Integrating mindfulness and connection practices into preservice teacher education results in durable automatic race bias reductions

Matthew J. Hirshberg [a,b,*], Lisa Flook [a], Evan E. Moss [c], Robert D. Enright [b], Richard J. Davidson [a]

[a] *Center for Healthy Minds, University of Wisconsin–Madison, USA*
[b] *Department of Educational Psychology, University of Wisconsin-Madison, USA*
[c] *Department of Curriculum and Instruction, University of Wisconsin-Madison, USA*

A B S T R A C T

Automatic race bias, which is the tendency to associate positive attributes more quickly with White as compared to Black faces, reflects enculturation processes linked to inequitable teaching behaviors. In sample of undergraduate preservice teachers ($N = 88$), we examined whether a novel mindfulness and connection practice intervention without anti-bias content incorporated into undergraduate teacher education would result in reduced automatic race bias favoring White faces. Random assignment to the intervention predicted significantly reduced race preference for White child faces immediately after the intervention. These significant reductions persisted at the 6-month follow-up, which are the most durable reductions in automatic race bias reported to date in adults. Data from semi-structured interviews indicated that the intervention enhanced self-awareness and self-regulation while reducing automatic responding among preservice teachers. These qualities are instrumental to adaptive teaching and putative mechanisms for reducing automatic race bias. The potential value of integrating mindfulness and connection practices into undergraduate preservice teacher education is discussed.

1. Introduction

Since the Coleman report's (Coleman et al., 1966) findings on K–12 racial educational achievement disparities, decades of research and policy have attempted to identify and rectify the contributing factors. Yet more than 50 years later, many educational achievement disparities remain (Snyder et al., 2016; Snyder & Dillow, 2012). Structural inequities within (e.g., educational financing) and outside (e.g., criminal justice statutes) of educational systems contribute to these persistent disparities, in part by producing disparate opportunities for educational attainment (Carter & Welner, 2013). Addressing them through policy reform is necessary to disrupt the conditions that perpetuate inequitable educational opportunities that result in racial achievement disparities. At the same time, policies are enacted by individuals. The proximal school context for student development is the classroom where individual biases may be wittingly or unwittingly played out.

Observers have noted that even policies designed to remediate inequities can ultimately further disparate outcomes (Holbein &

Ladd, 2017; Ladson-Billings, 2003; Markowitz, 2018). For example, the No Child Left Behind law was intended to reduce the racial achievement gap but instead led to diminished curricula that emphasized lower-order skills and in some cases reduced access to quality educational settings for the most under-resourced students with the greatest learning needs (Darling-Hammond, 2007). Although historical examples of the intentional undermining of policies intended to promote racial equality are plentiful (e.g., voter ID laws; Grimmer et al., 2018), at the level of the individual, inequitable behaviors need not be intentional. In the scientific literature, the distinction between witting and unwitting (or less witting) displays of inequitable racial attitudes and behaviors can be characterized by the measurement approach.

Explicit assessments of bias directly ask individuals to report on their attitudes about various races. Conversely, *indirect* or *implicit* assessments of bias typically utilize behavioral indices of automaticity that are under less conscious control. Implicit or what we will call automatic race bias is defined as the difference in the strength of positive versus negative associations between two races, often White and Black. Unlike explicit bias, these value associations are thought to occur automatically, or at least partially beneath conscious awareness, and are subject to limited self-regulation. In the most widely used measure of automatic race bias, the Implicit Association Test (IAT; Greenwald et al., 2003), the difference in the speed with which an individual associates positive and negative adjectives with Black and White faces provides the behavioral (i.e., automatic) index of race bias. When negative adjectives are more rapidly associated with Black than White faces and positive adjectives are more rapidly associated with White than Black faces, more automatic race bias is present. Because the difference in these associations is typically on the order of tens of milliseconds, the phenomenon has been described as automatic race bias. Although debate remains about the extent of these measures' automaticity and their accessibility to conscious awareness, it is clear that automatic forms of race bias are less easily regulated and vary in the strength of their relationship to explicit expressions of race bias (Charlesworth & Banaji, 2019; Nosek et al., 2007; Starck et al., 2020). Those who explicitly value multiculturalism and egalitarianism, which many teachers report that they do, are more motivated to control prejudicial views (Hachfeld et al., 2011). Explicit appreciation of multiculturalism does not necessarily translate to lower levels of automatic race bias, however. Any understanding of the impact of race bias on racial disparities in educational outcomes must be conditioned on both the level of analysis (i.e., systemic or individual) and type of measurement (i.e., explicit or automatic).

2. Prevalence of race bias in teachers

Until recently, there had been a notable paucity of research investigating the prevalence and impact of race bias in teaching. Recent research has demonstrated that both types of race bias (i.e., explicit and automatic) are expressed at similar levels in teachers as they are in the general population (Kumar et al., 2015; Starck et al., 2020). Starck et al. (2020) conducted the most comprehensive examination of bias in teachers to date. Using data collected online by Project Implicit (Nosek et al., 2007) on over 1.6 million American adults, the authors compared rates of explicit and automatic race bias between teachers (preK–12) and non-teachers. Among the 68,930 teachers in the sample, 30.3% reported explicit bias favoring White faces that was statistically different from zero, and 77% demonstrated automatic bias favoring White faces of a magnitude that was statistically different from zero. There was no substantive difference between teacher and non-teacher levels of explicit or automatic race bias.

Starck et al. (2020) findings are consistent with prior large-scale investigations of the prevalence of race bias among adults in America and other Western nations (Charlesworth & Banaji, 2019; Nosek et al., 2007) and with a smaller body of research on bias in teachers. For example, Kumar et al. (2015) studied White middle school teachers' ($N = 241$) explicit and automatic race bias toward Arab-Americans in a Midwestern American region with a large Arab-American population. Teacher levels of automatic race bias favoring adolescent White faces on both the Arab/White and Black/White IATs were comparable to general population estimates (e.g., Nosek et al., 2007; Starck et al., 2020). Automatic race bias favoring the White cultural majority over minoritized races/ethnicities (e.g., Turkish, Moroccan) has also been observed internationally in German preservice teachers (PST; Glock et al., 2013; Glock & Karbach, 2015) and in Dutch teachers (Van den Bergh et al., 2010). It is important to note that both in the general population and in teacher samples, automatic race bias favoring Whites is, on average, significantly lower among participants that identify as Black, Latino, or of other race/ethnicities (Starck et al., 2020). This is not surprising when considering that automatic forms of race bias reflect dominant cultural norms and structural inequities (Vuletich & Payne, 2019). We would expect that White teachers, in particular, reflect the evaluations of the dominant culture in which they live.

3. What effect does automatic race bias have on teaching behaviors?

The observation that teachers demonstrate equivalent levels of race bias to the general population is a cause for concern, but alone does not establish that the presence of bias is associated with inequitable teaching practices or student outcomes. There are a number of potential pathways whereby automatic race bias could affect teacher practice or student achievement. For example, teachers' expectations for student achievement are based on prior observations of students and other observable inputs (e.g., linguistic style, race, gender) that may trigger biases and stereotypes (Friedrich et al., 2015; Reyna, 2000, 2008). Teacher expectations in turn have been shown to affect student achievement in what is known as the Pygmalion effect (Rosenthal, 2010). For example, in a study of 73 5th grade teachers and 1289 of their students, Friedrich et al. (2015) found that teacher expectancy was associated with end of year math grades and math standardized assessment scores. Based on expectations, teachers may interact differently with students depending on student characteristics such as race or ethnicity. In four meta-analyses, Tenenbaum and Ruck (2007) found that teachers had higher expectations for Asian and White students than for Black and Latino students and engaged in higher levels of positive speech (e.g., encouragement) and less frequent negative speech (e.g., punitive discipline) with White than Black or Latino students. Such manifest forms of biased teaching behaviors have the potential to deleteriously impact student perceptions of themselves as learners (e.g.,

Friedrich et al., 2015), their teacher, and of the classroom/educational system as a safe space to learn.

When considering the dearth of research on teacher bias in general, it is unsurprising that research on the impacts of automatic race bias on teaching behaviors and student outcomes is limited. As more studies over the last decade or so have been conducted, evidence is mounting that automatic race bias impacts both teacher behaviors and teacher perceptions of students. Halberstadt et al. (2018) studied anger bias (i.e., a form of bias in which an individual wrongly interprets facial expressions as angry) in sample of 40 American pre-service teachers (PSTs; 80% White). PSTs were significantly more likely to interpret Black faces as angry and were most likely to evaluate Black male faces as angry. The inability to accurately detect the emotions of another person has important implications for how teachers perceive students and how students in turn perceive their teachers and experience their education (McKown, 2013). Similarly, in Halberstadt et al. (2020), PSTs (70% White) who were low on automatic race bias as assessed by the child Black/White race IAT (i.e., faces of children in place of adults) showed no difference in anger attributions between Black and White child faces. However, PSTs high in automatic race bias were significantly more likely to interpret Black child faces as angry.

Associations between teacher levels of automatic race/ethnicity bias on the IAT and student achievement have also been documented. For example, students learned the most in mathematics class when their teacher's automatic ethnic bias favored their ethnic group (Peterson et al., 2016). In a sample of 31 Dutch elementary school teachers and 434 of their students, Van den Bergh et al. (2010) reported that teacher automatic bias but not explicit ethnic bias (i.e., White versus Moroccan) predicted the degree of the ethnic minority achievement gap between classrooms. Taken together, these results may help explain the positive effect that same-race teacher match has on educational outcomes for students of color (e.g., Gershenson et al., 2018).

Experimental research provides additional evidence that higher teacher automatic race bias deleteriously impacts students of color by biasing teacher behaviors and expectations. Okonofua and Eberhardt (2015) recruited teachers to read records of ostensible student infractions. The experimental manipulation involved assigning the infractions to stereotypically Black and White male names. Teachers responded to stereotypically Black names with higher levels of feeling troubled and endorsing harsher disciplinary action for the exact same behavior. Similarly, preschool teachers who watched videos of four children interacting (one Black boy and girl and one White boy and girl) so that researchers could ostensibly learn about how teachers detect challenging behaviors looked at Black children significantly longer than White children, and Black boys longest of all. The videos were carefully curated to contain no instances of challenging behavior (Gilliam et al., 2016). These results are significant when considering that Black students are far more likely than White students to receive exclusionary discipline (i.e., discipline that involves removal from instructional time) for similar infractions (Skiba et al., 2014). Exclusionary discipline predicts poorer academic achievement, higher rates of dropping out of high school, and adult criminal convictions (Gregory et al., 2010; Skiba et al., 2014).

Today, more than 51% of American students are students of color but about 80% of teachers are White (U.S. Department of Education, 2018a). Most teachers enter the profession with the prosocial motivation to support student growth and development (Wolman, 2010). Yet, as Starck et al. (2020) suggested, it is possible "that well-intentioned teachers may be subject to biases they are less conscious of, limiting their capacity to facilitate racial equity despite their best intentions" (p. 273). This possibility highlights the need for continued research on the relationship between teacher biases, behaviors, and effective strategies to reduce automatic race bias.

4. Is reducing automatic race bias possible?

Traditional college or university-based teacher education programs (TEPs) continue to be the primary avenue for educating future teachers (U.S. Department of Education, 2018b). Many TEPs recognize the importance of educating future teachers about historical inequities based on race, gender, immigration status, and other characteristics. Social justice education and instruction in inclusive classroom practices and multicultural education are often mandatory (Ladson-Billings, 2003; Marx, 2006; Philip et al., 2019; Shah & Coles, 2020). Anti-bias psychoeducation may be effective in diminishing some forms of bias, but any effects on automatic bias are fleeting (Forscher et al., 2019; Lai et al., 2016). In fact, the longest reported adult reductions in IAT race bias following any type of intervention is 8 weeks (Devine et al., 2012). There currently is little evidence that durable reductions in automatic race bias are achievable, at least in adults.

We contend that durable reductions in automatic race bias are possible, but to instantiate them, interventions must impact the cognitive and affective mechanisms that underlie automaticity. The brief interventions favored in the bias reduction literature are unlikely to produce such changes. Conversely, interventions aimed at systematically developing attention, awareness, and prosocial dispositions are theorized to operate on the very cognitive and affective processes that underlie automatic bias (Dahl et al., 2015; Nosek et al., 2011; Shapiro et al., 2006; Vago & Silbersweig, 2012). They therefore offer the possibility of producing durable bias reductions by instantiating long-term change in these processes.

Mindfulness and connection practice interventions may be well suited to producing durable changes in race bias. Mindfulness is commonly described as paying attention, on purpose, to present moment experience with an attitude of nonjudgment or acceptance (Kabat-Zinn, 2013). Mindfulness is an umbrella term for a variety of techniques geared toward enhancing awareness of cognitive, affective, and sensory experience, strengthening regulatory processing (e.g., focused attention, cognitive flexibility), and cultivating the ability to remain unperturbed by thoughts, sensations, and emotions (i.e., equanimity). Through enhanced awareness and regulatory capacity, theorists contend that mindfulness practice can disrupt automatic patterns of thinking, feeling, and behaving (Dahl et al., 2015; Shapiro et al., 2006). Automatic race bias is a form of automatic behavior in which a stimulus (e.g., a Black or White face) elicits an association or a behavior (De Houwer, 2019) that is difficult to regulate and may be unconscious. Mindfulness practice may generate greater awareness of automatic responding which, when coupled with the increased self-regulatory skill thought to develop through mindfulness practice, provides greater volition over stimulus-dependent responding (Lindsay & Creswell, 2017).

Connection practices are a style of practice intended to develop prosocial tendencies (Hirshberg et al., 2018; Hofmann et al., 2011). For example, loving-kindness practice is intended to strengthen feelings of warmth, goodwill and connection to those with whom we do and don't identify, even including individuals we struggle with. Over time, loving-kindness practice is intended to construct a superordinate category consisting of all human beings (even all living beings) with whom we feel a basic kinship (Kabat-Zinn & Salzberg, 2004).

Whereas mindfulness practices may develop competencies that disrupt patterns of habit and reactivity, leading to greater equanimity (Kang et al., 2013), connection practices may systematically strengthen prosocial qualities *and* extend the boundaries of the in-group. As the in-group expands, so too might the preferential attitudes that attend to notions of in-group (Sherif, 1958). Both represent potential mechanisms by which automatic biases might be reduced. Because these practices are understood to produce benefits by strengthening competencies over time, effects may be more durable than reductions achieved by other means (Hofmann et al., 2011). In the current study, we examined the impact of a novel mindfulness and connection training intervention designed to be integrated into a preservice TEP. We hypothesized that learning mindfulness and connection practice at the same time as professional knowledge and skills might synergistically benefit learning to teach and processes such as automatic race bias that get in the way of enacting effective teaching practices.

Even brief mindfulness training has significantly reduced various forms of automatic bias (Hopthrow et al., 2017; Kiken & Shook, 2011; Lueke & Gibson, 2015, 2016). Furthermore, reduced automaticity/increased self-regulation and increased affiliation with out-group members have been identified as mechanisms of change. Lueke and Gibson (2015) found that reduced automaticity was a causal mechanism linking 10 min of mindfulness practice to automatic race bias reductions. Similarly, automatic race bias reduction following a connection induction were mediated by decreased automaticity as well as increased other-oriented concern (Stell & Farsides, 2016). A longer duration connection intervention (i.e., 6 weeks) resulted in similar automatic bias reductions (i.e., race and homelessness bias; Kang et al., 2014). None of these studies included follow-up assessments, however. It remains unknown whether such interventions produce durable change.

5. Mindfulness and connection training with teachers

Jennings and Greenberg (2009) proposed the Prosocial Classroom Model, a socio-ecological systems model of schools, to illustrate the potential benefits on school and classroom learning environments, and teacher well-being and practice, of developing social-emotional competencies such as mindfulness and empathy in teachers. They argue that teacher social-emotional competency levels are an important mediator of effective teaching. Equitable teaching practices are effective teaching practices (Ladson-Billings, 1995). In the school context, bias that manifests in inequitable behaviors represents ineffective teaching practice. We suggest that in addition to the potential benefits of mindfulness and connection practice training on teachers' own well-being, through bias reduction, the potential exists for these trainings to promote more equitable and therefore more effective teaching practices (e.g., discipline, student expectations).

We are not aware of a study examining whether mindfulness and connection practice training reduces automatic race bias in teachers or PSTs. However, results from mindfulness research with teachers are consistent with Jennings and Greenberg's (2009) theorizing. For example, in a randomized controlled trial (RCT) of in-service teachers, Roeser et al. (2013) reported significant improvements in mindfulness, self-compassion, stress, anxiety, depression, and occupational burnout that persisted at the 3-month follow-up. In another RCT of in-service teachers, Jennings et al. (2017) reported significant intervention group improvements in mindfulness, emotion regulation, psychological distress, and time urgency with some evidence for increases in emotionally supportive classroom practices assessed through a validated classroom observation protocol. Klingbeil and Renshaw's (2018) meta-analysis and Emerson et al.'s (2017) systematic review of mindfulness-based intervention studies with in-service teachers found moderate magnitude treatment effects on multiple teacher outcomes (e.g., mindfulness, psychological distress, emotion regulation). However, the authors of both papers noted several limitations in the published literature, including a reliance on self-report measures, rare reporting of treatment fidelity, and the potential for publication bias.

6. Present study

The present data come from a mixed-method investigation of the impact of integrating mindfulness and connection practices into undergraduate PST education. This study focuses on the secondary outcome of change in automatic race bias (https://clinicaltrials.gov/ct2/show/NCT02544412). We hypothesized that a novel mindfulness and connection intervention would result in automatic race bias reductions that persist over the 6-month follow-up in a sample of primarily White elementary education PSTs. Automatic race bias was assessed with both the child and adult versions of the Black/White IAT (Greenwald et al., 2003). The adult and child versions of the task were administered because motivation to change is an important ingredient in reducing bias (Devine et al., 2012). We reasoned that motivations toward adult and child faces might differ in a sample of PSTs training to teach preK–8th graders, possibly manifesting in greater bias reductions toward child faces. Quantitative analyses focus on group by time interactions in automatic race bias. The semi-structured interview protocol inquired into perceptions of mindfulness practices in PST education and did not ask about bias. However, we sought to contextualize IAT effects by examining follow-up interviews for statements reflecting our proposed mechanisms of change, including (a) increased awareness of thoughts, sensations and emotions; (b) reduced automaticity/increased self-regulation; and (c) extension of the in-group.

7. Method

7.1. Participants

Participants were recruited from an undergraduate elementary education program at a large Midwestern University. Nearly 60% of eligible PSTs enrolled and participated (98/176, 55.7%). One control cohort ($n = 10$) was removed from analyses as at the same time that they were serving as controls for the present study, they were recruited into a yoga-based intervention intended to increase well-being and reduce stress that included mindfulness practices. The research agreement with the TEP prohibited cohorts involved in this study from being recruited into mindfulness or similar intervention studies. In addition, the study protocol requested that participants refrain from engaging in similar interventions during the study period (however, participants already practicing mindfulness were not asked to stop). Participants in this control cohort took part in this alternative training, compromising this cohort's ability to serve as a control. In light of these violations to our study protocol and our inability to determine who from this cohort participated in the other intervention, we removed this control cohort from analyses. Sensitivity analyses including this cohort are reported in Supplemental Materials. The final sample analyzed was $N = 88$ ($n = 31$ Control; $n = 57$ Intervention).

The sample was representative of the demography of the teacher education program and more broadly of America's teaching corps (U.S. Department of Education, 2018a). Participants were 95.6% female ($n = 85$) and predominately White (86.36%, $n = 76$). About 6% of participants self-reported as Hispanic ($n = 5$), 3.41% as Asian ($n = 3$), another 3.41% as African American ($n = 3$), and one participant self-reported as mixed or other race ($n = 1$). The average age was 21.99 years old ($SD = 0.68$). Participants had substantial prior experience working with children in some capacity ($M = 58.57$ months, $SD = 49.06$), were happy with their choice to enter into teaching ($M = 8.81$, $SD = 1.14$, 10 = Very happy), and felt that teaching was a good career fit ($M = 4.44$, $SD = 0.62$, out of 5).

7.2. Power analysis

A priori power analysis conducted with G*Power 3.1 software (Faul et al., 2009) indicated that for planned one-tailed post-test minus pre-test and follow-up test minus pre-test difference score models (i.e., independent group *t*-tests of GROUP × time interaction effects from mixed Analysis of Variance [ANOVA]), with α set to 0.05, power (1- β error probability) = 0.80, $N = 102$ would be required to detect moderate ($d = 0.50$) magnitude effects. Based on effect sizes from related research on automatic race bias (e.g., Kang et al., 2014; $d = -0.60$) and pragmatic considerations (e.g., for logistical reasons involving eight cohorts over 2 years was the maximum possible duration for recruitment), the study was designed for an anticipated sample of $N = 100$, sufficiently powered to detect moderate magnitude and larger effects. Our recruited sample of $N = 98$ effectively achieved the planned sample from this power analysis. However, the unexpected need to remove one control cohort, reducing our analyzable sample to $N = 88$, means that the lower boundary of detectable effects based on this power analysis (i.e., $d = 0.50$) may go undetected.

7.3. Materials

Testing occurred on Dell Latitude E5550 15″ monitor laptop computers. EPrime psychological testing software version 2.1.0 (Psychological Software Tools Inc., 2012) was used to conduct IATs. Pre-, post-, and follow-up data from all study measures not reported here are presented separately (Hirshberg et al., 2020). Materials used in this study, other than the intervention, are widely available.

7.4. Race implicit association tests

Two standard Black/White race IATs were implemented following the procedures outlined in Lane et al. (2007) with stimuli procured from Project Implicit® (Nosek et al., 2007). In the adult version, participants are shown images of adult faces in the center of the screen (counterbalanced by race [Black/White] and gender [female/male]). The child version follows the same structure but replaces adult faces with child faces. In critical trials, positive (e.g., hope) and negative (e.g., evil) adjectives are presented on the upper right or left of the screen. In some blocks participants are asked to associate as quickly as possible positive attributes with Black faces and negative attributes with White faces. In other critical blocks, the opposite pairing is made.

The primary outcome score is the *D*-score. The *D*-score is computed by taking the difference in response time latencies for critical blocks divided by the pooled standard deviation of all correct trial latencies (Greenwald et al., 2003). A *D*-score of 0 reflects complete neutrality between race associations. A more negative score reflects greater bias toward White faces/preference for Black faces. A more positive score reflects greater bias toward Black faces/preference for White faces. Although no consensus standard exists, a score greater than 0.65 is often described as *strong*, scores between 0.65 and 0.35 described as *moderate*, and scores less than 0.35 as *slight* (Greenwald et al., 2003).

Prior research has shown that pairing positive White/negative Black first results in greater bias toward Black faces/preference for White faces (i.e., more positive D-scores) than the reverse (Nosek et al., 2005). We therefore used two orders for the adult and child IATs, corresponding to the initial pairing of race with good/bad adjectives, and counterbalanced the task order between the intervention and control groups such that the same proportion in each group received each order (i.e., approximately 50%). Participants received the same order at each timepoint. Experimenter error at follow-up resulted in six participants (five intervention, one control) completing the wrong task order, introducing within-subject measurement error and between group imbalance (at follow-up 62% of the intervention group as compared to 54% of controls first paired White/good and Black/bad). We therefore controlled for follow-up

task version in analyses.

Task validity has been established in a number of samples and in meta-analyses correlating task performance to relevant explicit forms of bias and to bias in real-world behaviors (Greenwald et al., 2015; Kurdi et al., 2018; McConnell & Leibold, 2001; Nosek, 2007). In this sample, the reliability of both the child and adult IATs was high (split-half reliability of both $r = 0.92$).

7.5. Qualitative interviews

The qualitative component of the study was designed by two experts in qualitative methods, one a qualitative researcher with over 20 years of experience in teacher education and the other a PhD student focusing on qualitative research methods. Qualitative methods were designed around the assumption that if aligned with specific instructional practices (i.e., integrated into teacher education), mindfulness and connection practices have the capacity to engage PSTs in relationship skills and instructional noticing. Given the lack of empirical evidence related to similar types of training with PSTs, this work was intended to be highly descriptive (see Supplemental Materials for interview protocol). Interview questions were developed to answer the overarching research question: How do actors experience the infusion of mindfulness practices into a PST education program?

Interviews were conducted at three timepoints throughout the study by three PhD level graduate students focusing on qualitative research methods, one of whom was the protocol designer. Interviews were 1 h and semi-structured, which allowed for a natural conversation to unfold. The first interview asked PSTs about how the mindfulness training was going, what tools they had developed to be more responsive to students, and about their own stress and emotion. The second interview centered around PSTs reflection of the mindfulness training, including eliciting connections being made between mindfulness constructs and teaching practice. The last interview, reported on here, explored PSTs evolving mindfulness practice, the implementation of mindfulness within teacher education, and final thoughts about connections between mindfulness and teaching.

In contrast to the quantitative methods, the qualitative component was not constructed around a priori hypotheses but instead was intended to identify how PSTs viewed the integration of mindfulness and connection practices into their learning. For example, although extant data suggest that these practices benefit in-service teachers, the extent to which PSTs would find these practices relevant or meaningful to apply in their training was not known.

Twenty-five intervention participants (96% Female, 88% White) were selected for interviews based on scoring in roughly the highest and lowest quartiles of stress and mindfulness at baseline. This stratification strategy was used because prior evidence of stress reduction from these practices suggests that sampling the least and most stressed participants might reveal interesting differences in perceptions of the intervention. Alternatively, consistent participant perceptions of the intervention would suggest perceived utility (if perceptions are positive) or lack of utility (if perceptions are negative) regardless of PST stress and mindfulness levels.

Within the qualitative or interpretive tradition, we chose to follow a case study approach, which guided us in all phases of our research, from the questions we asked to the way we interpreted data. A case study methodology helped us to study a complex phenomenon (i.e., how PSTs engaged in mindfulness) in actual, real-world contexts (Yin, 1994). The qualitative analysis reported here was a hybrid between phenomenological and inductive methods (Groenewald, 2004; Thomas, 2006). That is, we were interested in whether participant experiences reflected any of our three proposed mechanisms of bias reduction: (a) increased awareness of
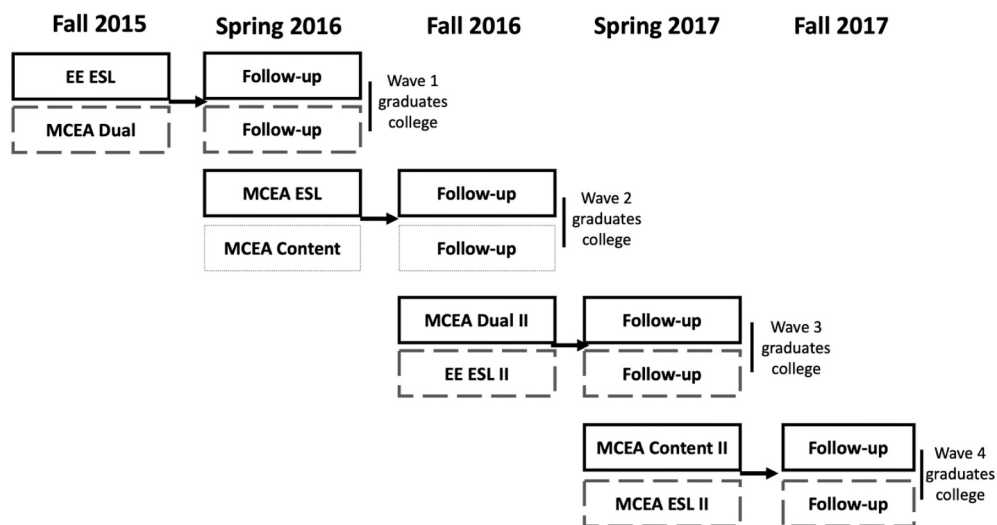

Fig. 1. Study design.
Note. Black outline = intervention cohort. Dashed outline = control cohort. Gray outline = control cohort removed for receiving yoga/mindfulness training. EE ESL = Early education English as Second Language certification cohort; MCEA = Middle Childhood / Early Adolescent; Dual = dual certification cohort; ESL = English as Second Language certification; Content = Content area certification. Reprinted with permission from Hirshberg et al. (2020).

thoughts, sensations, and emotions; (b) reduced automaticity/increased self-regulation; and (c) extension of the in-group, and we sought to summarize participant experiences to establish links to both the quantitative data and putative mechanisms of change.

The qualitative data analysis process began with audio-recording and then transcribing interviews. The protocol designers next used multiple coding methods to analyze the data, including inductive methods (i.e., descriptive coding using an open coding technique; Saldaña, 2021) and deductive coding using key concepts from research design (Creswell, 2013). Preliminary coding (i.e., classification) and interpretive memos written throughout this process showed emerging connected relationships that were identified as categories. An iterative process of additional coding passes, re-classifying categories, and identifying connected constellations of
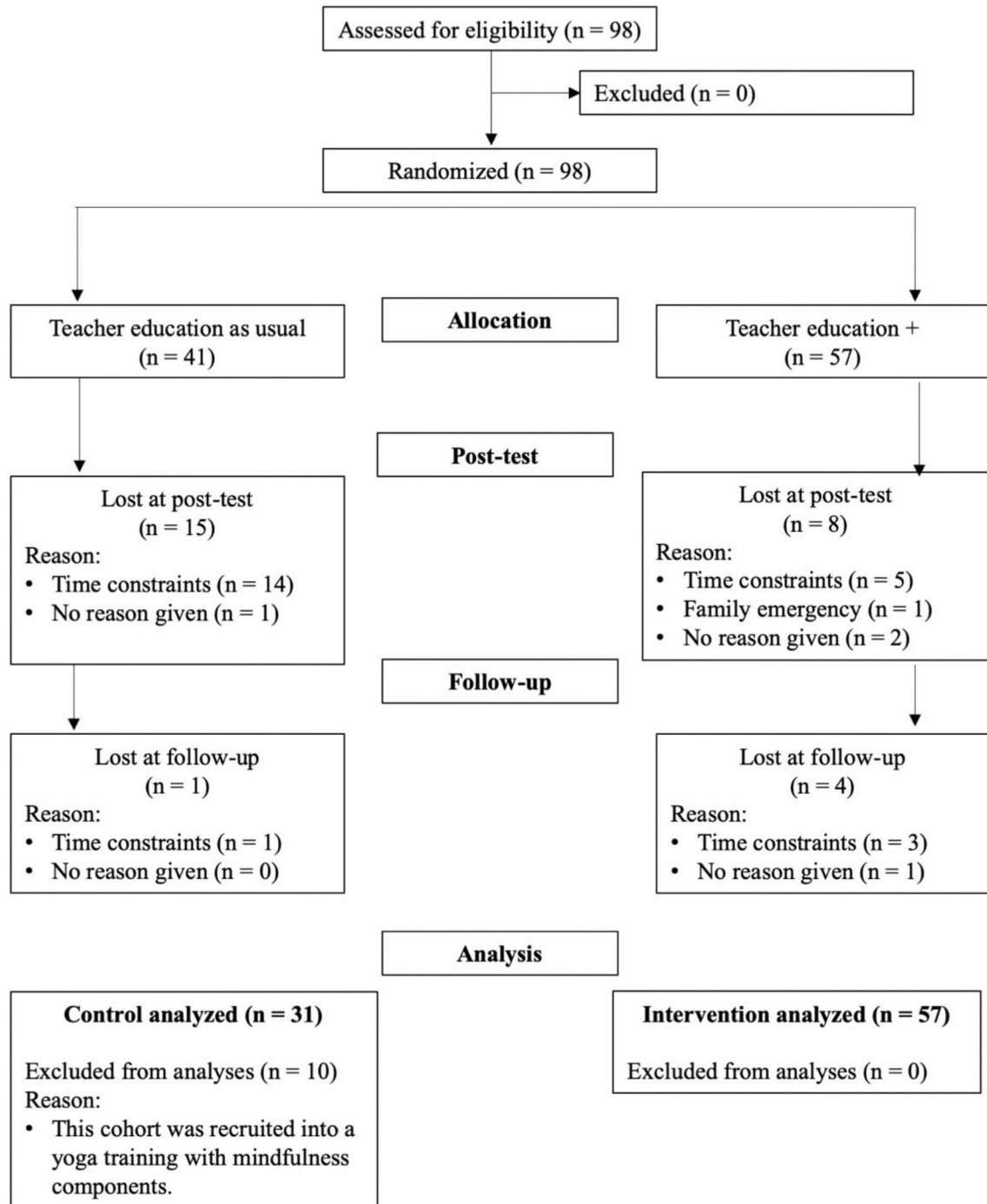

Fig. 2. Consolidated standards of reporting trials (CONSORT) study flow diagram.
Note. Reprinted with permission from Hirshberg et al. (2020).

categories helped the protocol designers chunk categories into themes.

Credibility checks were conducted with participants in real-time during interviews, with the interviewer often asking follow-up and clarification questions in subsequent interviews. In addition, the protocol designers engaged in inter-coder reliability checks through attention to intercoder agreement at all stages of analysis, from the initial coding pass through the development of larger themes. The process unfolded at regularly scheduled meetings focused on reaching consistency and agreement on the interpretation of excerpts, codes attached to those excerpts, as well as the organization of code families into categories and later, themes. Between meetings, shifting thinking was documented through on-going analytic and interpretive memo exchanges. In addition, the protocol designers were continuously reflecting and documenting through analytical memos how their positionality manifested in the way they interpreted, coded, and categorized data into themes. Researchers were intentional about looking for contrasting examples and searching for multiple meanings within the data.

7.6. Procedures

Participants from the competitive admission elementary education program (EEP) were recruited by a class visit and by email when entering their first semester of their senior year (i.e., third of four semesters in the program; see Fig. 1). The EEP follows a cohort model. Random assignment of individuals was not possible due to the integration of the intervention into weekly cohort seminar time. Thus, we block randomized by cohort type to ensure that at study end four cohorts were assigned to intervention and four to control (i.e., eight total), and of the four licensure tracks in the EEP, one of each was assigned to intervention and one to control (Fig. 1). The licensure tracks recruited into the study included (a) Early Childhood – English as Second Language, (b) Middle Childhood/Early Adolescent – English as Second Language, (c) Middle Childhood/Early Adolescent - Special Education, and (d) Middle Childhood/ Early Adolescent – Content Focused Minor. This design controlled for the possibility that licensure tracks enroll PSTs with different characteristics or systematically affect PSTs differently.

Participants signed up for a pre-test session at the beginning of the semester prior to assignment, read and signed the consent document, and then were guided through pre-testing by a trained undergraduate research assistant. At the end of the same semester (approximately 12 weeks later), participants completed a post-test assessment of the same measures. Approximately 6-months after post-test, which was around the time of full-time student teaching, measures were completed a third time (i.e., follow-up; see Fig. 2). Participants were not provided any indication of their performance on any outcome during or after the study. Roughly on the same timeframe, interviews were conducted. Participants were compensated $145 for complete participation. The study was preregistered prior to recruitment (https://clinicaltrials.gov/ct2/show/NCT02544412). All quantitative data and the R syntax used to conduct reported analyses are publicly available at https://osf.io/jcvb8/. Ethics board approval was obtained for all study procedures and materials.

7.7. Study design

This study was a cluster RCT involving four waves (Fig. 1). We wanted to explore the impacts of mindfulness and connection training integrated into teacher education and not separate from it. Researchers and EEP faculty reached an agreement wherein for the duration of the study, cohorts assigned to treatment would allot 30 min of weekly seminar time to intervention. Thus, all intervention cohort PSTs, regardless of status as research participants, received at least 30 min of intervention each week. Control cohort participants were asked to forgo training in mindfulness or similar practices until after follow-up testing. Research participants came an hour before or stayed an hour after the 30 min of integrated training (cohort schedule dependent) for a total of 90 min per week of contiguous intervention.

7.8. Intervention

A detailed description of the intervention as well as a sample lesson can be found in the Supplemental Materials. In short, the 9-week training was based on Mindfulness-Based Stress Reduction (Kabat-Zinn, 1982) and a traditional Tibetan Buddhist structure to mind training (Rinpoche, 1993). The intervention consisted of weekly 1.5-h classes along with two 4-h intensive practice days (21.5 h total class time). Over the follow-up period, an intervention instructor met with intervention cohorts once per month for 15 min to guide a short practice and discuss on-going practice related questions. During other follow-up weeks, intervention cohorts played a prerecorded, 15-min practice in a prescribed order. Mindfulness practices included breath awareness (monitoring the movement of the breath), body scan (scanning of body sensations), sound practice (using sounds as the anchor for attention), walking meditation, and mindfulness movement (using slow movement as the anchor for attention). Connection practices included generating feelings of warmth and well-wishing for oneself and others and contemplation of the caring intention to teach and strengthening that caring intention. Participants were asked to practice a mindfulness or connection practice for 10–20 min each day. All intervention cohorts were instructed by the same three individuals, each of whom had extensive classroom (> 40 years combined) as well as mindfulness teaching (> 5 years) and practice (> 10 years) experience.

7.9. Data analysis

Our preregistered analysis plan consisted of examining between-group changes over time (i.e., group by time interaction effects). We conducted intention-to-treat analyses with restricted maximum likelihood estimation using linear mixed effects models in which

participant linear change over time (pre-, post-, and follow-up) was regressed on the version of the IAT used at follow-up test, the group assignment variable (Control/Intervention), time (post-test/follow-up), and the interaction between group and time with participant-level random intercepts. The contrasts of interests were the group by post-test and group by follow-up test effects. Because the direction of IAT *D*-scores was meaningful, we included directional *D*-score estimates in analyses. We report simple unadjusted within-group percent change as an easily interpretable effect size (Baguley, 2009).

We conducted several types of sensitivity analyses to examine the robustness of our results. First, we compared results from complete case and intention-to-treat with maximum likelihood estimation in a linear mixed effect model. Second, because randomization occurred at the cohort level, not accounting for possible cohort effects on participant level outcomes can produce biased estimates (Clarke, 2008). However, hierarchal linear modeling (HLM) is not suggested with less than 10 level-two units (i.e., cohorts; Maas & Hox, 2005). To examine the possibility of bias resulting from cohort effects, we compared underpowered HLM models to our primary model and extract variance estimates from the HLM model to compute the intraclass correlation coefficients (ICC). The ICC provide an estimate of dependency that in this case can be understood as the correlation between two randomly selected individuals from within the same cohort. If the ICC are very small, substantive bias from not modeling cohort effects is unlikely. We also included cohort as a fixed effect in linear mixed effects analyses, comparing models with and without it.

Finally, we conducted pattern-mixture modeling sensitivity analyses on the missing at random (MAR) assumption (Iddrisu & Gumedze, 2019; Leurent et al., 2018). At post and follow-up test, missingness was significantly higher in the control group (38.71% to 15.79%, $p = .011$ and 45.15% to 21.05%, $p = .031$, respectively). We first multiply imputed 50 datasets while retaining the hierarchical structure of the data (i.e., timepoints within participants) using a clustered multivariate imputation through chained equations approach (Drechsler, 2015). We then scaled imputed values by assuming that missingness was 10%, 20%, and 30% worse than the imputed values (which are imputed based on the distribution of observed scores conditional on predictors in the imputation; i.e., MAR). We then estimated our linear mixed effect model on each scaled imputation separately, pooling the respective 50 results from each assumption (i.e., 10%, 20%, and 30% worse than MAR) according to Rubin's rules (2004). Detailed descriptions of each procedure are available in the Supplemental Materials.

7.10. Interview data coding

Interviews were audio recorded and later transcribed by undergraduate research assistants. Transcriptions were read and coded in a 3-stage process. In Step 1, the protocol designers coded data by using both inductive coding techniques such as descriptive and open coding, as well as a deductive coding process drawn from constructs aligned with the study design. The deductive codes were based on participant experiences related to mindfulness practice, mindful awareness in daily life (including in the classroom), and the intersections of mindfulness with teacher education. Researchers wrote on-going analytic memos to preserve shifting insights on emerging categories. In Step 2, the categories and patterns were organized into broader themes that captured the complexity and variability of PST experiences. Throughout this process, the excerpts, categories, and memos were discussed between the researchers to establish intercoder reliability. This process was iterative. Final excerpts coded by this iterative process, and highlighted in themes, were then provided to the first author of this manuscript. In Step 3, the first author used deductive coding to filter relevant excerpts into one of three general categories: (a) increased awareness of thoughts, sensations, and emotions; (b) increased ability to self-regulate or reduced automatic responding; and (c) reflections on race, bias, or group affiliation. We report the percent of participants that discussed each theme and provide exemplars quotes to illustrate participant experiences (Also see Supplemental Material Table S3).

8. Results

8.1. Group equivalence at baseline

There were no significant between group differences at baseline on the adult or child versions of the race IAT or any demographic variable ($ps > 0.05$). Descriptive statistics reporting performance on the IATs by group and time are presented in Table 1.

Table 1
Adult and child race implicit association tests' D-score means (SD) by group and time.

Measure	Treatment			Control		
	Mean (*SD*) Range			Mean (*SD*) Range		
	T1	T2	T3	T1	T2	T3
Child Race IAT	0.31 (0.29) −0.61, 0.80	0.22 (0.28) −0.43, 0.76	0.25 (0.33) −0.81, 0.89	0.19 (0.32) −0.49, 1.17	0.20 (0.35) −0.37, 1.17	0.26 (0.35) −0.52, 0.81
Adult Race IAT	0.36 (0.37) −0.89, 1.04	0.28 (0.37) −0.81, 0.88	0.27 (0.35) −0.54, 0.77	0.29 (0.38) 0.-53, 0.80	0.32 (0.33) −0.29, 0.84	0.37 (0.37) −0.65, 1.03

Note. IAT = Implicit association test. T1 = Pre-test. T2 = Post-test. T3 = Follow-up test. Descriptive statistics are from observed data. *D*-scores <0.34 are considered slight, ≥ 0.35 < 0.65 moderate, and ≥ 0.65 (Greenwald et al., 2003). A negative *D*-score indicates bias favoring Black face/disfavoring White faces whereas a positive *D*-score indicates bias favoring White faces/disfavoring Black faces.

8.2. IAT exclusions

IATs were examined for trial latencies <400 ms, > 10,000 ms, and error rates higher than 10%. One trial of one subject at follow-up test was removed for fast latency; no participants exceeded a 10% error rate on either IAT at any time point (Greenwald et al., 2003).

8.3. Treatment and implementation fidelity

We operationalized treatment fidelity as the rate of attendance and the amount of reported daily practice based on compensation guidelines for completion provided to participants during recruitment and consent. The compensation schedule defined full participation as greater than 85% attendance (i.e., > 8/9 classes). Seventy-two percent (41/57) of participants met full the criterion for full participation. Approximately 88% (50/57) of participants attended six or more intervention classes and at least one intensive day. Participants were asked to practice formally on their own for 10–20 min per day. Participants reported an average of 13.94 min per day (SD = 9.47).

We operationalized content fidelity as the percent of prescribed content delivered during the prescribed class. Instructors completed a checklist of planned and covered content following each class. Over 95% of content was delivered during the intended session, and there was close to total consistency (> 95%) on the content and timing between intervention waves. Over the follow-up period, the 15-min booster sessions were implemented weekly by all cohorts except one. This cohort held boosters once per month when instructors visited.

8.4. Change in automatic race bias

Full results are presented in Table 2. At both post-test and follow-up test, the intervention group demonstrated significant reductions on the child race IAT relative to the control group (post-test: Control = +8.51% / Intervention = −31.66%, β = −0.23, se = 0.10, $t(119.35)$ = −2.26, p = .026, and follow-up: Control = +38.30% / Intervention = −22.57%, β = −0.21, se = 0.10, $t(118.10)$ = −2.07, p = .040, respectively; see Fig. 3). The intervention group's automatic bias average moved from a moderate to a slight level of bias whereas the control group began and remained at an average slight level of bias on the child IAT. On the adult IAT, intervention group reductions were not significantly different than control group changes at post-test (Control = +11.42% / Intervention = −22.10%, β = −0.04, se = 0.12, $t(119.56)$ = −0.32, p = .753) or follow-up test (Control = +25.78% / Intervention = −24.86%, β = −0.07, se = 0.12, $t(118.67)$ = −0.599, p = .550). The intervention group's average moved from a moderate level of bias to a slight level of bias whereas the control group moved from a slight to a moderate level of bias on the adult IAT.

8.5. Sensitivity analyses

Full results from sensitivity analyses are reported in Supplemental Materials Tables S1–S3. Effect sizes from OLS regression complete case models, 2-level HLMs in which participants were nested within cohorts, and our primary intention-to-treat linear mixed effects models were equivalent. ICCs from the empty 2-level HLM with participants nested within cohorts were low (0.001–0.057), indicating that any bias resulting from potential cohort effects was minimal. Adding cohort as a fixed effect in ITT analyses did not substantively affect results; model comparison between ITT models with and without cohort favored the more parsimonious model without cohort (i.e., no significant difference between the models). Pattern mixture-model analyses examining MNAR assumptions produced equivalent results to ITT analyses. Intervention group changes on the child IAT were significantly different than controls at post-test and follow-up test when assuming missing data were 10%, 20%, and 30% worse than imputed values. We again observed no group differences on the adult IAT in MNAR analyses.

Table 2
Results from linear mixed effects models on the child and adult race IATs.

	Child Black/White Race IAT			Adult Black/White Race IAT		
	B	95% CI	*p*	*B*	95% CI	*p*
Fixed Effects						
(Intercept)	0.37	0.15, 0.59		0.37	0.11, 0.63	
Group	**0.19**	**0.02, 0.36**	**0.036**	−0.01	−0.21, 0.19	0.922
Time 2	0.06	−0.11, 0.22	0.519	−0.04	−0.24, 0.15	0.678
Time 3	0.08	−0.09, 0.24	0.376	−0.03	−0.23, 0.16	0.743
IAT version	**−0.13**	**−0.25, −0.01**	**0.038**	0.01	−0.13, 0.15	0.888
Group × Time 2	**−0.23**	**−0.42, −0.03**	**0.026**	−0.04	−0.27, 0.19	0.753
Group × Time 3	**−0.21**	**−0.40, −0.01**	**0.040**	−0.07	−0.30, 0.16	0.550
Random Effects	Variance					
Participant	0.035	0.13, 0.24		0.05	0.15, 0.28	
Residual	0.059	0.21, 0.27		0.08	0.25, 0.32	
Observations	183			184		

Note. All models controlled for IAT task version at follow-up test. Significant effects at p < .05 are bolded.
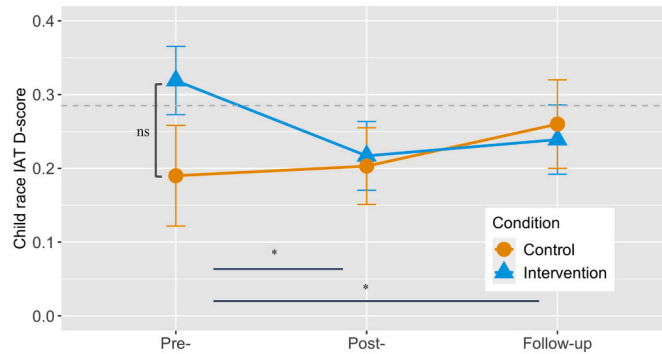
Fig. 3. Change over time on the child race IAT.

Note. Dashed line is the predicted teacher population *D*-score (i.e., bias) on the Child Race IAT. Observed data are plotted. At baseline the group difference on IAT D-scores was not statistically significant ($p > .05$). Error bars = SEM. ns $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$.

8.6. Probing effects on the child race IAT

Group scores on the child race IAT were not significantly different at baseline, but descriptively the intervention group had a higher *D*-score ($M = 0.32$, $SD = 0.29$; Control $M = 0.19$, $SD = 0.32$, $p = .063$, $d = 0.29$). We therefore probed the group by time interaction effects to examine regression to the mean as a possible explanation. Starck et al.'s (2020) large sample of teachers ($N = 68,930$) provides the best available estimate for the teacher population mean on the adult race IAT (D-score $= 0.32$, $SD = 0.45$). Nosek (2007) reported that average child race IAT bias was about 11% lower than bias on the adult IAT ($N = 28,816$). We therefore used 89.12% of Starck et al.'s (2020) reported average teacher score on the adult race IAT (i.e., *D*-score $= 0.29$) as our child race IAT population estimate. We first asked whether either group in our sample differed significantly from this population estimate of automatic bias. Using independent sample Welch's *t*-tests, we compared the intervention and control groups' baseline child race IAT *D*-score against this estimate (Intervention: $t(56.22) = 0.91$, $p = .367$; Control: $t(30.05) = -1.65$, $p = .109$). Next, we estimated within group change to determine if the observed effects were consistent with greater reductions in the intervention group as hypothesized, or unexpected increases in the control group that might be interpreted as regression to the mean. As expected, intervention group within group change at post-test was about 10 times as large in magnitude as the control group (Cohen's $d = -0.29$ and $d = 0.03$, respectively) and a little less than twice as large in magnitude at follow-up (Cohen's $d = -0.21$ and $d = 0.13$, respectively).

8.7. Interview data

A table with exemplar descriptions of participant experiences of increased awareness and reduced automaticity can be found in Supplemental Materials Table S4. All 25 (i.e., 100%) interviewed intervention participants described increased awareness of thoughts, sensations, or emotions as a result of the intervention at follow-up. Participants most commonly discussed this increased awareness in the context of classroom management when feelings of frustration or overwhelm would arise. Pointing toward reduced automaticity and enhanced self-regulation, most interviewees explained that, having noticed frustration or feeling overwhelmed, they would "take a step back" or "take a breath" to try and regain composure.

Similarly, nearly all interviewed intervention participants (22/25, 88%) commented on reduced automaticity of behaviors or increased self-regulation. Descriptions typically began with a clause reflecting enhanced awareness of thoughts or physical sensation that was followed by a top-down regulatory strategy. Using the language of the intervention, one participant described her growing recognition of automaticity in these words: "It's really easy to sort of auto pilot, put yourself on auto pilot mode, and just do things without thinking too much about it, so I have tried to do some of the breathing exercises both at school and at home."

None of the 25 interviewed participants described an enlarged sense of in-group affiliation. But although neither the intervention nor the interview protocol asked about issues of race and bias, which are phenomena at the heart of in-group preference, five interviewed participants (20%) reflected on the relationship of the intervention to these topics. Their reflections were consistent with Shah and Coles (2020) proposition of 'racial noticing,' or the expansion of teacher noticing to include race. When asked how the intervention may support instructional behaviors in the classroom, a White participant responded that,

> Ideally we would all take a step and think about our biases: but your first response is always going to be like, 'that's not true!' And I do think... mindfulness may have helped in that way... Being able to realize... maybe I am biased.

A different participant, also White, described the intersection of mindfulness and connection practices and reflection, rhetorically by asking herself, "How did I treat this student; was that equitable? And how do I change that for the next day?"

Perhaps most salient to teacher education, a third White participant described her frustration at the disconnect between learning about equitable practice and learning how to teach equitably:

> It's important to know of inequities in the education system, ...but... in a program where I had one class on inclusive schooling and nine classes on inequitable systems... when I'm about to go into a field where every day I will be trying to make education

accessible for all different kinds of students, I think that's a little silly… I would add more classes on mindfulness, learning how to be mindful… and more classes on how to be inclusive.

Another White PST concurred by asking,

How can we integrate that [discussion of race and equity] into my classroom and then have us practice enacting that? … Social justice and equity is so so so important… and especially our students are experiencing it, I just wish it was intertwined better.

9. Discussion

In a sample that reflects the racial/ethnic composition of PSTs nationally, we present evidence that automatic race bias reductions towards children of a similar age to those who participants are being certified to teach are possible and can be sustained for at least an average of six-months. The long-term reductions we observed are, to our knowledge, the most durable reductions in automatic race bias reported to date outside of children. We theorized that three capacities putatively learned through mindfulness and connection practice might support reduced automatic race bias: (a) increased awareness, (b) reduced automatic responding, and (c) a more inclusive definition of the in-group affiliation. Semi-structured interviews conducted at follow-up with 25 intervention participants provided evidence that participants uniformly experienced enhanced awareness of thoughts, emotions, and sensations after the intervention, and that in nearly all cases preservice teachers reported that this enhanced awareness was translated into exercising greater regulation over automatic or reactive responding in the classroom. Although interviews did not elicit commentary on notions of group affiliation, without prompting 20% of interviewees made connections between their experience of mindfulness and connection practices and issues of race and bias.

The extent to which reducing automatic bias may be causally related to more equitable teaching behaviors is unknown. However, it is clear that teachers demonstrate equivalent levels of automatic (and explicit) bias to the rest of society and that bias assessed on race IATs is associated with inequitable behaviors and perceptions in a variety of contexts, including teaching. For instance, teacher levels of automatic race bias are associated with the ethnic achievement gap between classrooms (Van den Bergh et al., 2010), inequitable perceptions of and expectations for students, and indicated in inequitable disciplinary actions that in turn predict a host of negative outcomes (Gregory et al., 2010). Many PST and professional development programs include psychoeducational anti-bias curricula. We theorized that durable change in automatic race bias requires affecting the cognitive and affective processes that undergird them, first by increasing awareness of the phenomenon, and then by strengthening self-regulatory capacity so that this increased awareness can be effectively acted on. The quantitative findings demonstrate that long-term reductions in automatic race bias are possible. The qualitative data suggest that the intervention increased awareness of thoughts, sensations, and emotions and that this increased awareness frequently stimulated self-regulatory processing to adaptively interact with internal (e.g., emotions) and external (e.g., a student) experience.

Learning environments and therefore schools are defined by the quality of interactions that occur within them. Reducing teacher levels of automatic race bias is critically important to the ultimate goal of improving teacher practice and student outcomes. Teachers' ability to adaptively self-regulate are associated with positive teacher outcomes (e.g., less stress and burnout, greater well-being) and better instructional practice (Klusmann et al., 2008). The primary outcome in the larger study from which these data were drawn was an observer rating of classroom teaching on which the intervention group showed significant improvements at follow-up (Hirshberg et al., 2020). Taken together, these results suggest that integrating mindfulness and connection practices into a PST program reduces automatic race bias, improves classroom teaching practices, and increases awareness and self-regulation. The potential impacts on school systems of training teacher responsivity, the tools to regulate reactivity, and reductions in automatic race bias are substantial for more equitable student academic achievement and also in terms of the potential social-emotional learning that might occur as a result of teachers modeling higher levels of awareness, attentiveness, and adaptive self-regulation.

Research replicating the present results and extending them by examining whether reduced bias is related to more equitable teaching, ideally using causal mediation modeling, is needed. In addition, further research is required to understand the discrepant findings between changes on the child and adult versions of the IAT. We are not aware of prior research examining in the same sample changes on the adult and child race versions of the IAT in response to an intervention. The observed discrepancy in changes on these measures may reflect the role of motivation in bias reduction (Devine et al., 2012; Jost, 2018). All participants in this study had selected into an elementary PST education program in order to earn state certification to teach Grades prek–8. We assume that they were motivated at least to a degree by the desire to support child development. One potential explanation then for the discrepant impacts on the IATs is that the intervention harnessed this motivation. Because the motivation to teach was specific to children, it may have manifested in reductions of automatic race bias toward child but not adult faces. Also relevant, the intervention group did reduce adult IAT *D*-scores at post- and follow-up test relative the control group, but the magnitude of the reductions was smaller and the variability in change larger than on the child version. Putting aside the binary of statistical significance, one interpretation of this pattern of results is that motivation might enhance the magnitude and consistency of intervention effects on bias of the most salient target (i.e., child faces), relative to no less important but less contextually salient other targets (i.e., adult faces).

When interpreting these results, the homogeneity of the gender composition of this sample (95.6% female) should be considered. Less than 5% of the sample were male, whereas around 23.5% of teachers nationally identify as male. Starck et al. (2020) reported that male teachers demonstrated higher levels of automatic and explicit bias. Continued research with greater numbers of male PSTs is needed to understand whether interventions like the one researched here impact male and female identifying PSTs equivalently.

Although this sample was predominantly White (86.36%), it is representative of the national teaching corps (i.e., 80% of teachers are White; U.S. Department of Education, 2018). Prior research has shown that White participants on average demonstrate preference

for White faces/bias for Black faces whereas Black participants do not show systematic Black/White biases (Nosek et al., 2007). Many studies using race IATs restrict sampling to White participants as a result, and most inequitable teaching practices associated with automatic race bias (e.g., discipline, student expectations) have been observed in the cultural majority race (e.g., White) but not in the cultural minority race (e.g., Black). Therefore, reductions of automatic race bias in this predominately White sample are particularly relevant for a majority of America's teachers, who are White.

9.1. Limitations

There are a few important limitations to consider. Protocol violations resulted in the removal of one control cohort from analyses, affecting the balance of the blocked randomization design. Although the reduction in sample size as well the effect on design symmetry should be considered when evaluating these results, we did not find evidence of group differences at baseline after removing this cohort. In addition, higher control group attrition coupled with a smaller sample to begin with resulted in a relatively high proportion of missingness in that group. Although our statistical methods are robust to MAR data and sensitivity analyses show that effects are robust to a variety of MNAR assumptions, these characteristics reinforce the importance of replication, in particular with large samples that will allow further investigation of putative mechanisms of change. A final limitation is that the interview protocol did not inquire into participant perceptions of in-group boundaries or bias. Future studies should probe participants to reflect on these areas to better understand whether mindfulness/connection practices elicit changing thoughts and perceptions about group affiliation and bias.

10. Conclusions

Automatic race bias is pervasive, especially among White Americans, and likely contributes to inequitable teaching behaviors and student outcomes (Okonofua et al., 2016; Peterson et al., 2016; Van den Bergh et al., 2010). This study provides evidence that automatic race bias favoring White and disfavoring Black child faces was reduced by 31.00% at post-test and 25.63% at follow-up in a sample of predominately White undergraduate PSTs. Intervention participants almost uniformly reported increased awareness and self-regulation, and reduced automatic responding in teaching contexts following the intervention. Coupled with prior research on the benefits of mindfulness/connection training for teachers, findings from this study suggest that integrating these practices into teacher education may improve outcomes. Further research on the potential benefits of mindfulness and connection practices during teacher education, including studies designed to examine putative mechanisms of change, are warranted.

Author note

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsp.2021.12.002.

References

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology, 100*(3), 603–617. https://doi.org/10.1348/000712608X377117

Carter, P. L., & Welner, K. G. (2013). *Closing the opportunity gap: What America must do to give every child an even chance.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199982981.001.0001

Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. long-term change and stability from 2007 to 2016. *Psychological Science, 30*(20), 174–192. https://doi.org/10.1177/0956797618813087

Clarke, P. (2008). When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiology & Community Health, 62*(8), 752–758. https://doi.org/10.1136/jech.2007.060798

Coleman, J., Campbell, E., Hobson, C., & MacPartland, J. (1966). *Equality of educational opportunity.* U.S. Government Printing Office.

Creswell, J. W. (2013). *Qualitative inquiry & research design. Choosing among five approaches* (3rd ed.). Sage Publications.

Dahl, C. J., Lutz, A., & Davidson, R. J. (2015). Reconstructing and deconstructing the self: Cognitive mechanisms in meditation practice. *Trends in Cognitive Sciences, 19*(9), 515–523. https://doi.org/10.1016/j.tics.2015.07.001

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind'. *Race Ethnicity and Education, 10*(3), 245–260. https://doi.org/10.1080/13613320701503207

De Houwer, J. (2019). Implicit bias is behavior: A functional-cognitive perspective on implicit bias. *Perspectives on Psychological Science, 14*(5), 835–840. https://doi.org/10.1177/1745691619855638

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology, 48*(6), 1267–1278. https://doi.org/10.1016/j.jesp.2012.06.003

Drechsler, J. (2015). Multiple imputation of multilevel missing data—Rigor versus simplicity. *Journal of Educational and Behavioral Statistics, 40*(1), 69–95. https://doi.org/10.3102/1076998614563393

Emerson, L.-M., Leyland, A., Hudson, K., Rowse, G., Hanley, P., & Hugh-Jones, S. (2017). Teaching mindfulness to teachers: A systematic review and narrative synthesis. *Mindfulness, 8*(5), 1136–1149. https://doi.org/10.1007/s12671-017-0691-4

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

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

Friedrich, A., Flunger, B., Nagengast, B., Jonkmann, K., & Trautwein, U. (2015). Pygmalion effects in the classroom: Teacher expectancy effects on students' math achievement. *Contemporary Educational Psychology, 41*, 1–12. https://doi.org/10.1016/j.cedpsych.2014.10.006

Gershenson, S., Hart, C., Hyman, J., Lindsay, C., & Papageorge, N. W. (2018). The long-run impacts of same-race teachers. *National Bureau of Economic Research.*. https://doi.org/10.3386/w25254

Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions. *Yale University Child Study Center, 9*(28). https://medicine.yale.edu/childstudy/zigler/publications/Preschool%20Implicit%20Bias%20Policy%20Brief_final_9_26_276766_5379_v1.pdf.

Glock, S., & Karbach, J. (2015). Preservice teachers' implicit attitudes toward racial minority students: Evidence from three implicit measures. *Studies in Educational Evaluation, 45*, 55–61. https://doi.org/10.1016/j.stueduc.2015.03.006

Glock, S., Kneer, J., & Kovacs, C. (2013). Preservice teachers' implicit attitudes toward students with and without immigration background: A pilot study. *Studies in Educational Evaluation, 39*(4), 204–210. https://doi.org/10.1016/j.stueduc.2013.09.003

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the implicit association test can have societally large effects. *Journal of Personality and Social Psychology, 108*(4), 553–561. https://doi.org/10.1037/pspa0000016

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

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

Grimmer, J., Hersh, E., Meredith, M., Mummolo, J., & Nall, C. (2018). Obstacles to estimating voter ID laws' effect on turnout. *The Journal of Politics, 80*(3), 1045–1051. https://doi.org/10.1086/696618

Groenewald, T. (2004). A phenomenological research design illustrated. *International Journal of Qualitative Methods, 3*(1), 42–55. https://doi.org/10.1177/160940690400300104

Hachfeld, A., Hahn, A., Schroeder, S., Anders, Y., Stanat, P., & Kunter, M. (2011). Assessing teachers' multicultural and egalitarian beliefs: The teacher cultural beliefs scale. *Teaching and Teacher Education, 27*(6), 986–996. https://doi.org/10.1016/j.tate.2011.04.006

Halberstadt, A. G., Castro, V. L., Chu, Q., Lozada, F. T., & Sims, C. M. (2018). Preservice teachers' racialized emotion recognition, anger bias, and hostility attributions. *Contemporary Educational Psychology, 54*, 125–138. https://doi.org/10.1016/j.cedpsych.2018.06.004

Halberstadt, A. G., Cooke, A. N., Garner, P. W., Hughes, S. A., Oertwig, D., & Neupert, S. D. (2020). Racialized emotion recognition accuracy and anger bias of children's faces. In *Emotion*. https://doi.org/10.1037/emo0000756. Advance online publication.

Hirshberg, M. J., Flook, L., Enright, R. D., & Davidson, R. J. (2020). Integrating mindfulness and connection practices into preservice teacher education improves classroom practices. *Learning and Instruction, 66*, Article 101298. https://doi.org/10.1016/j.learninstruc.2019.101298

Hirshberg, M. J., Goldberg, S. B., Schaefer, S. M., Flook, L., Findley, D., & Davidson, R. J. (2018). Divergent effects of brief contemplative practices in response to an acute stressor: A randomized controlled trial of brief breath awareness, loving-kindness, gratitude or an attention control practice. *PLoS One, 13*(12), Article e0207765. https://doi.org/10.1371/journal.pone.0207765

Hofmann, S. G., Grossman, P., & Hinton, D. E. (2011). Loving-kindness and compassion meditation: Potential for psychological interventions. *Clinical Psychology Review, 31*(7), 1126–1132. https://doi.org/10.1016/j.cpr.2011.07.003

Holbein, J. B., & Ladd, H. F. (2017). Accountability pressure: Regression discontinuity estimates of how no child left behind influenced student behavior. *Economics of Education Review, 58*, 55–67. https://doi.org/10.1016/j.econedurev.2017.03.005

Hopthrow, T., Hooper, N., Mahmood, L., Meier, B. P., & Weger, U. (2017). Mindfulness reduces the correspondence bias. *The Quarterly Journal of Experimental Psychology, 70*(3), 351–360. https://doi.org/10.1080/17470218.2016.1149498

Iddrisu, A.-K., & Gumedze, F. (2019). An application of a pattern-mixture model with multiple imputation for the analysis of longitudinal trials with protocol deviations. *BMC Medical Research Methodology, 19*(1), 10. https://doi.org/10.1186/s12874-018-0639-y

Jennings, P. A., Brown, J. L., Frank, J. L., Doyle, S., Oh, Y., Davis, R., … Greenberg, M. T. (2017). Impacts of the CARE for Teachers Program on teachers' social and emotional competence and classroom interactions. *Journal of Educational Psychology, 109*(7), 1010–1028. https://doi.org/10.1037/edu0000187

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research, 79*(1), 491–525. https://doi.org/10.3102/0034654308325693

Jost, J. T. (2018). The IAT is dead, long live the IAT: Context-sensitive measures of implicit attitudes are indispensable to social and political psychology. *Current Directions in Psychological Science, 28*(1), 10–19. https://doi.org/10.1177/0963721418797309

Kabat-Zinn, J. (1982). An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. *General Hospital Psychiatry, 4*(1), 33–47. https://doi.org/10.1016/0163-8343(82)90026-3

Kabat-Zinn, J. (2013). *Full catastrophe living, revised edition: How to cope with stress, pain and illness using mindfulness meditation. Hachette UK*.

Kabat-Zinn, J., & Salzberg, S. (2004). *Lovingkindness: The revolutionary art of happiness*. Shambhala Publications.

Kang, Y., Gray, J. R., & Dovidio, J. F. (2014). The nondiscriminating heart: Lovingkindness meditation training decreases implicit intergroup bias. *Journal of Experimental Psychology: General, 143*(3), 1306–1313. https://doi.org/10.1037/a0034150

Kang, Y., Gruber, J., & Gray, J. R. (2013). Mindfulness and de-automatization. *Emotion Review, 5*(2), 192–201. https://doi.org/10.1177/1754073912451629

Kiken, L. G., & Shook, N. J. (2011). Looking up: Mindfulness increases positive judgments and reduces negativity bias. *Social Psychological and Personality Science, 2*(4), 425–431. https://doi.org/10.1177/1948550610396585

Klingbeil, D. A., & Renshaw, T. L. (2018). Mindfulness-based interventions for teachers: A meta-analysis of the emerging evidence base. *School Psychology Quarterly, 33*(4), 501–511. https://doi.org/10.1037/spq0000291

Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Teachers' occupational well-being and quality of instruction: The important role of self-regulatory patterns. *Journal of Educational Psychology, 100*(3), 702–715. https://doi.org/10.1037/0022-0663.100.3.702

Kumar, R., Karabenick, S. A., & Burgoon, J. N. (2015). Teachers' implicit attitudes, explicit beliefs, and the mediating role of respect and cultural responsibility on mastery and performance-focused instructional practices. *Journal of Educational Psychology, 107*(2), 533–545. https://doi.org/10.1037/a0037471

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., & Banaji, M. R. (2018). Relationship between the implicit association test and intergroup behavior: A meta-analysis. *American Psychologist, 74*(5), 569–586. https://doi.org/10.31234/osf.io/582gh

Ladson-Billings, G. (1995). But that's just good teaching! The case for culturally relevant pedagogy. *Theory Into Practice, 34*(3), 159–165. https://doi.org/10.1080/00405849509543675

Ladson-Billings, G. (2003). Just what is critical race theory and what's it doing in a nice field like education? *International Journal of Qualitative Studies in Education, 16*(6), 883–886. https://doi.org/10.1080/714858243

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., … Burns, M. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General, 145*(8), 1001–1016. https://doi.org/10.31234/osf.io/ygscv

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the implicit association test: IV: What we know (so far) about the method. In B. Wittenbrink, & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 59–102). Guilford.

Leurent, B., Gomes, M., Faria, R., Morris, S., Grieve, R., & Carpenter, J. R. (2018). Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: A tutorial. *Pharmacoeconomics, 36*(8), 889–901. https://doi.org/10.1007/s40273-018-0650-5

Lindsay, E. K., & Creswell, J. D. (2017). Mechanisms of mindfulness training: Monitor and acceptance theory (MAT). *Clinical Psychology Review, 51*, 48–59. https://doi.org/10.1016/j.cpr.2016.10.011

Lueke, A., & Gibson, B. (2015). Mindfulness meditation reduces implicit age and race bias: The role of reduced automaticity of responding. *Social Psychological and Personality Science, 6*(3), 284–291. https://doi.org/10.1177/1948550614559651

Lueke, A., & Gibson, B. (2016). Brief mindfulness meditation reduces discrimination. *Psychology of Consciousness: Theory, Research and Practice, 3*(1), 34–44. https://doi.org/10.1037/cns0000081

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.85

Markowitz, A. J. (2018). Changes in school engagement as a function of no child left behind: A comparative interrupted time series analysis. *American Educational Research Journal, 55*(4), 721–760. https://doi.org/10.3102/0002831218755668

Marx, S. (2006). *Revealing the invisible: Confronting passive racism in teacher education.* Routledge.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*(5), 435–442. https://doi.org/10.1006/jesp.2000.1470

McKown, C. (2013). Social equity theory and racial-ethnic achievement gaps. *Child Development, 84*(4), 1120–1136. https://doi.org/10.1111/cdev.12033

Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science, 16*(2), 65–69. https://doi.org/10.1111/j.1467-8721.2007.00477.x

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*(2), 166–180. https://doi.org/10.1177/0146167204271418

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences, 15*(4), 152–159. https://doi.org/10.1016/j.tics.2011.01.005

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., & Greenwald, A. G. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*(1), 36–88. https://doi.org/10.1080/10463280701489053

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., 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

Peterson, E. R., Rubie-Davies, C., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction, 42*, 123–140. https://doi.org/10.1016/j.learninstruc.2016.01.010

Philip, T. M., Souto-Manning, M., Anderson, L., Horn, I., Carter Andrews, D., Stillman, J., & Varghese, M. (2019). Making justice peripheral by constructing practice as "core": How the increasing prominence of core practices challenges teacher education. *Journal of Teacher Education, 70*(3), 251–264. https://doi.org/10.1177/0022487118798324

Psychology Software Tools, Inc. (2012). *E-Prime 2.0.*

Reyna, C. (2000). Lazy, dumb, or industrious: When stereotypes convey attribution information in the classroom. *Educational Psychology Review, 12*(1), 85–110. https://doi.org/10.1023/A:1009307101170

Reyna, C. (2008). Ian is intelligent but Leshaun is lazy: Antecedents and consequences of attributional stereotypes in the classroom. *European Journal of Psychology of Education, 23*(4), 439–458. https://doi.org/10.1007/bf03172752

Rinpoche, P. (1993). *The heart treasure of the enlightened ones: The practice of view, meditation, and action.* Shambhala Publications.

Roeser, R. W., Schonert-Reichl, K. A., Jha, A., Cullen, M., Wallace, L., Wilensky, R., Oberle, E., Thomson, K., Taylor, C., & Harrison, J. (2013). Mindfulness training and reductions in teacher stress and burnout: Results from two randomized, waitlist-control field trials. *Journal of Educational Psychology, 105*(3), 787–804. https://doi.org/10.1037/a0032093

Rosenthal, R. (2010). Pygmalion effect. In *The Corsini encyclopedia of psychology* (pp. 1–2). American Cancer Society. https://doi.org/10.1002/9780470479216.corpsy0761.

Saldaña, J. (2021). *The coding manual for qualitative researchers.* Sage.

Shah, N., & Coles, J. A. (2020). Preparing teachers to notice race in classrooms: Contextualizing the competencies of preservice teachers with antiracist inclinations. *Journal of Teacher Education, 71*(5), 584–599. https://doi.org/10.1177/0022487119900204

Shapiro, S. L., Carlson, L. E., Astin, J. A., & Freedman, B. (2006). Mechanisms of mindfulness. *Journal of Clinical Psychology, 62*(3), 373–386. https://doi.org/10.1002/jclp.20237

Sherif, M. (1958). Superordinate goals in the reduction of intergroup conflict. *American Journal of Sociology, 63*(4), 349–356. https://doi.org/10.1086/222258

Skiba, R. J., Arredondo, M. I., & Williams, N. T. (2014). More than a metaphor: The contribution of exclusionary discipline to a school-to-prison pipeline. *Equity & Excellence in Education, 47*(4), 546–564. https://doi.org/10.1080/10665684.2014.958965

Snyder, T. D., de Brey, C., & Dillow, S. A. (2016). *Digest of education statistics 2014, NCES 2016-006.* National Center for Education Statistics. https://eric.ed.gov/?id=ED565675.

Snyder, T. D., & Dillow, S. A. (2012). *Digest of education statistics 2011.* National Center for Education Statistics.

Starck, J. G., Riddle, T., Sinclair, S., & Warikoo, N. (2020). Teachers are people too: Examining the racial bias of teachers compared to other American adults. *Educational Researcher, 49*(4), 273–284. https://doi.org/10.3102/0013189X20912758

Stell, A. J., & Farsides, T. (2016). Brief loving-kindness meditation reduces racial bias, mediated by positive other-regarding emotions. *Motivation and Emotion, 40*(1), 140–147. https://doi.org/10.1007/s11031-015-9514-x

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*(2), 253–273. https://doi.org/10.1037/0022-0663.99.2.253

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237–246. https://doi.org/10.1177/1098214005283748

U.S. Department of Education. (2018a). Digest of education statistics, 2016. https://nces.ed.gov/programs/digest/d17/tables/dt17_209.10.asp?current=yes.

U.S. Department of Education. (2018b). Characteristics of public school teachers who completed alternative route to certification programs. Spotlights—The condition of education 2018. https://nces.ed.gov/programs/coe/indicator_tlc.asp.

Vago, D. R., & Silbersweig, D. A. (2012). Self-awareness, self-regulation, and self-transcendence (S-ART): A framework for understanding the neurobiological mechanisms of mindfulness. *Frontiers in Human Neuroscience, 6*, 296. https://doi.org/10.3389/fnhum.2012.00296

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal, 47*(2), 497–527. https://doi.org/10.3102/0002831209353594

Vuletich, H. A., & Payne, B. K. (2019). Stability and change in implicit bias. *Psychological Science, 30*(6), 854–862. https://doi.org/10.1177/0956797619844270

Wolman, P. (2010). *Status of the American public school teacher, 2005-2006.* National Education Association Research Department. http://eric.ed.gov/?id=ED521866.

Yin, R. K. (1994). *Case study research: Design and methods.* Sage Publications.