend with stereotypes concerning their racial group but also are burdened by a history of disciplinary incidents, which together conspire to place them at an even greater risk for being labeled as a troublemaker.

Method

Participants

Our intent-to-treat (ITT) sample includes 2,328 seventh-grade students consented and assented by the second writing exercise across two waves of the self-affirmation intervention. The first cohort contains 1,055 student observations in all 11 schools in the district (64% of the district population) and the second cohort contains 1,280 additional unique student observations in the same schools a year later (73% of the district's seventh-grade population). The district has provided administrative data for each of these students, including their background information (i.e., race/ethnicity, gender, special education status, English learner status, and free or reduced-price lunch eligibility status) and disciplinary records that include suspensions imposed throughout their time in middle school (i.e., sixth through eighth grade). Attrition results from three main missing data elements. First, 97 students are missing measures of cumulative seventh- and eighth-grade suspensions. An additional 63 are missing a sixth-grade preintervention measure of suspensions. A final 19 students are identified as members of racial/ethnic groups too small for inclusion in the race-by-intervention analysis (i.e., Native American and Pacific Islander students).

After data attrition, the analytic sample includes 2,149 students: 963 from the first cohort of students and 1,186 from the second cohort of students. Within the

analytic sample, 50% are female, 53% are White, 19% are Black, 17% are Latinx, and 11% are Asian. Attrition does not vary meaningfully between control and treatment groups in the full sample or among racial subgroups, and balance for the baseline covariates is adequate for the full sample and among racial subgroups. The attrition rate from the ITT sample to the analytic sample is 8% and we identify no differences in rates of missing data between the treatment and control groups for the full sample ($\chi^2 = 0.34, p = .56$), the Cohort 1 sample ($\chi^2 = 0.01, p = .92$), the Cohort 2 sample ($\chi^2 = 0.49, p = .49$), nor within subsamples of students who are Black ($\chi^2 = 0.08, p = .78$), White ($\chi^2 = 0.05, p = .82$), Asian ($\chi^2 = 0.28, p = .60$), and Latinx ($\chi^2 = 0.23, p = .63$). Both overall and within Black and White student subgroups, statistical comparisons suggest baseline equivalence between the intervention and control groups on all background variables, the preintervention measure of suspensions, and academic achievement. We statistically adjust for any minor differences between the intervention and control groups in our analytic models by controlling for pretreatment covariates. See the Supplemental Materials in the online version of the journal for a discussion on sample attrition and sample balance statistics.

Procedure

The research team implemented the study across 11 racially diverse middle schools, representing all middle schools in the district (see Supplemental Table S1 in the online version of the journal for school-level racial and ethnic composition and baseline Black-White suspension and office disciplinary referral [ODR] gaps). Our materials and procedures closely replicate those previously used in middle schools by Cohen et al. (2006, 2009). In addition to this process of direct replication, we conducted a within-study replication across two independent cohorts of students.

Institutional review board review and approval were provided by the University of Wisconsin-Madison Education and Social/Behavioral Science Institutional Review Board (ID No. 2012-1055-CR007) and the study procedures were also reviewed and approved by the MMSD Education Research Committee. Parents were consented and students were assented at the beginning of the school year. After the consent and assent process was completed, students were randomly assigned to condition, blocking on school to ensure that half of the students in each school received the intervention materials and half in each school received the control materials. The team distributed three to four writing exercises over the course of the seventh-grade year in either language arts or homeroom classes (depending on each school's preferences). The first cohort involved 50 teachers in 80 classrooms and the second cohort involved 44 teachers in 77 classrooms. Since prior research emphasized the importance of the timing of self-affirmation interventions (Critcher et al., 2010), exercises were given to students within the first few weeks of the school year in order to prevent the early focus on negative identities.

The exercises were also administered later in the school year around stressful events (e.g., high-stakes tests), which trigger identity threats for negatively stereotyped students in school.

Teachers, who were blind to randomization and the true purpose of the exercises, distributed intervention materials as if they were normal classroom activities. The intervention group students received the self-affirmation exercise while control group students received a neutral writing exercise. All students remained blind to condition and to the hypotheses of the study. Students not involved in the study were given an alternative assignment that looked like the intervention and control exercises. While the exact wording of the exercises changed slightly over the course of the school year to maintain students' interest, the general theme of the prompts remained unchanged. Students took approximately 15 to 20 minutes to complete each exercise and wrote an average of 70 words responding to each exercise prompt. Time of administration and word count did not vary meaningfully between control and intervention conditions in either cohort (see Hanselman et al., 2017 for more details).

Intervention Exercise

Similar to the materials used by Cohen et al. (2006, 2009), all exercises had a general cover page. The first page of the self-affirmation exercise contained a list of values (e.g., friends, family, sports, creativity) and asked students to choose three of the listed values most important to them. On the second page, a prompt asked students to write about why the values they selected were personally important. Though the exercises were self-administered by students, teachers received a brief training of approximately 30 minutes prior to implementation. Teachers were also provided written materials, including a general script, a list of students' frequently asked questions and the research team's suggested responses to those questions. The training and materials helped teachers emphasize to their students that the exercises were a free-writing activity that would not be evaluated or graded, and that students should write as much as they like, and not worry about spelling or grammar.

Control Exercise

The control group students received a neutral, nonaffirming exercise with an identical cover sheet and an overall format that mirrored the intervention exercise. The first page of the exercise contained the same list of values as in the intervention exercise, but instead asked students to choose three of the listed values that were *not* important to them but could be important to someone else. The second page asked students to explain why these values could be important to someone else, and again emphasized the nonevaluative nature of the exercises. Control and treatment students completed the exercises during the same class period and all students were provided approximately 15 to 20 minutes to complete their responses.

Dependent Variable

The main outcome of interest is the total number of suspensions students received over the last 2 years of middle school, which come from school administrative suspension records. We consider data on students' suspensions for both seventh and eighth grades because previous studies of self-affirmation show lasting effects after the intervention (Borman et al., 2021; Cohen et al., 2009; Goyer et al., 2019). These data allow us to investigate whether any disciplinary benefits of self-affirmation persist for the remainder of middle school, both within and across Grades 7 and 8.

The suspension data are positively skewed; most students have no suspensions over the course of seventh and eighth grades, while very few have over 10. To account for the potential influence of outlier cases, we calculate two adjusted measures of suspensions. In our main analyses, we top-code suspensions at three standard deviations above the mean (rounded down to the nearest integer, which results in a top-coding of seven suspensions). The substantive conclusions are robust across specifications, including when top-coding the top 5% of suspension counts and when using the full range of suspensions (results not shown).

Independent Variables

The main independent variables of interest include each student's dummy-coded intervention status, with the control group coded as "0," and a categorical variable for race and ethnicity that includes codes for White, Black, Latinx, and Asian students, with "White" as the reference category. To increase the precision of treatment effect estimates, we also include a vector of control variables in our models. Gender, eligibility for free and reduced-price lunch (as a proxy for family income), special education designation, and limited English proficiency status are all dummy coded (0/1). The baseline categories are male, no free or reduced-price lunch eligibility, no special education designation, and no English learner designation.

Analytical Methods

We model the cumulative number of seventh- and eighth-grade suspensions in a hierarchical linear model framework with each student (i) nested within one of 11 schools (j):

$$Y_{ij} = \alpha_{ij} + \beta_1(Treatment_{ij}) + \beta_2(Race_{ij}) + \beta_3(Treatment_{ij} \times Race_{ij}) + \sum \phi X_{ij} + \varepsilon_i + \mu_j. \quad (1)$$

In the models, we include students' intervention status, $Treatment_{ij}$, race/ethnicity, $Race_{ij}$, an intervention-by-race/ethnicity interaction, ($Treatment_{ij} * Race_{ij}$), and a vector of covariates, X_{ij} , including the number of suspensions received in the sixth grade, gender, free or reduced-price lunch eligibility

(as an indicator of family poverty level), special education status, English learner status, and cohort fixed effects.³

Results

Self-Affirmation Effects on Suspensions: Main Effects and Effects by Race

The first model of Table 1 is a main effects model in which the dependent variable is a count of the number of suspensions a student received in seventh and eighth grades. This model suggests that there is no statistically discernable main effect of intervention on the number of suspensions for the average student in the sample ($\beta = -0.06$, $SE = 0.05$, $z = -1.18$, $p = .238$, $d = -0.04$). The second model in Table 1 shows results from the model described in Equation (1), in which we identify the effects of the intervention by race and ethnicity. Consistent with prior literature, there are statistically significant racial differences in the number of postintervention suspensions that Black students in the control group received versus control group students who are White ($\beta = 0.84$, $SE = 0.12$, $z = 7.25$, $p < .001$), all else equal. The treatment coefficient suggests that there is no intervention effect for the typical White student—the reference category ($\beta = 0.01$, $SE = 0.07$, $z = 0.12$, $p = .906$). Treatment-by-race coefficients suggest similar null effects for Asian students ($\beta = 0.16$, $SE = 0.17$, $z = 0.92$, $p = .358$) and Latinx students ($\beta = 0.11$, $SE = 0.15$, $z = 0.75$, $p = .456$) as compared with White students. There are statistically significant negative effects of the intervention on the suspension count for Black students ($\beta = -0.56$, $SE = 0.14$, $z = -4.00$, $p < .001$) compared with their White counterparts, after accounting for covariates. Postestimation linear combination tests confirm that the control-treatment group difference among Black students is statistically significant ($\beta = -0.55$, $SE = 0.12$, $z = -4.60$, $p < .001$, $d = -0.28$).

Figure 1 more clearly shows how the intervention reduces racial suspension gaps. The gap reduction occurs because self-affirmation results in a one half of a suspension reduction for typical Black students but has no effect for typical White students. As a consequence, the intervention eliminates two thirds of the Black-White suspension gap. Unadjusted means and standard deviations by treatment condition and subgroups of interest are shown in Supplemental Table S2 (in the online version of the journal).

These results in Table 1 are robust to several alternate model specifications. First, the treatment effect among Black students is consistent within each cohort when analyzing the two cohorts separately (see Supplemental Table S4 in the online version of the journal). Second, the treatment effect among Black students is consistent when examining seventh- and eighth-grade suspensions separately (see Supplemental Table S5 in the online version of the journal). Third, the treatment effects among Black students are consistent when using binary suspension outcome indicators, suggesting the intervention reduces the probability of having any suspension among

Table 1
Mixed-Effects Regression Estimates of Self-Affirmation Treatment Impacts on Suspensions

Independent Variable	(1)	(2)	(3)	(4)
Treatment group				
Race/ethnicity (reference = White)				
Asian	-0.06 (0.05)	0.01 (0.07)	-0.07 (0.06)	-0.02 (0.07)
Black	-0.13 (0.10)	-0.21* (0.13)	-0.13 (0.10)	-0.08 (0.13)
Latinx	0.54*** (0.09)	0.84*** (0.12)	0.54*** (0.09)	0.52*** (0.13)
Any prior suspensions	0.08 (0.10)	0.03 (0.12)	0.08 (0.10)	0.15 (0.13)
Treatment × Asian	1.95*** (0.09)	1.94*** (0.09)	1.92*** (0.12)	1.52*** (0.23)
Treatment × Black		0.16 (0.17)		-0.00 (0.17)
Treatment × Latinx		-0.56*** (0.14)		-0.26 (0.16)
Treatment × Prior suspensions		0.11 (0.15)	0.04 (0.16)	0.00 (0.15)
Asian × Prior suspension				0.60* (0.31)
Black × Prior suspension				-1.47*** (0.45)
Latinx × Prior suspension				1.23*** (0.29)
Treatment × Asian × Prior suspension				-0.51 (0.33)
Treatment × Black × Prior suspension				2.37*** (0.78)
Treatment × Latinx × Prior suspension				-1.33*** (0.40)
Constant	0.14* (0.07)	0.12 (0.08)	0.16** (0.07)	0.28 (0.49)
				0.15** (0.08)

Note. From the combined two-cohort intent-to-treat (ITT) samples, we retain students from the self-affirmation intervention for which all variables are present between seventh and eighth grades; $N = 2,149$. Estimates are derived using a mixed-effects model with students nested in 11 schools. The dependent variable is the number of suspensions received postintervention in seventh and eighth grades. Models include controls for gender, free or reduced-price lunch participation, limited English proficiency, special education status, and cohort fixed effects. Standard errors in parentheses. $*p < .10$. $**p < .05$. $***p < .01$.

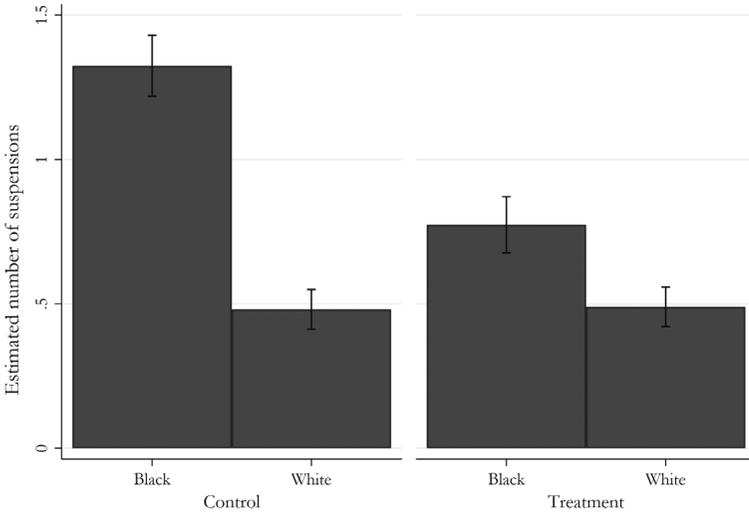


Figure 1. Intervention reduced the average Black-White suspension gap by 67%.

Note. Figures represent the marginal means for Black and White students derived from Model 2 in Table 1. The three standard deviation cutoff version of the suspension variable was used to address outliers. The error bars represent \pm one standard error around the means.

Black students by 6 percentage points and reduces the probability of having two or more suspensions among the same group by 10 percentage points (see Supplemental Table S6 in the online version of the journal). Finally, the statistically significant results among Black participants hold when using a zero-inflated negative binomial model to account for both the integer outcome and the prevalence of students in the sample with no suspensions (see Supplemental Table S7 in the online version of the journal).

Sensitivity Checks

Testing Heterogeneous Effects by Preintervention Suspensions

Another remaining question is, does the intervention more generally help all students with prior exposure to exclusionary discipline who may be at particular risk to be labeled as troublemakers, independent of their race or ethnicity? Or, is the intervention uniquely effective for Black students who must confront both stereotypes concerning their group and a prior history of disciplinary incidents? To investigate whether self-affirmation had greater benefits for students most at risk to be labeled as troublemakers, independent of their racial/ethnic background, we next test whether the intervention effect differs for students who are and are not suspended in middle school prior to

intervention (i.e., sixth grade). To examine this question, we estimate the following model:

$$Y_{ij} = \alpha_{ij} + \beta_1(\textit{Treatment}_{ij}) + \beta_2(\textit{Treatment}_{ij} \times \textit{Pre - suspension}_{ij}) + \sum \varphi X_{ij} + \varepsilon_i + \mu_j. \quad (2)$$

We then test whether a three-way interaction between treatment condition, race/ethnicity, and prior suspensions influences the efficacy of the treatment, by adding interaction terms to Equation (2):

$$Y_{ij} = \alpha_{ij} \dots + \beta_3(\textit{Treatment}_{ij} \times \textit{Race}_{ij}) + \beta_4(\textit{Race}_{ij} \times \textit{Pre - suspension}_{ij}) + \beta_5(\textit{Treatment}_{ij} \times \textit{Race}_{ij} \times \textit{Pre - suspension}_{ij}) \quad (3)$$

Finally, we use postestimation linear combination tests to determine both within-race and within-suspension history control-treatment differences.

The third model in Table 1 shows that there is no statistically significant difference in treatment effectiveness among students with no prior suspension history ($\beta = -0.07$, $SE = 0.06$, $z = -1.19$, $p = .235$, $d = -0.05$) and no statistically significant difference in coefficients between treatment group students with and without a prior suspension history ($\beta = 0.04$, $SE = 0.16$, $z = 0.23$, $p = .815$). A linear combination test shows no simple treatment effect among students with prior suspensions ($\beta = -0.03$, $SE = 0.15$, $z = -0.20$, $p = .84$, $d = -0.04$), suggesting that the intervention does not have stronger general impacts for students with prior suspensions from school.

The fourth model in Table 1 shows results from tests of heterogeneous treatment effects by race and prior suspension history. First, we urge caution in the interpretations of treatment effects among Asian, Latinx, and White students with prior suspensions due to low cell counts. For example, in Cohort 1, there are only four Asian students with prior suspensions and in Cohort 2 there are 10 Asian students with prior suspensions. One of the Asian treatment group students with prior suspensions has a large number of suspensions compared with one or zero suspensions for all other control and treatment group Asian students with prior suspensions. Similarly, there are 22 Latinx students with prior suspensions in Cohort 1 and 26 Latinx students in Cohort 2 with prior suspensions. There are 29 White students with prior suspensions in Cohort 1 and 35 White students with prior suspensions in Cohort 2. These small numbers of Asian, Latinx, and White students likely make these estimates unreliable.

With that in mind, we find no treatment effects among White students without prior suspensions ($\beta = -0.02$, $SE = 0.07$, $z = -0.34$, $p = .737$, $d = -0.03$) and the coefficients for the difference between White students with and without prior suspensions did not differ at the conventional $p < .05$ alpha level ($\beta = 0.60$, $SE = 0.31$, $z = 1.96$, $p = .050$). The treatment effect difference between White and Latinx students with prior suspensions is not statistically significant ($\beta = 0.28$, $SE = 0.49$, $z = 0.58$, $p = .562$) but the difference in

coefficients between White and Asian students with prior suspensions is statistically significant ($\beta = 2.37$, $SE = 0.78$, $z = 3.02$, $p = .002$). Postestimation linear combination tests suggest no statistically significant treatment effect for White students with prior suspensions at the conventional $p < .05$ alpha level ($\beta = 0.58$, $SE = 0.30$, $z = -1.94$, $p = .053$, $d = 0.21$) and statistically significant treatment effects among students with prior suspensions who are Latinx ($\beta = 0.86$, $SE = 0.35$, $z = 2.45$, $p = .014$, $d = 0.31$) and Asian ($\beta = 2.95$, $SE = 0.70$, $z = 4.19$, $p < .001$, $d = 2.21$). Treatment effects are not substantively or statistically significant for students with no prior suspensions who are Latinx ($\beta = -0.03$, $SE = 0.14$, $z = -0.17$, $p = .867$, $d = -0.03$) or Asian ($\beta = -0.03$, $SE = 0.14$, $z = -0.17$, $p = .867$, $d = -0.02$). These tests suggest that Asian and Latinx treatment students with prior suspensions receive more suspensions than their counterparts in the control group. Supplemental Table S3 (in the online version of the journal) shows that many of these subsample treatment effects differ between cohorts and again, we urge caution when interpreting these results due to very small subsample sizes.⁴

Subsample sizes are larger among Black students with prior suspensions in Cohort 1 ($n = 55$) and Cohort 2 ($n = 90$) than other racial or ethnic groups and there is acceptable treatment-control balance on prior suspension status among Black students ($\chi^2 = 0.58$, $p = .45$). Black students in the treatment group on average experience greater reductions in postintervention suspensions relative to White treatment group peers with prior suspensions (as indicated by the “Treat \times Black \times Prior suspension” interaction term; $\beta = -1.33$, $SE = 0.40$, $z = -3.35$, $p = .001$). A postestimation linear combination test indicates that the treatment effect is statistically significant for the group of Black students with prior suspensions ($\beta = -1.01$, $SE = 0.20$, $z = -5.01$, $p < .001$, $d = -0.39$) but is not statistically significant at the conventional $p < .05$ level among Black students without prior suspensions ($\beta = -0.28$, $SE = 0.15$, $z = -1.94$, $p = .052$, $d = -0.25$). The three quarters of a suspension difference between Black treatment group students with and without prior suspensions is statistically significant ($\beta = -0.72$, $SE = 0.25$, $z = -2.91$, $p = .004$).

Figure 2 more clearly captures the change in suspensions among Black students with and without prior suspensions who participate in the intervention. While the treatment has a statistically nonsignificant one quarter of a suspension reduction among Black students without prior suspensions, average treatment group Black students with a prior history of suspensions realize a one suspension reduction as a result of treatment compared with their peers in the control group.

Main and Heterogeneous Treatment Effects on Office Disciplinary Referrals

One critique of testing the effects of the intervention on suspension outcomes is that the decision to suspend is removed temporally from the disciplinary incident and is ultimately determined by someone other than the student and teacher who were originally involved in the incident. Other

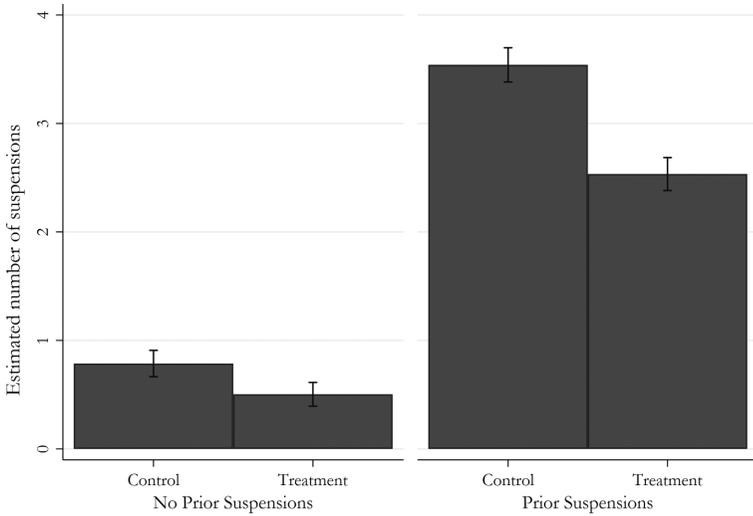


Figure 2. Average cumulative suspensions for Black students by intervention status and whether suspended in the year prior.

Note. The figures represent the marginal means derived from all Black students included in Model 4 of Table 1. The three standard deviation cutoff version of the suspension variable was used to address outliers. The error bars represent \pm one standard error around the means.

disciplinary actions, such as being sent out of class or being sent to the office, though less severe than suspension, are temporally and spatially closer to the infraction in question, and more clearly reflect the outcome of the teacher-student interaction. As a check for the sensitivity of the intervention on a broader record of student disciplinary incidents, we replace our suspension outcome with a count of the more commonly occurring ODR. ODRs are typically reported by a classroom teacher and involve sending a student to the office in response to a perceived rule infraction. These referrals most often do not lead to a suspension from school and can range from violent offenses or theft to simply having an argument with teachers or defying their directives. Many suspensions would not occur without a student first being sent to the office. Examining ODRs allows us to assess whether reductions in suspensions for Black recipients of self-affirmation may in part be the result of fewer trips to the office. Because ODRs are more common and widespread than suspensions, this analysis also helps us better describe the effects of the intervention on a wider range of students who contend with a “troublemaker” label.

Table 2 presents models estimating impacts of the intervention on a count of the number of ODRs a student receives in seventh and eighth grades. The

model specifications are otherwise identical to those in the previous models with suspensions as an outcome. Model 1 shows a statistically significant main treatment effect on the number of ODRs a student receives, suggesting that the treatment reduces the average students' ODR count by two thirds of an incident ($\beta = -0.67$, $SE = 0.24$, $z = -2.76$, $p = .006$, $d = -0.07$). However, Model 2 suggests that virtually all the treatment main effect can be explained by the impacts for Black treatment group students, who experience about three fewer ODRs in seventh and eighth grades due to the intervention compared with their White treatment group counterparts ($\beta = -3.15$, $SE = 0.64$, $z = -4.90$, $p < .001$). Postestimation linear combination tests show a statistically significant treatment effect for Black students of one-and-two-thirds fewer ODRs compared with their Black control group counterparts ($\beta = -1.66$, $SE = 0.86$, $z = -1.99$, $p = .046$, $d = -0.35$). These impacts among Black students represent a 66% reduction in the Black-White gap in office disciplinary referrals. The results are consistent when examining seventh- and eighth-grade ODRs separately (see Supplemental Table S5 in the online version of the journal) and when using zero-inflated negative binomial models (see Supplemental Table S7 in the online version of the journal).

When considering each student's history of ODRs in sixth grade, Model 3 suggests that there is no statistically significant average treatment effect among students with no prior ODRs ($\beta = -0.26$, $SE = 0.27$, $z = -0.92$, $p = .358$, $d = -0.04$) but a statistically significant difference between students with and without prior ODRs ($\beta = -1.67$, $SE = 0.56$, $z = -2.98$, $p = .003$). A linear combination test suggests a statistically significant average reduction of about two ODRs due to treatment among students with prior ODRs ($\beta = -1.93$, $SE = 0.49$, $z = -3.96$, $p < .001$, $d = -0.20$).

Model 4 shows that there is no statistically significant treatment effect among White students without prior ODRs ($\beta = -0.11$, $SE = 0.35$, $z = -0.32$, $p = .752$, $d = -0.05$). The treatment coefficients are not statistically different from those of White treatment group students with prior ODRs ($\beta = 0.16$, $SE = 0.95$, $z = 0.17$, $p = .865$). Relative to White treatment group students with ODRs, treatment interaction coefficients do not differ at the conventional $p < .05$ level for those with prior ODRs who are Asian ($\beta = 1.15$, $SE = 2.40$, $z = 0.48$, $p = .630$), Black ($\beta = -2.53$, $SE = 1.45$, $z = -1.74$, $p = .081$), or Latinx ($\beta = -1.75$, $SE = 1.57$, $z = -1.11$, $p = .266$). Postestimation linear combination tests suggest that there is a statistically significant four-ODR reduction due to treatment among Black students with prior ODRs ($\beta = -4.03$, $SE = 0.71$, $z = -5.66$, $p < .001$, $d = -0.37$) as well as a smaller but still statistically significant effect among Black students without prior ODRs ($\beta = -1.66$, $SE = 0.84$, $z = -1.99$, $p = .046$, $d = -0.33$). The difference between Black treatment group students with and without prior ODRs is statistically significant ($\beta = -2.37$, $SE = 1.10$, $z = -2.16$, $p = .031$). There are no statistically significant effects among those with prior ODRs who are Latinx ($\beta = -1.26$, $SE = 1.04$, $z = -1.21$, $p = .227$, $d = -0.05$) or Asian ($\beta = 1.42$, $SE = 2.07$, $z = 0.69$, $p = .492$,

Table 2
Mixed-Effects Regression Estimates of Self-Affirmation Treatment Impacts on Office Disciplinary Referrals (ODRs)

Independent Variable	(1)	(2)	(3)	(4)
Treatment group				
Race/ethnicity (reference = White)				
Asian	-0.67*** (0.24)	-0.11 (0.33)	-0.26 (0.28)	-0.11 (0.35)
Black	-0.50 (0.44)	-0.75 (0.60)	-0.52 (0.44)	-0.20 (0.62)
Latinx	3.03*** (0.41)	4.77*** (0.54)	3.02*** (0.41)	1.57** (0.72)
Any prior ODRs	0.26 (0.46)	0.31 (0.56)	0.25 (0.46)	-0.20 (0.61)
Treatment × Asian	7.09*** (0.32)	7.06*** (0.32)	7.96*** (0.43)	5.46*** (0.70)
Treatment × Black	0.45 (0.79)	0.45 (0.79)		0.22 (0.82)
Treatment × Latinx	-3.15*** (0.64)	-3.15*** (0.64)		-1.55* (0.91)
Treatment × Prior ODRs	-0.07 (0.68)	-0.07 (0.68)		0.44 (0.78)
Asian × Prior ODRs			-1.67*** (0.56)	0.16 (0.95)
Black × Prior ODRs				-4.84*** (1.66)
Latinx × Prior ODRs				6.37*** (1.07)
Treatment × Asian × Prior ODRs				2.15* (1.11)
Treatment × Black × Prior ODRs				1.15 (2.40)
Treatment × Latinx × Prior ODRs				-2.53* (1.45)
Constant	0.48 (0.39)	0.21 (0.40)	0.29 (0.39)	0.46 (0.40)

Note. From the combined two-cohort intent-to-treat (ITT) samples, we retain students from the self-affirmation intervention for which all variables are present between seventh and eighth grades; $N = 2,149$. Estimates are derived using a mixed-effects model with students nested in 11 schools. The dependent variable is the number of ODRs received postintervention in seventh and eighth grades. Models include controls for gender, free or reduced-price lunch participation, limited English proficiency, special education status, and cohort fixed effects. Standard errors in parentheses. * $p < .10$. ** $p < .05$. *** $p < .01$.

$d = 0.19$) or among those with no prior ODRs who are Latinx ($\beta = 0.33$, $SE = 0.70$, $z = 0.47$, $p = .637$, $d = 0.23$) or Asian ($\beta = 0.11$, $SE = 0.75$, $z = 0.14$, $p = .887$, $d = 0.16$). Unadjusted means and standard deviations by treatment condition are shown in Supplemental Table S2 (in the online version of the journal).

Variation in Student Discipline Explained by Postintervention GPA

Though prior evidence of impact on GPA reveals some inconsistencies across student cohorts within the current district context (Hanselman et al., 2017), the overall literature shows statistically significant and substantively meaningful achievement impacts of self-affirmation on students who are negatively stereotyped within the academic domain (Wu et al., 2021). Therefore, it could be that the intervention's impacts on suspensions and disciplinary referrals are explained by improvements in Black students' academic performances. Indeed, teachers may begin to see students as more able and worthy of attention and mentoring, thus amplifying the effects of the intervention via teacher expectancy effects (Purdie-Vaughns et al., 2009). Such changes in teachers' perspectives and expectancy effects, along with students' accumulating academic successes, may be mechanisms that explain decreases in students' suspensions and office disciplinary referrals. In this final sensitivity analysis, we assess the extent to which the students' GPA outcomes statistically explain these disciplinary outcomes.

To test this possibility, we compare suspension and ODR models from Tables 1 and 2 with those in which we also include postintervention GPA. This provides an assessment of the degree to which GPA mediates the effects of the intervention on disciplinary outcomes. In Supplemental Table S8 (in the online version of the journal), we first show that there is no statistically significant effect of the intervention on academic achievement for Black students across the two intervention cohorts ($\beta = 0.04$, $SE = 0.05$, $z = 0.65$, $p = .517$, $d < 0.01$), suggesting that GPA is unlikely to mediate the effects of the intervention on disciplinary outcomes. When postintervention GPA is included as a predictor of suspensions, it reduces the size of the intervention-by-Black student interaction coefficient by roughly 9%, and the interaction effect remains statistically significant. Much of this shift in the treatment-by-Black interaction coefficient comes from statistically nonsignificant effects among White students and not effects among Black students. Linear combination tests show that postintervention GPA reduces the size of the treatment effect on suspensions among Black students by about 4% (i.e., $0.527/0.551$). Using the same analytical approach for the ODR outcome, the results suggest that including postintervention GPA in the model reduces the magnitude of the treatment effect on ODRs among Black students by about 4%. This suggests that the self-affirmation intervention's effects on Black students' suspensions and ODRs are largely independent of effects on academic achievement.

Discussion

Suspension from school is strongly associated with academic motivation, achievement, dropout, criminal activity, and incarceration. The relationships between suspensions and these adverse outcomes are especially concerning for Black students, who are disciplined at a disproportionate rate relative to all other student groups, in this sample and nationally. Our study of the effects of self-affirmation on middle school student disciplinary involvement highlights three important findings. First, as hypothesized, the self-affirmation intervention we study is most beneficial for Black students, with no evidence of effects on other racial/ethnic groups. A typical Black student in the control group of our sample is suspended once per year, while a similar Black student assigned to self-affirmation writing exercises is suspended half as frequently. Second, we find that these treatment effects are more pronounced among Black students with prior suspensions, who experience a full one-suspension reduction due to treatment. Third, we find that Black treatment group students also experience fewer ODRs. While Black students with no prior ODRs see a one-and-two-thirds reduction in ODRs in middle school after the intervention, Black students with a prior history of ODRs have four fewer ODRs in seventh and eighth grades than their control group counterparts.

We conclude that the intervention cuts the middle school Black-White gap in suspensions by 67% across the school district, net of controls. In practical terms, these estimates suggest that a cohort of 150 Black middle-school students will receive approximately 82 fewer suspensions over the seventh and eighth grades. Supplemental results examining treatment effects on ODRs bolster these findings, suggesting a 66% reduction in the Black-White gap in ODRs, or 249 fewer ODRs among a cohort of 150 Black middle-school students, all else equal. Such reductions promise meaningful improvements for overall school and district climate, improved teaching and learning opportunities within classrooms, and more positive student-teacher relationships.

Toward an Improved Understanding of the Black-White Suspension Gap and Self-Affirmation's Role in Ameliorating It

There are a variety of reasons for disciplinary involvement in schools that should be understood by researchers and interventionists before considering potential solutions. According to Gregory et al. (2010):

Social class, immigrant status, racial and ethnic identity, neighborhood and familial diversity, and educator training and perspectives may all affect student behavior, teacher responses, or their interaction. Clearly, conducting research that could truly sort out the numerous and interacting sources of variance contributing to disciplinary disproportionality is challenging. Subtle and implicit processes related to racial bias, negative expectations, or stereotypes are not easily detected outside of controlled laboratory conditions, and it is not

a simple matter to observe the complex and interactive social processes that can contribute to an escalating sequence of actions and reactions during actual discipline encounters. (p. 64)

Of all the reasons the authors note, we believe that those involving “subtle and implicit processes” that produce racial biases are the most relevant teacher-specific pathways to disproportionality in disciplinary involvement. These complex processes of prejudice and implicit bias are among the most widely studied in the field of psychology (Cameron & Turner, 2010). However, prior work has also suggested that implicit bias and its real-world behavioral manifestations have been both challenging to understand (Gregory et al., 2010) and difficult to prevent (Cameron & Turner, 2010; Forscher et al., 2019).

Beyond biases held by teachers, we believe that the “complex and interactive social processes” involving students and teachers referenced by Gregory et al. (2010) provides a compelling explanation of how disproportionate disciplinary outcomes persist in schools. This “vicious cycle” causes educators to label Black students as “troublemakers” and to disproportionately and unfairly punish them (Okonofua, Walton, & Eberhardt, 2016). In response to this unjust behavior, Black students’ trust in their teachers and other school authorities wanes and, rather than being deterred by the sanctions, are instead motivated to exhibit increasing defiance.

Though the subtle and implicit biases exhibited by teachers and the vicious cycle of degrading teacher-student relationships contribute much to Black-White suspension gaps, our intervention focuses on students. We posit that Black students are psychologically threatened by the troublemaker identity imposed on them by others and experience internalized racial oppression as a result. These psychological mechanisms threaten students’ sense of moral and adaptive adequacy (Cohen & Sherman, 2014; Steele, 1988). In turn, these psychological responses can negatively affect Black students’ sense of well-being, their behaviors, and their relationships with teachers. The psychological and social harm of being labeled as a troublemaker based on stereotypes, while also having one’s group disproportionately disciplined, is a context in which most would react apprehensively, defensively, and with little trust (Steele, 1988). In the context of our study, we speculate that the double jeopardy of these stereotypes and prior experiences of being disciplined uniquely conspire to develop troublemaker identity threats among the Black students in our two samples through internalized racial oppression.

Our study and its findings situate self-affirmation as a student-focused strategy for reducing disciplinary incidents by helping Black students focus on positive aspects of their identities that can help them contend with a troublemaker label imposed by others. There are several mechanisms explaining the intervention’s effectiveness in improving Black students’ identities. The intervention expands Black students’ self-concepts by prompting them to

focus on valued aspects of their identities beyond the stereotyped domain (Critcher & Dunning, 2015), which in turn reduces attention to and saliency of the troublemaker identity. With less attention diverted to the harm imposed by the troublemaker identity, Black students' sense of school belonging increases, which tempers feelings of marginalization (Shnabel et al., 2013). Reduced saliency of the troublemaker identity and tempered feelings of marginalization, in turn, reduce students' defensiveness and mistrust (Schmeichel & Vols, 2009). Though we do not actively observe or measure these psychological mechanisms in this large field trial, we surmise that some or all of them explain the observed impacts on Black students' suspension and ODR outcomes.

As theorized, threats to identity appear to be muted by the self-affirmation intervention. Our results are consistent with prior research that suggests that the strategy we employ counteracts identity threats and promotes a more resilient and positive concept of the self as good, competent, and efficacious (Steele, 1988). The intervention does not erase the bias and discrimination students encounter in society, nor does it necessarily remove the environmental threats associated with this labeling. Instead, as Schmeichel and Vols (2009) articulate, "self-affirmation acts as a powerful salve for negative feedback and other threats to the self, such that self-affirmed individuals forego defensive, self-protective responses to threat in favor of more open and evenhanded responses" (p. 778). Such reflection can reduce defensiveness and mistrust (Sherman & Cohen, 2006), providing an avenue through which the student-teacher interaction can be improved (Okonofua, Walton, & Eberhardt, 2016). By reducing defensiveness and mistrust and promoting a more broadly construed and positive sense of self, the affirmations can help Black students facing troublemaker identity threats—or yet another disciplinary encounter—navigate the teacher-student interaction more productively and, potentially, avoid a trip to the office or temporary removal from school.

Limitations and Future Directions

Self-affirmation theory suggests that the exercises positively impact Black students' disciplinary outcomes through a number of potential mechanisms, including their psychological responses to identity threats, the nature of their interactions and relationships with their teachers and other school authority figures, and, ultimately, their behaviors. Nevertheless, as Gregory et al. (2010) suggest, the measurement of these complex social and psychological phenomena, and their potential interactions, is difficult if not impossible in a large field-based trial such as this study. The "vicious cycle" theory posits that many Black students, by virtue of being both stereotyped as guilty and receiving a disproportionate number of suspensions, are labeled as troublemakers, lose trust in their teachers, and grow defensive (Okonofua, Walton, & Eberhardt, 2016). Our findings suggest that changes induced by self-

affirmation help Black students more readily navigate the threats and injustices that they encounter in school and less often fall victim to the vicious cycle of mistrust, thus substantially reducing suspensions by way of decreases in the total number of office disciplinary referrals they receive.

To more clearly investigate these hypotheses, future studies of self-affirmation and discipline should collect data on student-teacher relationships, teachers' views of students, and students' views of themselves to determine the specific social and psychological constructs that are most important for preventing the development and persistence of troublemaker labels and identities. Self-affirmed Black students are provided resources to help combat stereotyping and navigate the vicious cycle through a variety of likely mechanisms—from improved teacher-student social interactions, to heightened psychological resources to combat internalized racial oppression. Efforts to uncover more evidence concerning these mediating factors are needed.

Because education is a highly applied field and because these results have implications for further scale-up, some practical, policy-based considerations deserve attention. First, our disciplinary outcomes suggest that only Black students benefit from the self-affirmation intervention. We observe no benefits for other historically underserved racial and ethnic student subgroups such as Latinx students. Second, Black students who have a history of prior disciplinary infractions particularly benefit, suggesting that when the psychological toll of the troublemaker identity threat is especially salient—as indicated by both the broader societal stereotypes against Black students and an individual history as a Black student having been previously punished—the self-affirmation is particularly effective.

Third, though we find relatively consistent impacts across the 11 schools within this context, prior research on self-affirmation's achievement impacts have revealed school-to-school differences. This is of particular importance for schools in which students of color represent the clear minority and in which the achievement gaps are particularly large, which are schools where there have been the largest academic impacts of self-affirmation interventions (Borman et al., 2018; Hanselman et al., 2014). Further research is needed to determine whether implementations of self-affirmation in other district or school contexts reveal additional replicated outcomes for disciplinary impacts.

Finally, though the intervention can be relatively easily administered with minimal investment of additional time and resources, there are some subtle aspects regarding implementation that are important. One consideration is that the timing of the intervention appears crucial. The intervention should be delivered as early in the school year as possible (Cook et al., 2012)—as was the case in our study—and before students experience adversity and corresponding negative outcomes (Cohen & Sherman, 2014; Critcher et al., 2010). A second consideration is the way the intervention is framed can affect how students perceive and respond to the writing activities. For example, when the exercises are introduced as beneficial, or “good for you,” it can

suggest that the activity is intended to be therapeutic and to “cure” something that is “wrong” with the participants, which can undermine the effectiveness of affirmations (Sherman et al., 2009). Finally, the exercises seem most beneficial when introduced as normal classroom activities delivered by teachers who are seeking to know more about their students’ most important values (Cohen et al., 2006; Cohen & Sherman, 2014; Silverman et al., 2013).

These issues of how, for whom, in which school contexts, and under what conditions of implementation deserve continued exploration, especially regarding the disciplinary outcomes studied here. We certainly do not believe that the pervasive biases and stereotypes in schools should be ignored because self-affirmation can help Black students effectively navigate them. Quite the contrary; we believe that simultaneous efforts to affect the mindsets of teachers and inoculate students against racial oppression may provide the strongest solutions to closing the Black-White suspension gap at scale. Evidence from a wide range of disciplines has established that such a union of proposed approaches, which simultaneously target beliefs and behaviors of both teachers and students, produces the largest and most lasting effects on mitigating racial inequities, even more so than each of its parts (Okonofua, Walton, & Eberhart, 2016).

Self-affirmation is not a panacea, and it is not likely to solve all disciplinary inequities in U.S. schools. However, the opportunity costs for using self-affirmation interventions, as opposed to engaging in a similar alternative intervention or creative writing activity, are negligible. Because students already engage in numerous writing activities over the course of the school year in their classes, teachers could integrate (with the help of researchers) the 45 to 60 minutes of self-affirmation exercises during the school year in ways that are unlikely to noticeably interfere with typical class routines. Racially inequitable exclusionary discipline practices interfere with Black students’ opportunities to learn, and portend a host of negative life outcomes, including dropout, unemployment, and incarceration. Given these significant and life-changing associations between suspensions and school, career, and life outcomes, this intervention has the potential for great benefit, at almost no risk. Its implementation may reduce the number of suspensions a student receives; if not, the only cost would be 45 to 60 minutes of creative writing practice.

ORCID iDs

Geoffrey D. Borman  <https://orcid.org/0000-0002-7039-8208>

Jaymes Pyne  <https://orcid.org/0000-0003-2570-5384>

Christopher S. Rozek  <https://orcid.org/0000-0003-4389-8538>

Supplemental Material

Supplemental material for this article is available online.

Notes

This research was supported by the Institute of Education Sciences, United States Department of Education (R305A110136 and R305B150003) and the Spencer Foundation (201500044). The opinions expressed are those of the authors and do not represent views of the funders. We thank the teachers, administrators, and students of the Madison Metropolitan School District (Wisconsin) for their support of and cooperation with this project.

¹Throughout this article, we conform to the American Psychological Association (APA) guidelines for writing about racial and ethnic identity (<https://apastyle.apa.org/style-grammar-guidelines/bias-free-language/racial-ethnic-minorities>). The APA guidelines recommend parallel designations and capitalization of “Black” and “White” when describing racial groups by color.

²These findings may have arisen because the sampled schools were predominantly Black institutions. In highly segregated racial contexts such as these, racial minority students tend to experience relatively low levels of stereotype threat due to the simple lack of other distinct racial/ethnic groups to which their behaviors and outcomes can be compared and stereotyped. Borman et al. (2018) and Hanselman et al. (2014) found self-affirmation’s effects on achievement are largest in so-called “high-threat” schools, defined as those schools in which a marginalized racial group is a minority of the student body and relatively underperforming. Because these school contextual cues have been theorized to amplify stereotype threat and, thus, increase the salience and corresponding strength of the intervention, the Black students in Dee’s (2015) sample may have realized few benefits at all for this reason alone.

³In preliminary analyses using a cross-level treatment-by-school interaction, we found that less than 1% of the variance in treatment effect was between schools.

⁴For example, we again note that one Asian student assigned to the treatment group had a large number of suspensions in seventh and eighth grades. When this outlier case was removed, the large, anomalous result among Asian students with prior suspensions was greatly reduced and no longer statistically significant.

References

- Belmi, P., Barragan, R. C., Neale, M. A., & Cohen, G. L. (2015). Threats to social identity can trigger social deviance. *Personality and Social Psychology Bulletin*, *41*(4), 467–484. <https://doi.org/10.1177/0146167215569493>
- 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>
- Borman, G. D., Choi, Y., & Hall, G. J. (2021). The academic impacts of a brief middle-school self-affirmation intervention help propel African American and Latino students through high school. *Journal of Educational Psychology*, *113*(3), 605–620. <https://doi.org/10.1037/edu0000570>
- 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*, 1773–1784. <https://doi.org/10.1177/0956797618784016>
- Bushaw, W. J., & Lopez, S. J. (2010). A time for change: The 42nd annual Phi Delta Kappa/Gallup poll of the public’s attitudes toward the public schools. *Phi Delta Kappan*, *92*(1), 8–26. <https://doi.org/10.1177/003172171009200103>
- Cameron, L., & Turner, R. N. (2010). The application of diversity-based interventions to policy and practice. In R. J. Crisp (Ed.), *Social issues and interventions*. *The*

- psychology of social and cultural diversity* (pp. 322–351). Wiley Blackwell. <https://doi.org/10.1002/9781444325447.ch14>
- Cohen, G. L., & Garcia, J. (2014). Educational theory, practice, and policy and the wisdom of social psychology. *Policy Insights from the Behavioral and Brain Sciences*, 1(1), 13–20. <https://doi.org/10.1177/2372732214551559>
- 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., 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, 333–371. <https://doi.org/10.1146/annurev-psych-010213-115137>
- 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), 479–496. <https://doi.org/10.1037/a0026312>
- Critcher, C. R., & Dunning, D. (2015). Self-affirmations provide a broader perspective on self-threat. *Personality and Social Psychology Bulletin*, 41(1), 3–18. <https://doi.org/10.1177/0146167214554956>
- Critcher, C. R., Dunning, D., & Armor, D. A. (2010). When self-affirmations reduce defensiveness: Timing is key. *Personality and Social Psychology Bulletin*, 36(7), 947–959. <https://doi.org/10.1177/0146167210369557>
- David, E. J. R., Schroeder, T. M., & Fernandez, J. (2019). Internalized racism: A systematic review of the psychological literature on racism's most insidious consequence. *Journal of Social Issues*, 75(4), 1057–1086. <https://doi.org/10.1111/josi.12350>
- 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>
- Dover, T. L., Major, B., & Kaiser, C. R. (2016). Members of high-status groups are threatened by pro-diversity organizational messages. *Journal of Experimental Social Psychology*, 62, 58–67. <https://doi.org/10.1016/j.jesp.2015.10.006>
- Duguid, M. M., & Thomas-Hunt, M. C. (2015). Condoning stereotyping? How awareness of stereotyping prevalence impacts expression of stereotypes. *Journal of Applied Psychology*, 100(2), 343–359. <https://doi.org/10.1037/a0037908>
- Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.
- Easterbrook, M. J., & Hadden, I. R. (2021). Tackling educational inequalities with social psychology: Identities, contexts, and interventions. *Social Issues and Policy Review*, 15(1), 180–236. <https://doi.org/10.1111/sipr.12070>
- Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., & Booth, E. A. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement*. Council of State Governments Justice Center.
- Fisher, C. B., Wallace, S. A., & Fenton, R. E. (2000). Discrimination distress during adolescence. *Journal of Youth and Adolescence*, 29(6), 679–695. <https://doi.org/10.1023/A:1026455906512>
- Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology*, 117(3), 522–559. <https://doi.org/10.1037/pspa0000160>

- Gabbidon, S. L., Greene, H. T., & Young, V. D. (2002). *African American classics in criminology and criminal justice*. Sage.
- Gibbs, J. P. (1975). *Crime, punishment, and deterrence*. Elsevier.
- Goff, P. A., Jackson, M. C., Di Leone, B. A. L., Culotta, C. M., & DiTomasso, N. A. (2014). The essence of innocence: Consequences of dehumanizing Black children. *Journal of Personality and Social Psychology, 106*(4), 526–545. <https://doi.org/10.1037/a0035663>
- Gopalan, M., & Nelson, A. A. (2019). Understanding the racial discipline gap in schools. *AERA Open, 5*(2). <https://doi.org/10.1177/2332858419844613>
- Goyer, J. P., Cohen, G. L., Cook, J. E., Master, A., Apfel, N., Lee, W., Henderson, A. G., Reeves, S., 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>
- Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher, 39*(1), 59–68. <https://doi.org/10.3102/0013189X09357621>
- Gregory, A., & Weinstein, R. S. (2008). The discipline gap and African Americans: Defiance or cooperation in the high school classroom. *Journal of School Psychology, 46*(4), 455–475. <https://doi.org/10.1016/j.jsp.2007.09.001>
- 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., Grigg, J., Pyne, 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>
- Jordan, W. J., Lara, J., & McPartland, J. M. (1996). Exploring the causes of early dropout among race-ethnic and gender groups. *Youth & Society, 28*(1), 62–94. <https://doi.org/10.1177/0044118X96028001003>
- Kupchik, A. (2010). *Homeroom security: School discipline in an age of fear*. NYU Press. <https://doi.org/10.18574/nyu/9780814748206.001.0001>
- Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of antiprejudice messages: How motivational interventions can reduce (but also increase) prejudice. *Psychological Science, 22*(12), 1472–1477. <https://doi.org/10.1177/0956797611427918>
- Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology, 74*(6), 1421–1436. <https://doi.org/10.1037/0022-3514.74.6.1421>
- Logel, C., & Cohen, G. L. (2012). The role of the self in physical health: Testing the effect of a values-affirmation intervention on weight loss. *Psychological Science, 23*(1), 53–55. <https://doi.org/10.1177/0956797611421936>
- Losen, D. (2011). *Discipline policies, successful schools, and racial justice*. National Education Policy Center.
- Mickelson, R. A. (1990). The attitude-achievement paradox among Black adolescents. *Sociology of Education, 63*(1), 44–61. <https://doi.org/10.2307/2112896>

- Mizel, M. L., Miles, J. N., Pedersen, E. R., Tucker, J. S., Ewing, B. A., & D'Amico, E. J. (2016). To educate or to incarcerate: Factors in disproportionality in school discipline. *Children and Youth Services Review, 70*, 102–111. <https://doi.org/10.1016/j.childyouth.2016.09.009>
- Morris, E. W., & Perry, B. L. (2016). The punishment gap: School suspension and racial disparities in achievement. *Social Problems, 63*(1), 68–86. <https://doi.org/10.1093/socpro/spv026>
- Noltemeyer, A., Ward, R. M., & McLoughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review, 44*(2), 224–240. <https://doi.org/10.17105/spr-14-0008.1>
- Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science, 26*(5), 617–624. <https://doi.org/10.1177/0956797615570365>
- Okonofua, J. A., Paunesku, D., & Walton, G. M. (2016). Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proceedings of the National Academy of Sciences, 113*(19), 5221–5226. <https://doi.org/10.1073/pnas.1523698113>
- Okonofua, J. A., Walton, G. M., & Eberhardt, J. L. (2016). A vicious cycle: A social-psychological account of extreme racial disparities in school discipline. *Perspectives on Psychological Science, 11*(3), 381–398. <https://doi.org/10.1177/1745691616635592>
- Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science, 26*(6), 784–793. <https://doi.org/10.1177/0956797615571017>
- Pearman, F. A., Curran, F. C., Fisher, B., & Gardella, J. (2019). Are achievement gaps related to discipline gaps? Evidence from national data. *AERA Open, 5*(4). <https://doi.org/10.1177/2332858419875440>
- Purdie-Vaughns, V., Cohen, G., Garcia, J., Sumner, R., Cook, J., & Apfel, N. (2009). Improving minority academic performance: How a values-affirmation intervention works. *Teachers College Record*. <https://www.tcrecord.org/content.asp?contentid=15774>
- Pyne, J. (2019). Suspended attitudes: Exclusion and emotional disengagement from school. *Sociology of Education, 92*(1), 59–82. <https://doi.org/10.1177/0038040718816684>
- Robbins, C. G. (2008). *Expelling hope: The assault on youth and the militarization of schooling*. State University of New York Press.
- Schmeichel, B. J., & Vohs, K. (2009). Self-affirmation and self-control: Affirming core values counteracts ego depletion. *Journal of Personality and Social Psychology, 96*(4), 770–782. <https://doi.org/10.1037/a0014635>
- Sherman, D. K. (2013). Self-affirmation: Understanding the effects. *Social and Personality Psychology Compass, 7*(11), 834–845. <https://doi.org/10.1111/spc3.12072>
- Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology, 38*, 183–242. [https://doi.org/10.1016/S0065-2601\(06\)38004-5](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), 745–764. <https://doi.org/10.1037/a0015451>
- 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>
- Shollenberger, T. L. (2014). Racial disparities in school suspension and subsequent outcomes. In D. J. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 31–43). Teachers College Press.
- Silverman, A., Logel, C., & Cohen, G. L. (2013). Self-affirmation as a deliberate coping strategy: The moderating role of choice. *Journal of Experimental Social Psychology*, 49(1), 93–98. <https://doi.org/10.1016/j.jesp.2012.08.005>
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality. *Urban Review*, 34(4), 317–342. <https://doi.org/10.1023/A:1021320817372>
- Sparks, S. D. (2020, November 20). Training bias out of teachers: Research shows little promise so far. *Education Week*. <https://www.edweek.org/leadership/training-bias-out-of-teachers-research-shows-little-promise-so-far/2020/11>
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology*, 21, 261–302. [https://doi.org/10.1016/S0065-2601\(08\)60229-4](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), 613–629. <https://doi.org/10.1037/0003-066X.52.6.613>
- U.S. Department of Education Office for Civil Rights. (2016). *2013-2014 civil rights data collection: A first look*. U.S. Government Printing Office.
- 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>
- Way, S. M. (2011). School discipline and disruptive classroom behavior: The moderating effects of student perceptions. *Sociological Quarterly*, 52(3), 346–375. <https://doi.org/10.1111/j.1533-8525.2011.01210.x>
- Wentzel, K. R. (2002). Are effective teachers like good parents? Teaching styles and student adjustment in early adolescence. *Child Development*, 73(1), 287–301. <https://doi.org/10.1111/1467-8624.00406>
- Wolf, K. C., & Kupchik, A. (2017). School suspensions and adverse experiences in adulthood. *Justice Quarterly*, 34(3), 407–430. <https://doi.org/10.1080/07418825.2016.1168475>
- Wu, Z., Spreckelsen, T. F., & Cohen, G. L. (2021). A meta-analysis of the effect of values affirmation on academic achievement. *Journal of Social Issues*. Advance online publication. <https://doi.org/10.1111/josi.12415>
- Yeager, D. S., Purdie-Vaughns, V., Hooper, S. Y., & Cohen, G. L. (2017). Loss of institutional trust among racial and ethnic minority adolescents: A consequence of procedural injustice and a cause of life-span outcomes. *Child Development*, 88(2), 658–676. <https://doi.org/10.1111/cdev.12697>
- Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, 81(2), 267–301. <https://doi.org/10.3102/0034654311405999>

Manuscript received April 28, 2020

Final revision received April 7, 2021

Accepted July 24, 2021