



# A Cluster Randomized Pilot Trial of the Equity-Explicit Establish-Maintain-Restore Program among High School Teachers and Students

Mylien T. Duong<sup>1</sup> · Larissa M. Gaias<sup>2</sup> · Eric Brown<sup>3</sup> · Sharon Kiche<sup>1</sup> · Lillian Nguyen<sup>1</sup> · Catherine M. Corbin<sup>1</sup> · Cassandra J. Chandler<sup>1</sup> · Joanne J. Buntain-Ricklefs<sup>1</sup> · Clayton R. Cook<sup>4</sup>

Accepted: 21 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Student–teacher relationships are important to student outcomes and may be especially pivotal at the high school transition and for minoritized racial/ethnic groups. Although interventions exist to improve student–teacher relationships, none have been shown to be effective among high school students or in narrowing racial/ethnic disparities in student outcomes. This study was conducted to examine the effects of an equity-explicit student–teacher relationship intervention (Equity-Explicit Establish Maintain Restore, or E-EMR) for high school teachers and students. A cluster-randomized pilot trial was conducted with 94 ninth grade teachers and 417 ninth grade students in six high schools. Teachers in three schools were randomized to receive E-EMR training and follow-up supports for one year. Teachers in three control schools conducted business as usual. Student–teacher relationships, sense of school belonging, academic motivation, and academic engagement were collected via student self-report in September and January of their ninth-grade year. Longitudinal models revealed non-significant main effects of E-EMR. However, there were targeted benefits for students who started with low scores at baseline, for Asian, Latinx, multicultural, and (to a lesser extent) Black students. We also found some unexpected effects, where high-performing and/or advantaged groups in the E-EMR condition had less favorable outcomes at post, compared to those in the control group, which may be a result of the equity-explicit focus of E-EMR. Implications and directions for future research are discussed.

**Keywords** Student–teacher relationships · Professional development · High school · Equity · Race/ethnicity

## A Cluster Randomized Controlled Trial of Establish-Maintain-Restore Among High School Teachers and Students

A large body of research shows that when students have close, trusting relationships with their teachers, they are more likely to exhibit a host of other positive social, emotional, and academic outcomes (Quin, 2017; Roorda et al., 2011). Meta-analyses, conducted with students ranging from pre through high school, found student–teacher relationships (STR) to be associated with a number of student outcomes, including engagement, sense of belonging, self-esteem and social skills, in addition to improved behavior, academic achievement, lower suspension rates, and reduced risk of dropout (Quin, 2017; Roorda et al., 2011). These associations have been documented cross sectionally and longitudinally, with some studies showing effects up to four years later (e.g., Valiente et al., 2019). Further, evidence

---

✉ Mylien T. Duong  
myliend@uw.edu

<sup>1</sup> Department of Psychiatry and Behavioral Medicine, University of Washington, Seattle, WA, US

<sup>2</sup> Department of Psychology, University of Massachusetts, Lowell, Lowell, MA, US

<sup>3</sup> Department of Public Health Sciences, University of Miami, Miami, FL, US

<sup>4</sup> Department of Educational Psychology, University of Minnesota, Minneapolis, MN, US

from a recent meta-analysis of the intervention literature revealed that universal programs aimed at improving STR have an average effect size of  $d = 0.26$  on relationship quality, with some programs having substantially large effects (Kincade et al., 2020). Notably, 20 of 21 studies included were conducted with elementary school students; only one was conducted in middle school. No studies were conducted with high school students, leaving a significant gap in the student–teacher relationship literature. In addition to the lack of intervention research at the high school level, there are notable disparities in STR among racially minoritized students (Hughes et al., 2005; Thijs et al., 2012), warranting programs that are equity-explicit (Gregory et al., 2016). Given this, the purpose of this study was to conduct a randomized pilot trial of an existing student–teacher relationship program adapted to be equity-explicit for use with ninth students transitioning into high school.

## STR in High School

At the transition to high school, academic engagement and performance drop, and risk of dropout increases, for many students (Kennelly & Monrad, 2007). Not surprisingly, this transition portends the greatest risk for students with existing risk factors and those from minoritized racial/ethnic groups (EPE Research Center, 2006; Wheelock & Miao, 2005). The difficulties exhibited by students at this transition reflect the multitude of challenges facing entering 9th graders. The typical high school is larger and more bureaucratic than middle schools, which can lead to a sense of depersonalization and a lack of community (Lee & Smith, 2001). Classrooms are more socially comparative and competitive (Roeser et al., 2002), and there are increased demands to prepare for college or career (Williamson et al., 2015). Students at the transition, especially those who eventually drop out, report liking school less, feeling less connected to school and teachers, and generally feeling unmotivated and disengaged (Gruman et al., 2008). At a time when teacher support may be especially helpful, students report spending less time with teachers and feeling poorly supported and monitored by adults in their school (Barber & Olsen, 2004). Yet, how students adjust to the high school transition has significant implications for their long-term academic success (Archambault et al., 2009). Because of the challenges associated with the high school transition, interventions may be particularly powerful at this critical juncture. As Hertzog and Morgan (1999) noted, “students will decide during the first few weeks of their freshman year if they intend to be engaged in high school” (p. 27). Well-timed interventions during this sensitive period can trigger a series of reciprocally reinforcing positive interactions between the youth and their social system (Yeager & Walton, 2011).

Despite these risks, there is a dearth of STR interventions that are appropriate for high school students. This mirrors developmental patterns in social, emotional, and behavioral research in education in general, where the vast majority of studies focus on the elementary and middle school levels (e.g., Durlak et al., 2011). In addition, the dominant paradigm in STR research is grounded in attachment theory and conceptualizes STR quality along three dimensions: closeness, conflict, and dependency (Pianta et al., 2002). These aspects of STR are theoretically most important to development in early and middle childhood (O’Connor et al., 2011; Pianta & Nimetz, 1991). Compared to younger students, high school students have better-developed self-regulatory (Raffaelli et al., 2005) and social perspective-taking skills (Diazgranados et al., 2016). They rely less on dyadic interactions with teachers to co-regulate and moderate conflict in the classroom (Kurki et al., 2016). Thus, dependency in the STR may be less relevant for adolescents (Koomen & Jellesma, 2015). In contrast, constructs such as control, democratic socialization, maturity, autonomy support, competence, and relatedness may be more relevant for high schoolers (Brinkworth et al., 2018).

## STR for Minoritized Racial/Ethnic Groups

Although the associations between STR and student outcomes have been established across demographic groups and contexts, positive STR may be especially protective for minoritized youth (McHugh et al., 2013). Yet, teachers are least likely to have positive cross-racial relationships with students (Thijs et al., 2012). Many scholars have attributed these racial disparities to the cultural mismatch between students and teachers, which are compounded by a lack of effective professional development (Milner, 2005). More than 80% of US teachers are White and female, which is a mismatch for the increasingly diverse student population (Howard, 2016). Additionally, many teacher candidates have limited exposure to cultural diversity, endorse skepticism that racism exists, and are concerned that discussing race may be harmful (Gay, 2010; Kyles & Olafson, 2008). These dispositions lend to “color-blind” approaches and low efficacy integrating race and culture into their teaching (Gay, 2010; Kyles & Olafson, 2008).

The racial disparities in STR have consequences for students’ school outcomes. The quality of the relationship predicts teachers’ perceptions of student academic ability, which may contribute to racial/ethnic differences in academic outcomes (Kellow & Jones, 2008; van den Bergh et al., 2010). STR represent one point in a recursive cycle that exaggerates student differences over time. Students who receive more positive attention demonstrate better behavior and improved academics over time, which in turn

enhances teachers' perception of their behavior and capability (Wang & Eccles, 2012).

Such findings suggest that any STR intervention needs to be equity-explicit. Fortunately, a number of equity-explicit interventions are emerging, primarily targeting classroom management practices with the aim of reducing racial/ethnic disparities in discipline referrals. For instance, Greet-Stop-Prompt (GSP) trains teachers to recognize how implicit biases may impact their responses to student misbehavior and to regulate their behavior to prevent exclusionary discipline. GSP was found to be effective at reducing relative risk ratios of office referrals for Black boys (Cook, Duong, et al., 2018a). Similarly, Double Check aims to improve teachers' culturally responsive classroom practices by having teachers examine how their own identity and cultural background interact with their students' identities and impact their responses to student behavior. Double Check has led to improved teacher classroom management practices and student behavior (Bradshaw et al., 2018). In sum, training teachers to recognize and mitigate implicit bias is a promising avenue for improving the equity of teacher-delivered interventions.

## Previous Research with EMR

In previous research on EMR, we tested the intervention's effects among elementary and middle school teachers and students in two randomized-controlled trials. In both trials, we found that EMR improved STR, and observer-rated disruptive behavior and academically engaged time, with large effect sizes (*ds* ranging from 0.61 to 1.32 over a control group) (references removed for blinding). However, EMR's developmental appropriateness and effectiveness for high school students have not been determined. Further, our previous study showed that, although student race/ethnicity did not moderate EMR effects, there were significant differences in STR quality between White and non-White students at baseline that remained at post-intervention (reference removed for blinding) (Duong, et al., 2019; Gaias, et al., 2020). This suggests that an explicit equity component may be needed to enhance the intervention's ability to disrupt extant disparities in STR. We previously developed an equity-explicit version of Establish-Maintain-Restore (E-EMR). A pre-post pilot study of E-EMR in one high school demonstrated promising results for reducing disparities for students of color and for improving equity in teacher relationship practices (reference removed for blinding). However, there is a need to more rigorously examine the effects of E-EMR on student outcomes and on racial/ethnic disparities in these outcomes.

## The Current Study

This study reports the effects of a randomized controlled pilot trial testing the impact of E-EMR on student social, emotional, and behavioral outcomes, including STR, school belonging, academic motivation, and engagement. Pilot trials play an essential role in intervention science; they assess feasibility and allows investigators to detect signals of possible effects that can be tested in more robust, larger-scale efficacy trials. In addition to examining the main effects of E-EMR, we also examine for whom E-EMR is most effective by testing the moderating effects of students' race/ethnicity and their baseline scores.

## Method

### Setting and Participants

Participants included 94 teachers and 412 students recruited from 6 public high schools in the Pacific Northwest region of the USA. A nearest neighbor approach was used to block schools into pairs based on similarity in school size, percent of the student body who identified as Asian, Latinx, Black, Multiracial, and Other (including American Indian/Alaskan Native), and percent of the student body eligible for free or reduced priced lunch (FRL). We randomly assigned schools within pairs to the intervention or waitlist control condition. High schools served a relatively diverse student population racially ( $M = 52.3\%$  non-White) and socioeconomically ( $M = 27.5\%$  receiving FRL).

### Teacher Participants

Within both intervention and control schools, all 9th grade teachers were recruited to participate. Teachers had to meet two eligibility criteria: (1) be general education certified, and (b) 50% or more of their class rosters had to consist of 9th grade students. Of the 104 teachers eligible, 94 (90.4%) agreed to participate. Table 1 summarizes teacher-level demographics. As seen, teachers were primarily White (80.9%), with a mean age of 36.6 ( $SD = 10.14$ ) and an average of 9.3 years ( $SD = 7.98$ ) of teaching experience.

### Student Participants

A random subset of students was selected for recruitment based on class rosters provided by schools. For each participating teacher, we aimed to recruit five students from their classroom roster. Because we were interested in the impact of E-EMR on equity, we over-sampled non-White

**Table 1** Participant demographic characteristics

Demographic Variable	Teachers						Students					
	All		E-EMR		Control		All		E-EMR		Control	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<b>Gender</b>												
Male	38	40.4	17	39.5	21	41.2	204	49.2	93	49.2	111	49.1
Female	56	59.6	26	60.5	30	58.8	200	48.2	92	48.7	108	47.8
Other	0	0	0	0	0	0	5	1.2	2	1.1	3	1.3
Missing	0	0	0	0	0	0	6	1.4	2	1.1	4	1.8
Total	94	100	43	100	51	100	415	100	189	100	226	100
<b>Race/ethnicity</b>												
White	76	80.9	38	88.4	38	74.5	123	29.6	59	31.2	64	28.3
Asian	3	3.2	1	2.3	2	3.9	60	14.5	28	14.8	32	14.2
Latinx	1	1.1	0	0	1	2	57	13.7	23	12.2	34	15
Black	1	1.1	0	0	1	2	26	6.3	10	5.3	16	7.1
Multiracial	10	10.6	4	9.3	6	11.8	96	23.1	48	25.4	48	21.2
Other	2	2.1	0	0	2	3.9	21	5.1	5	2.6	16	7.1
Missing	1	1.1	0	0	1	2	32	7.7	16	8.5	16	7.1
Total	94	100	43	100	51	100	415	100	189	100	226	100
<b>FRL</b>												
Eligible	–	–	–	–	–	–	99	23.9	34	18	65	28.8
Not eligible	–	–	–	–	–	–	166	40.0	82	43.4	84	37.2
Missing	–	–	–	–	–	–	150	36.1	73	38.6	77	34.1
Total	–	–	–	–	–	–	415	100	189	100	226	100
<b>Home language</b>												
English	–	–	–	–	–	–	307	74.0	144	76.2	163	72.1
Not English	–	–	–	–	–	–	96	23.1	39	20.6	57	25.2
Missing	–	–	–	–	–	–	12	2.9	6	3.2	6	2.7
Total	–	–	–	–	–	–	415	100	189	100	226	100
<b>Highest degree</b>												
Associate's	1	1.1	0	0	1	2	–	–	–	–	–	–
Bachelor's	24	25.5	10	23.3	14	27.5	–	–	–	–	–	–
Master's	68	72.3	32	74.4	36	70.6	–	–	–	–	–	–
Doctoral	1	1.1	1	2.3	0	0	–	–	–	–	–	–
Total	94	100	43	100	51	100	–	–	–	–	–	–

students. To prevent cross-classification, we eliminated already-selected students from the next teacher's classroom roster. Of 470 students recruited, 412 (87.7%) agreed to participate. Table 1 summarizes student demographics. The majority of students self-identified as either White (29.6%), Multiracial (23.1%), Asian (14.5%), or Latinx (13.7%). There were relatively fewer Black students (6.3%) and students who identified as other (5.1%). In comparison, the general enrollment for the participating schools, based on public records, was 53.4% White, 15.8% Latinx, 15.8% Asian, 8.0% Multiracial, and 5.4% Black. Chi-square and *t*-tests examined baseline equivalency of the recruited sample across conditions. No significant differences were found on any demographic characteristics, except eligibility for FRL, which was entered as a covariate in all models.

## Procedure

All study procedures were approved by the University Institutional Review Board and school district research committees. Recruitment began in Spring 2019 via meetings with school and district leaders. Schools who agreed to participate provided a list of teachers whose teaching load was at least 50% ninth grade students. Teachers were oriented to the study procedures in person by research staff and those who agreed to participate provided written consent. In Fall of 2019, classroom rosters for participating teachers were obtained from school records. School administrators sent an information sheet describing the study to parents of ninth grade students with instructions to contact the researchers if they wish to opt out. From the students who remained eligible, a subset of five students per participating teacher

was selected from class rosters. Student informed assent occurred in person at each participating school.

E-EMR training occurred at the beginning of the 2019–2020 academic year, during protected professional development time prior to the return of students in August and early September, 2019. Separate trainings were conducted for each school. There were three waves of data collection for both students and teachers. Time 1 (T1) took place over a 2 week period from the end of September to beginning of October 2019. T2 occurred in late January 2020, prior to the end of first semester for all schools. For these first two waves of data collection, electronic surveys were sent to all teachers and students. Regular reminders were sent to teachers, and an in-person “make-up” data collection was conducted at each school for students. Between T2 and T3, the COVID-19 pandemic caused schools to move to distance learning. Leadership and teachers from all high schools expressed interest in continuing the study. However, E-EMR was developed and previously tested as an in-person intervention, making it difficult to measure fidelity of the intervention in distance settings. T3 data collection in May 2020 was completed virtually only, resulting in significantly lower response rates for students without the aid of an in-person make-up. Because the disruptions caused by COVID-19 compromised the validity of T3 data, the present study utilizes only the first two time points of data collected.

## Intervention Condition

### E-EMR Condition

EMR is a framework that provides shared language and concrete practices to promote STR. It also includes implementation supports such as training, a “playbook” with implementation scripts, structured professional learning community (PLC) protocols, and weekly email prompts. Teachers are trained on the three distinct, inter-related phases of a relationship: Establish, Maintain, and Restore. These phases are arranged sequentially as a heuristic that guides teachers’ practices, but they can fluctuate over time depending on changes in the relationship (e.g., teacher discipline of the student may present a need to restore the relationship).

### Establish practices

The initial phase in any relationship, and the first intervention component of EMR, involves intentional efforts to establish relationships with each student. The goal is to ensure that all students feel a sense of connection characterized by trust, respect, and mutual understanding. There are eight establish practices: banking time, getting to know the details of students’ lives, expressing high expectations and high support, proactively and reactively offering help,

providing opportunities for voice, interspersing opportunities to exercise choice, positive greetings and farewells, and relaying positive information through other adults.

### Maintain

Once established, active effort is required to maintain a relationship. Without proactive maintenance, relationship quality can deteriorate over time as the ratio of positive to negative interactions naturally diminishes (Steinberg, 2001). The primary practice associated with the Maintain Phase is the 5-to-1 ratio of positive to negative interactions. To achieve the 5-to-1 ratio, teachers learn strategies for non-contingent and contingent positive interactions. In addition, EMR provides guidance for preventing negative interactions by responding to unwanted behavior progressively and with empathy.

### Restore

Conflict within relationships is common. When left unattended, however, negative interactions can weaken the relationship, leaving the student less engaged in class, less responsive to efforts to correct problem behaviors, and more challenging to motivate to take on academic work they perceive as challenging or boring. The aim is for teachers to follow negative interactions with students by reconnecting and repairing relational rifts and returning the relationship to its previous (Maintain) phase. Teachers are trained on four Restore strategies: letting go, taking ownership, mutual problem-solving, and expressing care.

### Equity Levers

Four equity levers were added to make EMR equity-explicit, and were designed to provide teachers with concrete strategies for interrupting implicit biases while building relationships with students and increase teachers’ capacity to understand student behavior beyond their own cultural lens.

### Seek Commonalities

The goal of the Seek Commonalities equity lever is to minimize the perceived differences between teachers and their students, which can contribute to negative out-group biases and stereotyping. Teachers were encouraged to better understand students’ preferences, values, and experiences and identify which of these they may share with the student. Teachers are then encouraged to actively center the commonalities they share with students when establishing or restoring relationships.

### Gain Perspective

The goal of the Gain Perspective equity lever is to understand how a student perceives their experience in the classroom overall and how they may perceive negative interactions (e.g., disciplinary incidents) in particular. Teachers reflect on how their background shapes their expectations for student behavior and how students' cultural or personal backgrounds shape the ways in which they engage in the classroom and school. It is critical, however, for teachers not to assume students' perspectives based on their racial/ethnic or cultural identity, but to find ways to elicit this knowledge through discussions with students.

### Gather Facts to Disprove Assumptions

The goal of this equity lever is for teachers to seek objective and counter-stereotypical information about individual students. This requires teachers to (1) reflect on their assumptions of certain students and what behaviors they expect them to show, (2) determine what information needs to be gathered to disprove those assumptions, and (3) conduct objective observations to test their assumptions. Teachers can then engage in relationship building strategies with students having checked their assumptions regarding students and their behavior.

### Know your Vulnerabilities

The goal of the Know your Vulnerabilities equity lever is for teachers to recognize when they are most likely to respond in a biased manner and pay particular attention to interrupting those reactions in those moments. Teachers reflect on the situations and behaviors that are most likely to trigger an unskilled and/or biased interaction with a student and how their implicit or explicit biases may be most in play during those moments.

### Training

The initial training was a six-hour interactive group-based experience that utilized an elicit-provide-elicited structure to promote teacher engagement. Two trainers who were involved in the development of E-EMR conducted the trainings. The trainings included a rationale for the importance of STR, particularly for ninth grade students, the science behind implicit bias, and an introduction to principles of equitable implementation of universal programs. Teachers were given the opportunity to see and practice skills in each of the three relationship phases. Finally, teachers were oriented to the rationale and structure of ongoing implementation supports, including the PLC and embedded reflection form, the E-EMR playbook, and the reminder emails.

Trainings were videotaped and coded for fidelity. In terms of adherence, trainers delivered 97–100% of the active components as planned across all trainings. Teacher behavioral engagement was coded via 10 randomly selected 5 min video clips of the training. Two raters coded the videos using a 15 s partial interval recording and operational definitions of behavioral engagement. Raters overlapped on 30% of the observations; inter-rater agreement was 96%. Overall, behavioral engagement was 96% of the observed intervals.

### PLCs

A primary E-EMR implementation support is a structured PLC protocol, used monthly in small groups. The protocol guides teachers through six steps: (1) Review the previous month's plans and progress, (2) Using a roster, identify students who are in the Establish or Restore phases, (3) Reflect on equity by examining how student demographics are associated with relationship phase, (4) Set goals to establish or restore relationships with students, (5) Develop student-specific action plans to achieve the goals within the next month, and (6) Reflect on the PLC process, including what worked well and what could be improved.

### Weekly emails

To encourage and remind teachers to use E-EMR practices, weekly prompts were emailed to participating teachers. Emails aimed to increase motivation or featured specific practices that were particularly relevant based on timing in the school year. An "equity tip" was included with every email, which explicitly tied the practice to the experiences of students of color, provided implementation tips to enhance culturally responsiveness for students of color, or reminded teachers of relevant equity levers and how they may be applied to the practice.

### Control Condition

Ninth grade teachers in the control condition delivered practices as usual, with no formal training or technical assistance.

### Measures

#### Psychological Sense of School Membership (PSSM)

The PSSM is a validated student self-report measure (Goodenow, 1993) that assesses a young person's global sense of belonging and school connectedness. The PSSM consists of 18 items (e.g., I feel proud of belonging to my school) rated on a 5-point Likert scale (0 = *not at all true* to 4 = *completely true*). Internal reliability estimates and 4 week test-retest

reliability are adequate (Goodenow, 1993; Hagborg, 1994). The PSSM had good internal reliability in this sample ( $\alpha=0.89$ ).

### Academic Motivation Scale (AMS)

The AMS (Vallerand et al., 1992) is a 28-item questionnaire with seven subscales, each designed to assess different motivational states. This study included only five of the subscales deemed most relevant to STR: amotivation (i.e., unmotivated to engage in academic pursuits), external regulation (i.e., academic behaviors motivated to obtain a positive outcome or avoid a negative one), identified regulation (i.e., attributing personal value to academics), introjected regulation (i.e., behavior motivated to achieve or avoid favorable or unfavorable internal states), and (e) intrinsic regulation (i.e., a state of being intrinsically motivated by academics). Items were rated on a 7-point scale (1 = *does not correspond* to 7 = *correspond exactly*). The AMS has shown good reliability and criterion and concurrent validity (e.g., Cokley et al., 2001; Fairchild et al., 2005; Vallerand et al., 1993). Internal reliability for this sample ranged from  $\alpha=0.81$ –0.94 for all subscales.

### Student Engagement Instrument (SEI)

The SEI includes 33 items measuring cognitive and affective engagement (Appleton et al., 2006). Although the SEI has five subscales, only three were used in this study: (a) teacher-student relationships (9 items), control and relevance of school work (9 items), and future aspirations and goals (5 items). Items are rated on a 4-point scale (1 = *strongly agree* to 4 = *strongly disagree*). These three subscales have shown evidence of construct (Betts et al., 2010) and predictive validity (Appleton et al., 2006). Internal reliability for this sample ranged from  $\alpha=0.79$  to 0.88 for all subscales.

### Strengths and Difficulties Questionnaire (SDQ)

The SDQ is a standardized behavior rating scale for 3- to 16-year-olds that assesses psychosocial functioning and consists of four problem behavior subscales and one prosocial behavior subscale. Problem behavior subscales can be aggregated to yield a “total problems” score. Respondents complete items on a 3-point scale (0 = *not true* to 2 = *certainly true*). The SDQ has demonstrated acceptable internal consistency, reliability, and validity (Goodman, 2001) and has been shown to be at least as good as the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) at detecting conduct and emotional problems (Goodman & Scott, 1999). Internal reliability in this sample varied by subscale:  $\alpha=0.73$ –0.74 for emotional problems,  $\alpha=0.64$ –0.69 for conduct problems,  $\alpha=0.71$ –0.75 for

hyperactivity/inattention problems,  $\alpha=0.57$  for peer relationship problems,  $\alpha=0.62$ –0.65 for prosocial behavior, and  $\alpha=0.82$ –0.83 for total problems.

### Analytic Plan

We conducted tests of baseline equivalence, examining differences in student and teacher demographics and baseline levels of outcome variables, across the treatment and control groups. The two groups were not significantly different on any demographic or outcome variables, except that the control group had a significantly higher proportion of students eligible for FRL. Thus, FRL status was entered as a covariate in all inferential models.

We calculated intraclass correlation coefficients (ICCs) and design effects (Huang, 2018; Musca et al., 2011; Raudenbush & Bryk, 2002). ICCs ranged from 0 to 48% (average = 8%) at Time 1 and 0–15% (average = 3%) at Time 2; design effects ranged from 1.02–2.86 (average = 1.32) at Time 1 and 1.03–2.25 (average = 1.13) at Time 2. Although the design effects suggest clustering may be ignorable, some of our ICCs exceed 0.05, suggesting a need for clustering. We opted for a more conservative approach and conducted our inferential analyses accounting for nesting.

Three sets of models were run to test our research questions. First, we examined the main effects of E-EMR condition (1 = E-EMR, 0 = Control) on each outcome at T2, controlling for baseline (T1) levels of the outcome and FRL status. Baseline scores were centered; thus, the effect of E-EMR could be interpreted for a student with an average level of the variable at T1. Effect sizes were calculated using a delta score which accounts for relative change over time for both the intervention and control group  $d_{\text{change}} = (M_{\text{change-t}}/SD_{\text{pre-t}}) - (M_{\text{change-c}}/SD_{\text{pre-c}})$ , as recommended by Feingold (2009). Second, we examined whether the effect of E-EMR on T2 outcomes were moderated by T1 scores by including a multiplied term between the centered T1 variable and the E-EMR condition variable. Significant ( $p < 0.05$ ) interaction terms were probed using the Johnson-Neyman technique, where the effect of E-EMR on the T2 outcome was plotted according to continuous levels of the adjusted T1 score. This allows for the identification of “regions of significance”—specific values of T1 where the effect of E-EMR was significantly different from zero. Third, we examined whether student's race/ethnicity moderated the effects of E-EMR on T2 outcomes. Dummy variables representing each racial/ethnic category were entered; White was used as a reference. Interaction terms between the five categories (i.e., Asian, Latinx, Black, Multiracial, Other) and E-EMR condition were added to the model. Although the Other race group was included as a dummy code, we did not interpret interactions or simple slopes for this group

because of very small cell sizes (intervention  $n = 5$ , control  $n = 16$ ). Otherwise, any significant interactions were probed using the model constraint command in Mplus to generate simple slopes. For any probed interaction, an additional model was run where the second-largest racial category (i.e., Multiracial) was removed to acquire a simple slope for White students.

All models were run using Mplus 8.1 (Muthén & Muthén, 2017). Missing data ranged from 6.7 to 10.1% at T1 and from 9.8 to 13.0% at T2. Missing analyses revealed that multiple student characteristics, including race, gender, whether a student had failed a class the previous year, and GPA, predicted missingness; thus, all missing data was imputed using demographic characteristics as auxiliary variables. Analyses were run using TYPE = COMPLEX at the classroom level to address clustering of students within teachers.

## Results

### Main Effects of E-EMR

Table 2 summarizes the main effects of E-EMR and average scores at pre- and post-intervention for students in E-EMR and control conditions separately. As shown, none of the main effects reached statistical significance, although changes were in the expected direction.

### Moderation by Baseline Scores

There was a two-way interaction between condition and baseline scores for three variables: STR ( $B = -0.16(0.08)$ ,  $p = 0.05$ ), conduct problems ( $B = -0.24(0.10)$ ,  $p = 0.02$ ), and total problems ( $B = -0.15(0.07)$ ,  $p = 0.047$ ). Johnson-Neyman plots probing these interactions are shown in Fig. 1. As illustrated in Panel A, the effect of E-EMR was significant

**Table 2** E-EMR intervention effects on T2 outcomes and pre- and post-test means for each condition

	Baseline	E-EMR <sup>a</sup> (1 = Intervention)		FRL (1 = Eligible)		E-EMR condition		Control condition	
Outcome variable	<i>B</i> ( <i>SE</i> )	<i>B</i> ( <i>SE</i> )	<i>delta</i>	<i>B</i> ( <i>SE</i> )		Pre-test <i>M</i> ( <i>SD</i> )	Post-test <i>M</i> ( <i>SD</i> )	Pre-test <i>M</i> ( <i>SD</i> )	Post-test <i>M</i> ( <i>SD</i> )
School belonging	0.70 (.04) <sup>***</sup>	0.09 (.06)	.18	0.02 (.06)		3.71 (.66)	3.76 (.66)	3.74 (.64)	3.67 (.65)
Amotivation	0.55 (.06) <sup>***</sup>	-0.11 (.11)	-.14	0.08 (.17)		2.06 (1.32)	2.16 (1.34)	1.88 (1.22)	2.15 (1.37)
External regulation	0.46 (.06) <sup>***</sup>	-0.02 (.14)	.10	-0.06 (.16)		5.31 (1.43)	5.35 (1.44)	5.58 (1.36)	5.48 (1.54)
Identified regulation	0.49 (.07) <sup>***</sup>	-0.04 (.13)	-.05	0.02 (.16)		5.69 (1.12)	5.48 (1.43)	5.79 (1.34)	5.61 (1.35)
Introjected regulation	0.54 (.04) <sup>***</sup>	-0.05 (.14)	.07	0.17 (.17)		4.65 (1.68)	4.43 (1.67)	5.02 (1.62)	4.69 (1.66)
Intrinsic regulation	0.56 (.05) <sup>***</sup>	-0.07 (.14)	-.07	0.19 (.20)		4.85 (1.53)	4.58 (1.64)	4.86 (1.61)	4.69 (1.75)
Student-teacher relationships	0.68 (.05) <sup>***</sup>	0.04 (.04)	.05	0.05 (.06)		2.11 (.44)	2.09 (.44)	2.07 (.44)	2.03 (.51)
Control of schoolwork	0.76 (.04) <sup>***</sup>	0.04 (.04)	.14	0.01 (.06)		1.93 (.49)	1.87 (.54)	1.97 (.49)	1.85 (.56)
Future aspirations & goals	0.67 (.05) <sup>***</sup>	0.05 (.04)	.15	-0.08 (.06)		2.51 (.44)	2.47 (.48)	2.50 (.48)	2.39 (.54)
Emotional problems	0.70 (.04) <sup>***</sup>	0.04 (.04)	-.02	0.07 (.05)		.77 (.51)	.83 (.51)	.70 (.50)	.77 (.50)
Conduct problems	0.65 (.06) <sup>***</sup>	-0.05 (.03)	-.20	0.04 (.04)		.34 (.35)	.29 (.33)	.32 (.35)	.34 (.39)
Hyperactivity/inattention	0.70 (.03) <sup>***</sup>	-0.03 (.03)	-.11	0.00 (.04)		.89 (.50)	.91 (.47)	.85 (.47)	.92 (.44)
Peer problems	0.62 (.05) <sup>***</sup>	0.01 (.03)	.03	-0.02 (.04)		.39 (.33)	.41 (.35)	.42 (.34)	.43 (.36)
Prosocial behavior	0.56 (.05) <sup>***</sup>	-0.05 (.03)	-.16	-0.02 (.04)		1.61 (.35)	1.56 (.38)	1.59 (.33)	1.60 (.33)
Total problems	0.74 (.04) <sup>***</sup>	-0.01 (.02)	-.10	0.02 (.03)		.60 (.32)	.61 (.31)	.57 (.30)	.61 (.30)

FRL free & reduced lunch

<sup>a</sup>Effects of E-EMR should be interpreted for students with average levels of baseline scores who are not eligible for free and reduced lunch

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$



CONSORT diagram detailing study procedures

RANDOMIZATION

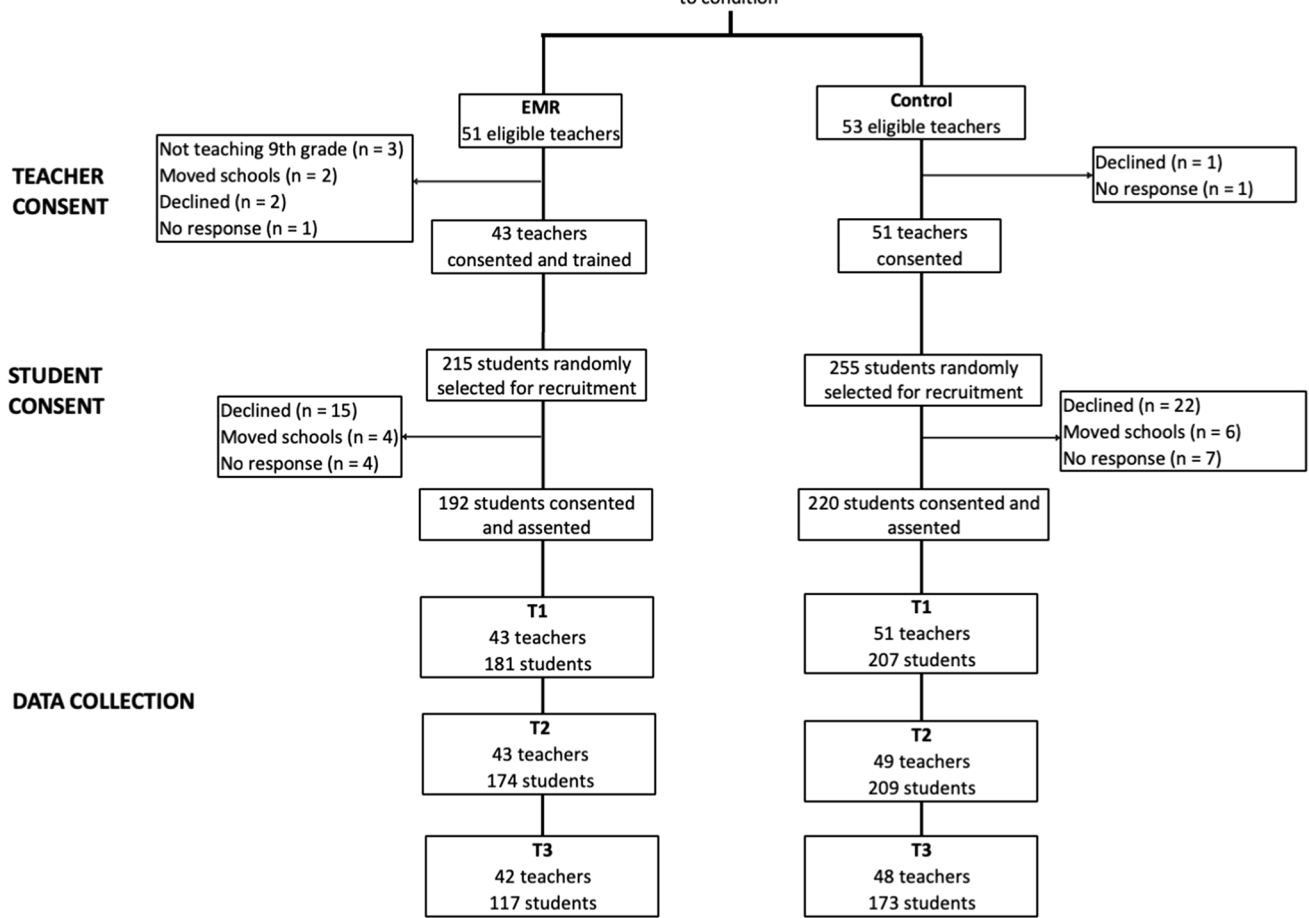


Fig. 1 CONSORT diagram detailing study procedures

and in the expected direction for students whose STR scores were  $-0.53 SD$  or below the mean at baseline, but not students who scored above this cutoff at T1. Panel B shows a similar plot for conduct problems. As seen, a significant expected effect of E-EMR was found for students who scored  $0.1 SD$  above the mean at baseline. However, a significant effect of E-EMR in the unexpected direction was found for students who scored  $1.6 SD$  below the mean at baseline (i.e., students who began with fewer conduct problems than their peers). Finally, for total problems, the plot in Panel C indicates a significant effect of E-EMR in the expected direction for students at  $0.95 SD$  above the mean at baseline.

**Moderation by Student Race/Ethnicity**

Significant interactions by race were identified for nine of the E-EMR outcomes. Interaction coefficients are summarized in Table 3 and can be interpreted as the difference between the effect of E-EMR for White students and

for students from Asian, Latinx, Black, and Multiracial students, respectively. Simple slopes can be interpreted as the difference between intervention and control students within each racial group.

**Asian Students**

Asian race significantly interacted with E-EMR condition for school belonging ( $B = 0.37(0.13)$ ,  $p = 0.01$ ) and hyperactivity/inattention ( $B = 0.23(0.10)$ ,  $p = 0.02$ ). Simple slopes revealed that E-EMR had a significant positive effect on T2 school belonging for Asian ( $B = 0.28(0.11)$ ,  $p < 0.01$ ) but not White students ( $B = -0.10(0.09)$ ,  $p = 0.28$ ). Simple slopes for hyperactivity/inattention revealed a significant effect in the unexpected direction for Asian students ( $B = 0.16(0.07)$ ,  $p = 0.03$ ) and a non-significant effect for White students ( $B = -0.07(0.06)$ ,  $p = 0.26$ ).

**Table 3** Interactions between baseline scores and condition and interactions between race/ethnicity and condition

Outcome variable	Baseline moderation	Race moderation			
	Baseline x condition	Asian x condition	Black x condition	Latinx x condition	Multiracial x condition
	<i>B</i> ( <i>SE</i> )	<i>B</i> ( <i>SE</i> )	<i>B</i> ( <i>SE</i> )	<i>B</i> ( <i>SE</i> )	<i>B</i> ( <i>SE</i> )
School belonging	− 0.01 (.08)	0.37 (.13)*	0.26 (.18)	0.27 (.16)	0.27 (.11)*
Amotivation	− 0.05 (.12)	0.00 (.30)	0.46 (.60)	− 0.18 (.35)	0.09 (.32)
External regulation	0.00 (.12)	− 0.61 (.37)	− 0.31 (.66)	− 0.31 (.40)	− 0.03 (.39)
Identified regulation	0.06 (.13)	− 0.04 (.34)	0.27 (.49)	0.82 (.38)*	0.57 (.35)
Introjected regulation	0.03 (.09)	− 0.65 (.41)	0.06 (.69)	0.12 (.41)	0.50 (.37)
Intrinsic regulation	− 0.03 (.09)	− 0.08 (.40)	0.48 (.66)	0.79 (.43)	0.74 (.40)
Student–teacher relationships	− 0.16 (.08)	0.08 (.10)	0.23 (.16)	0.20 (.12)	0.06 (.09)
Control & relevance of schoolwork	− 0.06 (.08)	0.01 (.12)	0.55 (.19)**	0.23 (.13)	0.09 (.11)
Future aspirations	− 0.15 (.10)	0.21 (.11)	0.28 (.18)	0.34 (.13)*	0.23 (.11)*
Emotional problems	− 0.07 (.07)	0.12 (.12)	− 0.05 (.15)	− 0.04 (.10)	0.06 (.10)
Conduct problems	− 0.24 (.10)*	0.06 (.09)	0.04 (.16)	− 0.09 (.09)	− 0.07 (.08)
Hyperactivity/inattention	− 0.01 (.06)	0.23 (.10)*	− 0.11 (.14)	− 0.06 (.10)	0.07 (.08)
Peer problems	− 0.12 (.09)	0.04 (.09)	− 0.04 (.14)	− 0.17 (.09)	− 0.08 (.09)
Prosocial behavior	0.06 (.10)	0.16 (.09)	0.01 (.14)	0.04 (.09)	0.15 (.09)
Total problems	− 0.15 (.07)*	0.11 (.07)	− 0.05 (.10)	− 0.09 (.06)	− 0.01 (.06)

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ 

### Latinx Students

A significant interaction between Latinx race and E-EMR condition was found for identified regulation ( $B = 0.82(0.38)$ ,  $p = 0.03$ ) and future aspirations and goals ( $B = 0.34(0.13)$ ,  $p = 0.01$ ). For identified regulation, E-EMR showed effects in opposing directions for Latinx students ( $B = 0.52(0.28)$ ,  $p = 0.06$ ) and White students ( $B = -0.28(0.24)$ ,  $p = 0.23$ ), although neither effects reached statistical significance. Similarly, simple slopes revealed that the effect of E-EMR on future aspirations and goals was significant for Latinx ( $B = 0.24(0.12)$ ,  $p = 0.04$ ), but not White students ( $B = -0.12(0.08)$ ,  $p = 0.13$ ).

### Black Students

A significant interaction of E-EMR was found for Black students relative to White students on only one outcome, control and relevance of schoolwork ( $B = 0.55(0.19)$ ,  $p = 0.003$ ). Simple slopes revealed a positive effect of E-EMR for Black students ( $B = 0.51(0.16)$ ,  $p = 0.002$ ) that was not significant for White students ( $B = -0.07(0.08)$ ,  $p = 0.18$ ).

### Multiracial Students

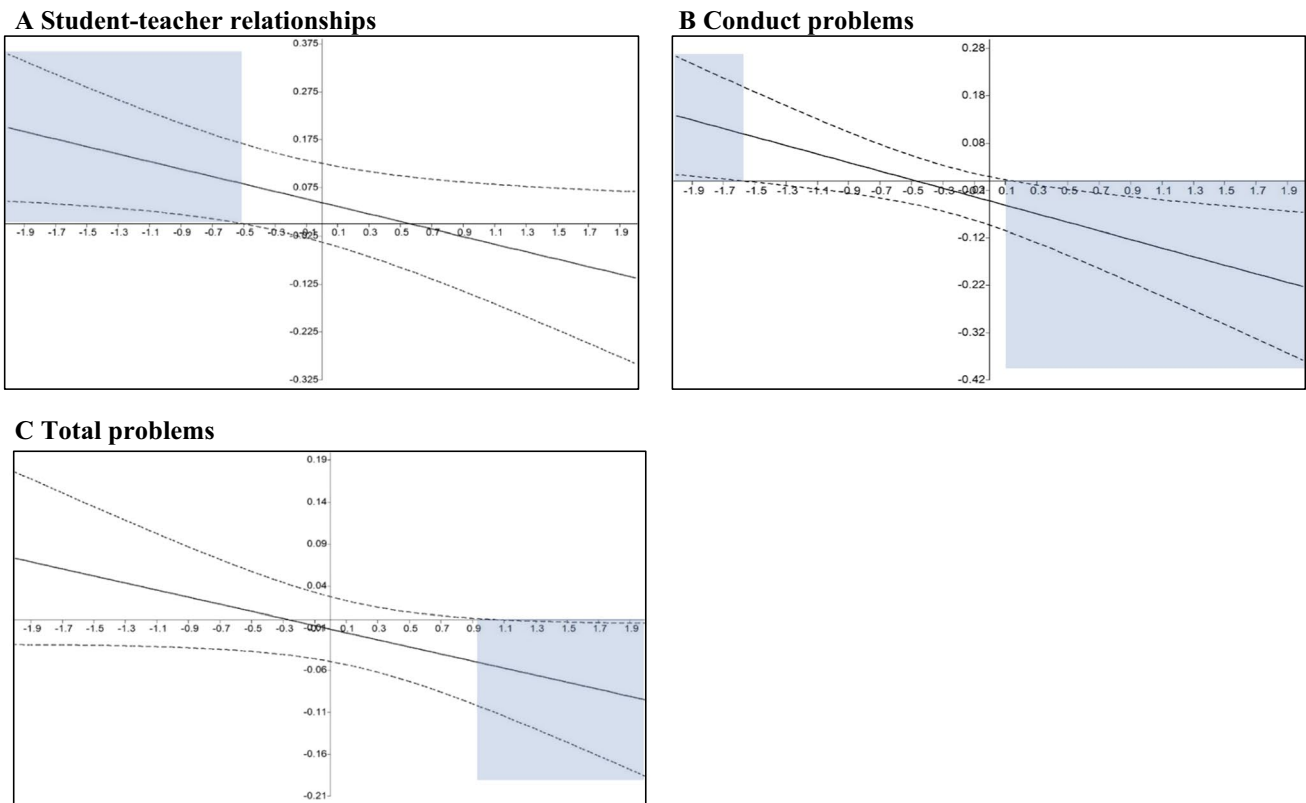
Significant interactions emerged for school belonging ( $B = 0.27(0.11)$ ,  $p = 0.02$ ) and future aspirations and goals ( $B = 0.23(0.11)$ ,  $p = 0.04$ ). E-EMR had a significant positive effect on school belonging for Multiracial ( $B = 0.18(0.09)$ ,

$p = 0.04$ ), but not White students ( $B = -0.10(0.09)$ ,  $p = 0.28$ ). Similarly, simple slopes for future aspirations and goals showed changes in opposing directions for Multiracial ( $B = 0.14(0.08)$ ,  $p = 0.09$ ) and White students ( $B = -0.12(0.08)$ ,  $p = 0.13$ ), although neither reached statistical significance.

### Discussion

This study built on previous research on Establish-Maintain-Restore with elementary and middle school students by evaluating its effectiveness among high school students and teachers in a cluster randomized controlled trial. The version of EMR tested in this study had been iteratively adapted to optimize its developmental and contextual appropriateness for high school and to make it equity-explicit (reference removed for blinding). Overall, we found no significant main effects, which is common in evaluations of universal prevention programs (Greenberg & Abenavoli, 2017). However, there were targeted benefits for students who started with low scores at baseline, for Asian, Latinx, Multiracial, and (to a lesser extent) Black students. We also found some unexpected effects, where high-performing and/or advantaged groups in the E-EMR condition had less favorable outcomes at post, compared to those in the control group, which may be a result of the equity-explicit focus of E-EMR (Fig. 2).

Two principles guided our interpretation of the findings. First, this study was a pilot trial. Pilot trials play a critical role in the translational research process because they



**Fig. 2** Johnson-Neyman Plots outlining the change in slope between E-EMR and T2 outcomes (y-axes) at varying levels of standardized baseline scores (x-axes) of each outcome. Solid lines represent the effect of E-EMR on T2 outcomes, whereas dashed lines represent the

confidence interval around that effect. The effect of E-EMR is significant when the confidence interval does not include zero along the y-axis. Blue shading indicates regions of significance

are designed to assess feasibility for and detect potential signals that could inform planning of a larger, more rigorous, and adequately powered study (Eldridge et al., 2016). Second, E-EMR is a universal intervention. In a review of 74 meta-analyses of universal intervention programs for school-aged youth, Tanner-Smith et al. (2018) found that the median average effect sizes of such programs tended to fall between Cohen’s  $d=0.07$  and Cohen’s  $d=0.16$ . Those authors concluded that the statistical benchmarks outlined in by Cohen (1988) for small (0.20), medium (0.50), and large (0.80) effects are inappropriate to use in interpreting findings from universal interventions.

**Main Effects of E-EMR**

We did not find significant main effects of E-EMR, although changes were in the hypothesized direction. The practices in E-EMR are mostly dyadic. For instance, one of the primary Establish practices is to “bank time” with individual students by engaging in student-centered conversation. This may reflect E-EMR’s origins in elementary schools, and the field’s focus on STR among younger students (Kincade et al., 2020). The dyadic nature of such

practices render them very difficult for high school teachers to implement with all their students, given the limited resources and time that results from teaching many more students across multiple class periods (Deschenes et al., 2010). In fact, we acknowledged the limitations on resources and time in E-EMR training. A fundamental part of E-EMR implementation is a “triage” process that helps teachers focus their attention on the students who need it most. Therefore, E-EMR practices, particularly at the high school level, are not intended to nor can feasibly be distributed equally across all students. It is also possible that we did not detect more effects because the five students randomly selected to be part of the study procedures did not match the students that teachers were focused on. Future research should examine the fidelity of E-EMR implementation at the dyadic level and how dyadic fidelity corresponds to benefits in student outcomes. Further, dyadic interactions between teachers and students in high schools have a strong academic focus. It may be worthwhile to examine the effects of E-EMR in concert with a whole-class intervention such as My Teaching Partner-Secondary (Liu et al., 2018).

## Moderation by Baseline Scores

We found significant moderation by baseline scores for STRs, conduct problems, and total problems. For all three variables, E-EMR had the strongest effect among those who scored worse at baseline. These findings are consistent with other universal interventions, which often show increased benefits for students who are in the most need (Greenberg & Abenavoli, 2017). The findings are particularly encouraging, given that STR were reported by students, who were not aware of intervention condition. Further, conduct problems and total problems represent more distal outcomes of E-EMR, and are not directly targeted by the intervention. The impact of E-EMR on these outcomes are consistent with previous research suggesting that positive STR are protective against a variety of independently observed problem behaviors (Cook, Coco, et al., 2018b).

## Moderation by Student Race/Ethnicity

With White students as the reference group, we found some targeted benefits for Asian, Latinx, Black, and Multiracial students. Among Asian students, for example, those whose teachers participated in E-EMR reported significantly larger increases in school belonging from September to January, compared to the control group. Although Asian students are not disadvantaged relative to their white peers on many educational outcomes, including achievement (Hsin & Xie, 2014), they are often subject to experiences of racial discrimination at school that undermine their sense of belonging (Byrd & Andrews, 2016; Montoro et al., 2021). These psychological outcomes are often overlooked for Asian students (Wing, 2007). Seen as the “model minority,” they are often perceived as high-performing and needing minimal support (Whaley & Noel, 2013; Yoo et al., 2015). However, a strong sense of school membership is protective across a range of psychosocial outcomes (Wagle et al., 2018).

Among Latinx students, E-EMR led to significant benefits on students’ future aspirations and goals, and trends in improved STR and perceived control and relevance of schoolwork. These patterns suggest E-EMR may have promise for mitigating long-standing barriers for Latinx students. Prior work has found that teachers have diminished expectations for Latinx students, are less likely to praise them, even for correct answers, and have fewer positive interactions with Latinx students (Tenenbaum & Ruck, 2007). As a consequence of these systemic barriers, Latinx students are less likely to feel connected to school, score lower than their White peers on standardized achievement tests, and drop out of high school at greater rates (Vera et al., 2018). It is important for future research to examine whether the positive trends noted in this study also manifest in indicators of achievement.

We found relatively fewer positive effects of E-EMR among Black high school students. Indeed, the only significant finding that emerged was that Black students in the intervention condition perceived greater control over and relevance of their schoolwork. It may be that anti-Black racism needs to be more explicitly addressed in interventions such as E-EMR (Wun, 2016). Additionally, our small subgroup sample size likely undermined our ability to detect effects. Future research with a larger sample size will be better positioned to tease out competing hypotheses.

Among Multiracial students, we found a relative benefit of E-EMR on sense of school belonging. These findings should be interpreted cautiously, given the heterogeneity of the “multiracial” category. Despite the growing multiracial population in the USA (Mordechay & Orfield, 2017), the field lacks consensus around how to measure and categorize multiracial individuals (Charmaraman et al., 2014). Consequently, inconsistent findings have been documented in the literature (Shih & Sanchez, 2005; Udry et al., 2003). Previous studies have adopted a range of approaches, including not allowing participants to report more than one race, not reporting data from multiracial participants, and combining all mixed subgroups into a single “multiracial” category (Charmaraman et al., 2014). In the current study, we adopted the latter approach. The 96 students within our multiracial category represented 30 unique racial combinations, with the largest groups identifying as Latinx and White ( $n=26$ ), Black and White ( $n=14$ ), Asian and White ( $n=11$ ), Native American and White ( $n=5$ ), and Black and Latinx ( $n=4$ ). Some researchers have hypothesized that mixed-race children may occupy positions of power and privilege that are somewhere between those of monoracial white and monoracial ethnic minority peers (Binning et al., 2009). Others have noted that mixed-race youth face unique challenges and opportunities, such as conflictual cultural messages at home and access to multiple communities (Shih & Sanchez, 2005). In terms of interventions like E-EMR, which are focused on teacher behavior, a student’s phenotype (which influences racial identification by others) may be important to consider in addition to self-identification (AhnAllen et al., 2006).

## Unexpected Findings

We found some unexpected changes attributable to E-EMR. For instance, among students who scored low on conduct problems relative to their peers at baseline, students in E-EMR, compared to those in the control group, showed an *increase* in conduct problems. It should be noted, however, that even with this apparent increase, these students’ scores at post were 0.12 *SD* above the mean, well within the normative range on this measure. Additionally, E-EMR may have equalized outcomes by influencing subgroup scores in unintended directions. For example, Asian students

increased in hyperactivity/inattention, and White students decreased in prosocial behavior. In both of these instances, students' scores at post still fell well within the normal range ( $z = 0.96$  for hyperactivity/inattention and  $z = 1.48$  for prosocial behavior, respectively).

These unintended effects are small and limited and we are cautious about over-interpreting the findings, particularly in the context of a pilot trial. It should also be noted that these changes were observed over a short intervention period of four months, and we do not know whether these changes will continue and for how long. Nonetheless, the findings raise important questions about desired and acceptable outcomes of doing equity work. Some might expect equity to entail improving outcomes for “disadvantaged” subgroups while maintaining or improving outcomes for advantaged groups. However, this may not be feasible or enable disparities to be dismantled. If all tides rise in response to universal supports, gaps will continue to exist between advantaged and disadvantaged groups. As a field and society, is it acceptable if equity work equalizes outcomes not only by bringing up disadvantaged groups but also by bringing down advantaged groups? How much of a loss to advantaged subgroups are we willing to tolerate in the context of gains for disadvantaged groups? We must grapple with such questions.

## Limitations

This study has several limitations not already discussed above. First, we had only a small number of schools. This, combined with randomization at the school level, resulted in some imbalance in our sample at baseline that had to be adjusted for statistically. It is also important to note that although randomization occurred at the school level, our analyses were conducted at the student level, impacting our degrees of freedom (Raudenbush & Bryk, 2002). If randomization occurs at the school level, a true test of treatment effects would require measurement of all teachers and all students within the school (Raudenbush & Bryk, 2002). As a pilot trial and due to the pandemic, we had only a short follow-up period of four months, and were not adequately powered to test mechanisms (e.g., implementation fidelity as a mediator of outcomes). Relatedly, we did not assess the extent to which teachers in the control condition implemented practices similar to E-EMR. Finally, future research needs to examine E-EMR's impact on behavioral and academic outcomes such as attendance, discipline, and grades.

## Implications and Conclusions

In this study, we extended the research on EMR by adapting it to be equity-explicit and developmentally appropriate for high school students and testing it to detect signals of evidence to inform future research. Together, findings from

our previous studies demonstrate the promise of E-EMR for improving teachers' skills for building relationships and for enhancing equity in student outcomes. E-EMR shares a common component with other promising equity-focused programs, which is the opportunity for teachers to learn about, reflect on, and create action plans to disrupt their biases (see Bradshaw et al., 2018). Importantly, the equity-explicit components were not siloed from other intervention components, but was interwoven throughout each EMR practice. That is, we made it clear that teachers can establish, maintain, and restore relationships in a color-blind manner, or that they can establish, maintain, and restore relationships with equity front and center, and we gave specific instructions for doing the latter. We believe these concrete recommendations were pivotal in the effectiveness of E-EMR. Relatedly, the equity lens was interwoven throughout the monthly professional learning communities. This monthly cycle of goal setting, action, and reflection was designed to promote behavioral habits (taking equity-explicit actions to cultivate relationships) as well as dispositions (reflecting on one's teaching through an equity lens). At this stage, E-EMR is poised for a large-scale efficacy trial that establishes more robust evidence of the effects of E-EMR on student social, emotional, and behavioral outcomes, including the prevention of high school dropout.

**Funding** This research was supported by a grant from the Institute of Education Sciences (R305A170458).

## Declarations

**Conflict of interest** The last author receives compensation for providing training on EMR.

**Ethical Approval** All procedures were approved by University of Washington IRB (Study No. 02720).

**Informed Consent** Written, informed consent was obtained from adult participants. Participants under 18 years of age provided written, informed assent.

## References

- Achenbach, T. M., & Rescorla, L. (2001). Manual for the ASEBA school-age forms & profiles. University of Vermont, Research Center for Children, Youth, & Families.
- AhnAllen, J. M., Suyemoto, K. L., & Carter, A. S. (2006). Relationship between physical appearance, sense of belonging and exclusion, and racial/ethnic self-identification among multiracial Japanese European Americans. *Cultural Diversity and Ethnic Minority Psychology, 12*(4), 673–686. <https://doi.org/10.1037/1099-9809.12.4.673>
- Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of

- the student engagement instrument. *Journal of School Psychology*, 44(5), 427–445. <https://doi.org/10.1016/j.jsp.2006.04.002>
- Archambault, I., Janosz, M., Fallu, J.-S., & Pagani, L. S. (2009). Student engagement and its relationship with early high school dropout. *Journal of Adolescence*, 32(3), 651–670. <https://doi.org/10.1016/j.adolescence.2008.06.007>
- Barber, B., & Olsen, J. A. (2004). Assessing the transitions to middle and high school. *Journal of Adolescent Research*, 19(1), 3–30. <https://doi.org/10.1177/0743558403258113>
- Binning, K. R., Unzueta, M. M., Huo, Y. J., & Molina, L. E. (2009). The interpretation of multiracial status and its relation to social engagement and psychological well-being. *Journal of Social Issues*, 65(1), 35–49. <https://doi.org/10.1111/j.1540-4560.2008.01586.x>
- Bradshaw, C. P., Pas, E. T., Bottiani, J. H., Debnam, K. J., Reinke, W. M., Herman, K. C., & Rosenberg, M. S. (2018). Promoting cultural responsiveness and student engagement through double check coaching of classroom teachers: An efficacy study. *School Psychology Review*, 47(2), 118–134.
- Brinkworth, M. E., McIntyre, J., Juraschek, A. D., & Gehlbach, H. (2018). Teacher-student relationships: The positives and negatives of assessing both perspectives. *Journal of Applied Developmental Psychology*, 55, 24–38. <https://doi.org/10.1016/j.appdev.2017.09.002>
- Byrd, C. M., & Andrews, D. J. (2016). Variations in students' perceived reasons for, sources of, and forms of in-school discrimination: A latent class analysis. *Journal of School Psychology*, 57, 1–14. <https://doi.org/10.1016/j.jsp.2016.05.001>
- Charmaraman, L., Woo, M., Quach, A., & Erkut, S. (2014). How have researchers studied multiracial populations? A content and methodological review of 20 years of research. *Cultural Diversity & Ethnic Minority Psychology*, 20(3), 336–352. <https://doi.org/10.1037/a0035437>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). USA: Lawrence Earlbaum Associates.
- Cokley, K. O., Bernard, N., Cunningham, D., & Motoike, J. (2001). A psychometric investigation of the academic motivation scale using a United States sample. *Measurement and Evaluation in Counseling and Development*, 34(2), 109–119. <https://doi.org/10.1080/07481756.2001.12069027>
- Cook, C. R., Duong, M. T., McIntosh, K., Fiat, A. E., Larson, M., Pullmann, M. D., & McGinnis, J. (2018). Addressing discipline disparities for black male students: Linking malleable root causes to feasible and effective practices. *School Psychology Review*, 47(2), 135–152.
- Cook, C. R., Coco, S., Zhang, Y., Fiat, A. E., Duong, M. T., Renshaw, T. L., Long, A. C., & Frank, S. (2018). Cultivating positive teacher-student relationships: Preliminary evaluation of the establish-maintain-restore (EMR) method. *School Psychology Review*, 47(3), 226–243.
- Deschenes, S. N., Arbreton, A., Little, P. M., Herrera, C., Grossman, J. B., & Weiss, H. B. (2010). *Engaging Older Youth: Program and City-Level Strategies to Support Sustained Participation in Out-of-School Time*. US: Harvard Family Research Project.
- Diazgranados, S., Selman, R. L., & Dionne, M. (2016). Acts of social perspective taking: A functional construct and the validation of a performance measure for early adolescents. *Social Development*, 25(3), 572–601. <https://doi.org/10.1111/sode.12157>
- Duong, M. T., Pullmann, M. D., Buntain-Ricklefs, J., Lee, K., Benjamin, K. S., Nguyen, L., & Cook, C. R. (2019). Brief teacher training improves student behavior and student-teacher relationships in middle school. *School Psychology*, 34(2), 212–221.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432.
- Eldridge, S. M., Lancaster, G. A., Campbell, M. J., Thabane, L., Hopewell, S., Coleman, C. L., & Bond, C. M. (2016). Defining feasibility and pilot studies in preparation for randomised controlled trials: Development of a conceptual framework. *PLoS ONE*, 11(3), e0150205. <https://doi.org/10.1371/journal.pone.0150205>
- EPE Research Center. (2006). *Diplomas count: An essential guide to graduation rates and policies*. (41 No. 25; Special Issues, Education Week, pp. 4–17). EPE Research Center. Retrieved from <http://www.edweek.org/ew/toc/2006/06/22/index.html>
- Fairchild, A. J., Horst, S. J., Finney, S. J., & Barron, K. E. (2005). Evaluating existing and new validity evidence for the academic motivation scale. *Contemporary Educational Psychology*, 30(3), 331–358. <https://doi.org/10.1016/j.cedpsych.2004.11.001>
- Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, 14(1), 43–53. <https://doi.org/10.1037/a0014699>
- Gaias, L. M., Cook, C. R., Nguyen, L., Brewer, S. K., Brown, E. C., Kiche, S., Shi, J., Buntain-Ricklefs, J., & Duong, M. T. (2020). A mixed methods pilot study of an equity-explicit student-teacher relationship intervention for the ninth-grade transition. *Journal of School Health*, 90(12), 1004–1018. <https://doi.org/10.1111/josh.12968>
- Gay, G. (2010). Acting on beliefs in teacher education for cultural diversity. *Journal of Teacher Education*, 61, 143–152. <https://doi.org/10.1177/0022487109347320>
- Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools*, 30(1), 79–90. [https://doi.org/10.1002/1520-6807\(199301\)30:1%3c79::AID-PITS2310300113%3e3.0.CO;2-X](https://doi.org/10.1002/1520-6807(199301)30:1%3c79::AID-PITS2310300113%3e3.0.CO;2-X)
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337–1345. <https://doi.org/10.1097/00004583-200111000-00015>
- Goodman, R., & Scott, S. (1999). Comparing the strengths and difficulties questionnaire and the child behavior checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, 27(1), 17–24. <https://doi.org/10.1023/A:1022658222914>
- Greenberg, M. T., & Abenavoli, R. (2017). Universal interventions: Fully exploring their impacts and potential to produce population-level impacts. *Journal of Research on Educational Effectiveness*, 10(1), 40–67. <https://doi.org/10.1080/19345747.2016.1246632>
- Gregory, A., Hafen, C. A., Ruzek, E., Mikami, A. Y., Allen, J. P., & Pianta, R. C. (2016). Closing the racial discipline gap in classrooms by changing teacher practice. *School Psychology Review*, 45(2), 171–191.
- Gruman, D. H., Harachi, T. W., Abbott, R. D., Catalano, R. F., & Fleming, C. B. (2008). Longitudinal effects of student mobility on three dimensions of elementary school engagement. *Child Development*, 79(6), 1833–1852. <https://doi.org/10.1111/j.1467-8624.2008.01229.x>
- Hagborg, W. J. (1994). An exploration of school membership among middle- and high-school students. *Journal of Psychoeducational Assessment*, 12(4), 312–323. <https://doi.org/10.1177/073428299401200401>
- Hertzog, C. J., & Morgan, P. L. (1999). Making the transition from middle level to high school. *High School Magazine*, 6(4), 26–30.
- Howard, G. R. (2016). *We can't teach what we don't know: White teachers, multiracial schools (Third)*. US: Teachers College Press.
- Hsin, A., & Xie, Y. (2014). Explaining Asian Americans' academic advantage over whites. *Proceedings of the National Academy of Sciences*, 111(23), 8416–8421. <https://doi.org/10.1073/pnas.1406402111>

- Huang, F. L. (2018). Multilevel modeling myths. *School Psychology Quarterly*, 33(3), 492.
- Hughes, J. N., Gleason, K. A., & Zhang, D. (2005). Relationship influences on teachers' perceptions of academic competence in academically at-risk minority and majority first grade students. *Journal of School Psychology*, 43(4), 303–320. <https://doi.org/10.1016/j.jsp.2005.07.001>
- Kellow, J. T., & Jones, B. D. (2008). The effects of stereotypes on the achievement gap: Reexamining the academic performance of African American high school students. *Journal of Black Psychology*, 34(1), 94–120. <https://doi.org/10.1177/0095798407310537>
- Kennelly, L., & Monrad, M. (2007). Approaches to Dropout Prevention: Heeding Early Warning Signs with Appropriate Interventions. In *American Institutes for Research* (p. 33). American Institutes for Research. <https://eric.ed.gov/?id=ED499009> Washington, DC: National High School Center at the American Institutes for Research.
- Kincade, L., Cook, C., & Goerd, A. (2020). Meta-analysis and common practice elements of Universal approaches to improving student-teacher relationships. *Review of Educational Research*, 90(5), 710–748. <https://doi.org/10.3102/0034654320946836>
- Koomen, H. M. Y., & Jellesma, F. C. (2015). Can closeness, conflict, and dependency be used to characterize students' perceptions of the affective relationship with their teacher? Testing a new child measure in middle childhood. *British Journal of Educational Psychology*, 85(4), 479–497. <https://doi.org/10.1111/bjep.12094>
- Kurki, K., Järvenoja, H., Järvelä, S., & Mykkänen, A. (2016). How teachers co-regulate children's emotions and behaviour in socio-emotionally challenging situations in day-care settings. *International Journal of Educational Research*, 76, 76–88. <https://doi.org/10.1016/j.ijer.2016.02.002>
- Kyles, C. R., & Olafson, L. (2008). Uncovering preservice teachers' beliefs about diversity through reflective writing. *Urban Education*, 43(5), 500–518. <https://doi.org/10.1177/0042085907304963>
- Lee, V. E., & Smith, J. B. (2001). *Restructuring High Schools for Equity and Excellence: What Works*. US: Teachers College Press.
- Liu, P. P., Savitz-Romer, M., Perella, J., Hill, N. E., & Liang, B. (2018). Student representations of dyadic and global teacher-student relationships: Perceived caring, negativity, affinity, and differences across gender and race/ethnicity. *Contemporary Educational Psychology*, 54, 281–296. <https://doi.org/10.1016/j.cedpsych.2018.07.005>
- McHugh, R. M., Horner, C. G., Colditz, J. B., & Wallace, T. L. (2013). Bridges and barriers: Adolescent perceptions of student-teacher relationships. *Urban Education*, 48(1), 9–43. <https://doi.org/10.1177/0042085912451585>
- Milner, H. R. (2005). Stability and change in US prospective teachers' beliefs and decisions about diversity and learning to teach. *Teaching and Teacher Education*, 21(7), 767–786. <https://doi.org/10.1016/j.tate.2005.05.010>
- Montoro, J. P., Kilday, J. E., Rivas-Drake, D., Ryan, A. M., & Umaña-Taylor, A. J. (2021). Coping with discrimination from peers and adults: implications for adolescents' school belonging. *Journal of Youth and Adolescence*, 50(1), 126–143. <https://doi.org/10.1007/s10964-020-01360-5>
- Mordechay, K., & Orfield, G. (2017). Demographic transformation in a policy vacuum: The changing face of US metropolitan society and challenges for public schools. *The Educational Forum*, 81(2), 193–203.
- Musca, S. C., Kamiejski, R., Nugier, A., Meot, A., Er-Raify, A., & Brauer, M. (2011). Data with hierarchical structure: Impact of intraclass correlation and sample size on Type-I error. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2011.00074>
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide* (8th ed.). Muthén & Muthén.
- O'Connor, E. E., Dearing, E., & Collins, B. A. (2011). Teacher-child relationship and behavior problem trajectories in elementary school. *American Educational Research Journal*, 48(1), 120–162. <https://doi.org/10.3102/0002831210365008>
- Pianta, R. C., la Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal*, 102(3), 225–238. <https://doi.org/10.1086/499701>
- Pianta, R. C., & Nimetz, S. L. (1991). Relationships between children and teachers: Associations with classroom and home behavior. *Journal of Applied Developmental Psychology*, 12(3), 379–393. [https://doi.org/10.1016/0193-3973\(91\)90007-Q](https://doi.org/10.1016/0193-3973(91)90007-Q)
- Quin, D. (2017). Longitudinal and contextual associations between teacher-student relationships and student engagement: A systematic review. *Review of Educational Research*, 87(2), 345–387. <https://doi.org/10.3102/0034654316669434>
- Raffaelli, M., Crockett, L. J., & Shen, Y.-L. (2005). Developmental stability and change in self-regulation from childhood to adolescence. *The Journal of Genetic Psychology*, 166(1), 54–76. <https://doi.org/10.3200/GNTP.166.1.54-76>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). US: SAGE Publications.
- Roeser, R. W., Strobel, K. R., & Quihuis, G. (2002). Studying early adolescents' academic motivation, social-emotional functioning, and engagement in learning: Variable- and person-centered approaches. *Anxiety, Stress, & Coping*, 15(4), 345–368. <https://doi.org/10.1080/1061580021000056519>
- Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*, 81(4), 493–529. <https://doi.org/10.3102/0034654311421793>
- Shih, M., & Sanchez, D. T. (2005). Perspectives and research on the positive and negative implications of having multiple racial identities. *Psychological Bulletin*, 131(4), 569–591. <https://doi.org/10.1037/0033-2909.131.4.569>
- Steinberg, L. (2001). We know some things: Parent-adolescent relationships in retrospect and prospect. *Journal of Research on Adolescence*, 11(1), 1–19. <https://doi.org/10.1111/1532-7795.00001>
- Tanner-Smith, E. E., Durlak, J. A., & Marx, R. A. (2018). Empirically based mean effect size distributions for universal prevention programs targeting school-aged youth: A review of meta-analyses. *Prevention Science*, 19(8), 1091–1101. <https://doi.org/10.1007/s1121-018-0942-1>
- 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.
- Thijs, J., Westhof, S., & Koomen, H. (2012). Ethnic incongruence and the student-teacher relationship: The perspective of ethnic majority teachers. *Journal of School Psychology*, 50(2), 257–273. <https://doi.org/10.1016/j.jsp.2011.09.004>
- Udry, J. R., Li, R. M., & Hendrickson-Smith, J. (2003). Health and behavior risks of adolescents with mixed-race identity. *American Journal of Public Health*, 93(11), 1865–1870. <https://doi.org/10.2105/ajph.93.11.1865>
- Valiente, C., Parker, J. H., Swanson, J., Bradley, R. H., & Groh, B. M. (2019). Early elementary student-teacher relationship trajectories predict girls' math and boys' reading achievement. *Early Childhood Research Quarterly*, 49, 109–121. <https://doi.org/10.1016/j.ecresq.2019.05.001>
- Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education.

- Educational and Psychological Measurement*, 52(4), 1003–1017. <https://doi.org/10.1177/0013164492052004025>
- Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1993). On the assessment of intrinsic, extrinsic, and amotivation in education: Evidence on the concurrent and construct validity of the academic motivation scale. *Educational and Psychological Measurement*, 53(1), 159–172. <https://doi.org/10.1177/0013164493053001018>
- 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>
- Vera, E. M., Polanin, J. R., Polanin, M., & Carr, A. L. (2018). Keeping latina/o students in school: Factors related to academic intentions of students at risk for school dropout. *Journal of Latina/o Psychology*, 6(1), 34–48. <https://doi.org/10.1037/lat0000082>
- Wagle, R., Dowdy, E., Yang, C., Palikara, O., Castro, S., Nylund-Gibson, K., & Furlong, M. J. (2018). Preliminary investigation of the psychological sense of school membership scale with primary school students in a cross-cultural context. *School Psychology International*, 39(6), 568–586. <https://doi.org/10.1177/0143034318803670>
- Wang, M.-T., & Eccles, J. S. (2012). Social support matters: Longitudinal effects of social support on three dimensions of school engagement from middle to high school. *Child Development*, 83(3), 877–895. <https://doi.org/10.1111/j.1467-8624.2012.01745.x>
- Whaley, A. L., & Noel, L. T. (2013). Academic achievement and behavioral health among Asian American and African American adolescents: Testing the model minority and inferior minority assumptions. *Social Psychology of Education*, 16(1), 23–43. <https://doi.org/10.1007/s11218-012-9206-2>
- Wheelock, A., & Miao, J. (2005). The ninth-grade bottleneck: An enrollment bulge in a transition year that demands careful attention and action. *School Administrator*, 62(3), 36.
- Williamson, A. A., Modecki, K. L., & Guerra, N. G. (2015). *SEL programs in high school*. US: The Guilford Press.
- Wing, J. Y. (2007). Beyond black and white: The model minority myth and the invisibility of Asian American students. *The Urban Review*, 39(4), 455–487. <https://doi.org/10.1007/s11256-007-0058-6>
- Wun, C. (2016). Unaccounted foundations: Black girls, anti-black racism, and punishment in schools. *Critical Sociology*, 42(4–5), 737–750. <https://doi.org/10.1177/0896920514560444>
- 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>
- Yoo, H. C., Miller, M. J., & Yip, P. (2015). Validation of the internalization of the model minority myth measure (IM-4) and its link to academic performance and psychological adjustment among Asian American adolescents. *Cultural Diversity and Ethnic Minority Psychology*, 21(2), 237–246. <https://doi.org/10.1037/a0037648>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.