FRONT MATTER

Title

- A Scalable Empathic-Mindset Intervention Reduces Group Disparities in School Suspensions
- Empathic-Mindset Intervention Reduces Suspensions

Authors

Jason A. Okonofua^{1*}, J. Parker Goyer², Constance A. Lindsay³, Johnetta Haugabrook⁴, & Gregory M. Walton²

Affiliations

¹University of California, Berkeley; Berkeley, USA

²Stanford University; Stanford, USA

³University of North Carolina; Chapel Hill, USA.

⁴Multi-Tiered System of Supports; USA.

*Corresponding author. Email: okonofua@berkeley.edu.

Abstract

Suspensions remove students from the learning environment at high rates throughout the United States. Policy and theory highlight social groups that face disproportionately high suspension rates—racial-minoritized students, students with a prior suspension, and students with disabilities. We used an active placebo-controlled, longitudinal field-experiment ($N_{teachers}=66$, $N_{students}=5,822$) to test a scalable “empathic-mindset” intervention, a 70-minute online exercise to refocus middle-school teachers on understanding and valuing the perspectives of students and on sustaining positive relationships even when students misbehave. In pre-registered analyses, this exercise reduced suspension rates especially for Black and Hispanic students, cutting the racial disparity over the school year by 45%. Significant reductions were also observed for other groups of concern. Moreover, reductions persisted through the next year when students interacted with new teachers, suggesting that empathic treatment with even one teacher in a critical period can improve students’ trajectories.

Teaser

An “empathic-mindset” intervention with teachers reduces suspensions and disparities in them.

MAIN TEXT

Introduction

The widespread systemic exclusion of children from school in the United States through suspensions (1-3) is not only associated with poor academic performance and negative school and friendship experiences (4-5), but predicts longer run negative outcomes. For students,

suspensions are associated with school dropout (6), and, in turn, lower earning potential (7) and greater substance use (8), crime (9), and risk of incarceration (10). Moreover, high rates of suspensions pose a fundamental threat to equal opportunity, as racial disparities in suspensions have been observed consistently for decades, as early as when schools were first desegregated (11). The latest national records find that Black students, for instance, are nearly four times more likely to be suspended than White students (3). They are also at greater risk of suffering negative life outcomes as a consequence of those suspensions (12).

High rates of suspension are also a problem for teachers and society at large. For teachers, difficulty managing conflict in class can undermine their ability to promote student learning (13). It can also be disheartening, and contribute to stress, burnout, low job satisfaction, and ultimately attrition from the profession (14-15). Such attrition is a national concern with declines in teacher supply in school districts across the United States (16). Society at large bears the costs of increased health problems, substance abuse, and crime. In economic terms, suspensions recorded for a single cohort of 10th grade students in California and Florida produced an estimated 67,000 dropouts in the United States who, over the course of their adult lives, were estimated to generate more than \$35 billion in additional government costs, including increased expenditures on health and social services and reduced tax revenue (17).

What causes high rates of suspensions, especially for students of color? Growing evidence points to racial bias. For instance, laboratory experiments find that teachers are more likely to knit together a series of misbehaviors as a pattern, to view a student as a troublemaker, and to punish them more severely, if the student is Black as compared to White (18). Moreover, correlational research finds an association between county aggregates of anti-Black/pro-White implicit and explicit bias and racial disparities in in- and out-of-school suspensions (19).

Counties with more anti-Black/pro-White bias have larger disparities in suspension rates between Black and White students. Quasi-experimental evidence from North Carolina demonstrates a significant relationship between having a same race teacher and reduced exclusionary discipline (e.g., suspensions) for Black students in particular (20). This reduction is driven mainly by fewer referrals for interpersonal offenses that require the subjective judgment of the teacher, such as willful defiance, pointing to a role for racial bias (21).

In general, strategies to contend with racial bias have aimed to reduce this bias, which, in theory, would mitigate disparities in suspension rates. While bias-reduction approaches have not yet been tested in the context of school discipline, their effectiveness elsewhere has been notably limited. Typical and theory-based approaches show weak and short-lived effects at best (22-23). Moreover, reductions in anti-Black bias have not predicted changes in behaviors or outcomes (24).

Typical approaches to racial disparities in discipline have focused on prescriptive policies or heavy-touch skill-building programs, each with mixed results. First, some states have banned common interpersonal offenses for which racial disparities are largest (e.g., “willful defiance”) as a basis for suspensions (25-27). These approaches may reduce this classification of offenses but they do not necessarily prevent the offenses or exclusion from the classroom by other means (e.g., in-school suspensions). As a result, racial disparities in exclusionary discipline can remain. Second, many districts have adopted Positive Behavioral Intervention Supports (PBIS), which uses multi-tier models that call for professional “behavioral coaches” and individualized trainings for students to learn better behavior (28-29). PBIS has reduced overall suspension rates in elementary schools but is rarely effective at higher grade levels when suspension rates and racial disparities both spike (30). Other efforts to reduce suspensions have primarily benefited

White students and have thus increased racial disparities over time (31). This lack of consistent effectiveness in mitigating racial disparities may be due to a lack of direct focus on psychological mechanisms by which race-ethnicity affects teacher-student interactions and relationships.

Our theoretical approach foregrounds teacher-student interactions, and bias that arises in the course of these interactions, in contributing to high and disproportionate suspension rates. Even as racial stereotypes can lead teachers to be more likely to view racially stigmatized students' misbehaviors as fixed aspects of their character (18), awareness of the same stereotypes can lead Black students to fear and expect unfair treatment and thus to react to any perceived bias and disengage in a manner that can be interpreted as misbehavior (32). These perspectives between teachers and students interact and contribute to a recursive cycle of growing mistrust and disrespect, which can ultimately result in large racial disparities in suspensions between racial minoritized and White students (33).

To effectively address this cycle, we do not seek to reduce bias in teachers but to sideline it, to make bias less controlling of teachers' behavior. To do so, we elevate an ideal self in teachers, one that prioritizes strong working relationships with students for which bias would be maladaptive (34-35). In theory, this can lead misbehaving students to feel more respected by and, thus, respect for their teachers in circumstances of misbehavior and more motivation to behave well in class (36).

This approach complements and extends prior research with teachers. One study found that middle and high school teachers ($N_{teachers}=82$ for $N_{students}=979$) who participated in My Teacher Partner (MTP), a year-long intensive and individualized training program designed to help teachers reflect on their approaches to and relationships with students, reduced racial

disparities in teachers' office referrals as compared to a randomized control group (37). This program, however, involved a heavy-touch including multiple components, reducing its potential for scalability and scientists' ability to isolate the causal effect. In one component, teachers watched videos of their own classrooms with trained coaches throughout the year (38). The study also did not examine school suspensions. Nonetheless, this research suggests that teachers' mindsets (beliefs or attributions) about their relationships with their students are connected to their discipline practices and, moreover, that shifts in those mindsets can reduce racial disparities in disciplinary citations.

Most directly relevant, in another study a ~70-minute online empathic-mindset intervention, which encouraged middle school teachers to value students' perspectives during circumstances of misbehavior, cut yearlong suspension rates by 4.8 percentage points in a sample of 1,682 predominantly Hispanic students (36). This approach was tested with only a modest sample of teachers (e.g., just 31 teachers were randomized to condition) and the student sample was not diverse (e.g., just 7% White and 2% Black students), so its potential to mitigate racial disparities has yet to be determined. Yet, it suggests that a brief psychological intervention can mitigate actual student suspension rates.

The current research addresses four questions. First, it provides a critical replication of the empathic-discipline intervention with a far larger and more diverse student sample ($n_{teachers}=66$; $n_{students}=5,822$); indeed, this is the largest test to date of any randomized intervention with teachers aimed at reducing student suspensions. The opportunity for a brief, online and low-cost approach to reduce student suspensions is significant for theory, practice, and policy. It is essential to test the replicability of the effect in another region and district and with a different student and teacher sample.

Second, the larger and more diverse student sample allows us to evaluate whether the empathic-discipline intervention is most effective for racial minoritized students and can thus reduce racial disparities in school suspensions.

Third, the current research explores more deeply processes by which the empathic-discipline intervention reduces student suspensions. These processes, moreover, have implications for the duration of the effect as students move from grade to grade. While both MTP and the empathic-discipline intervention are randomized to teachers with students and other teachers unaware of this condition assignment, our theory understands suspensions as emanating not from the simple behavior of teachers alone but from patterns of interaction that develop between teachers and students, including trust or mistrust that each forms about the other, and thus cycles of behavior that improve or worsen over time (33). From this perspective, teachers are just one avenue into a system of relationships between students and teachers and other avenues should be possible too. Indeed, addressing students' trust of teachers and feelings of belonging in school can also reduce disciplinary citations among racial minoritized youth (39). In one study, an intervention aimed at bolstering students' sense of belonging at the beginning of middle school reduced disciplinary citations among Black boys through the end of high school. It achieved this long-term reduction by preventing the development of adverse patterns of teacher-student interaction in 6th and 7th grades and, in turn, supporting a stronger sense of belonging in school and reduced fears about being stereotyped with time (32).

Does the empathic-discipline intervention similarly shift students' behaviors and, thus, improve their outcomes even beyond the treated teacher? The previous trial provides initial evidence for this theorizing. Specifically, even as only students' math teachers were randomized to condition, suspensions could result from incidents that occur at any time or place throughout

each school day (e.g., science class, hallway, recess, school bus, etc.). In the one school that tracked the source of the suspension ($n=559$ middle school students), only 7.4% of referrals were from teachers eligible to participate in the intervention (36); even entirely eliminating these referrals from analyses, the treatment effect reducing students' risk of suspension with treatment remained. The results imply that more positive treatment from one teacher triggered by the empathic-discipline intervention caused a broader improvement in students' behavior throughout the school and school day. Here we ask: Might this process also be developmental, a shift in fundamental ways students interact at school, and thus extend into the subsequent academic year, even after students have little or no further regular exposure to the treated teacher (32)? With the larger sample and by including data from the next school year we are able to address this question.

Finally, the large and diverse sample allows for exploration of the intervention's efficacy with other policy-relevant groups, in particular students with a history of suspensions and students with a special education status. Recent national policy highlights elevated suspension rates for both groups (3) and both may be subject to similar adverse recursive teacher-student dynamics that we theorize give rise to high rates of suspension among racial-minoritized students. Teachers may become more punitive toward students with a history of misbehavior (18;33), expect more misbehavior, and experience greater emotional exhaustion and burnout (40). They may also be more concerned that students with disabilities could disrupt their classes and want them removed, as compared to typically developing students (41). Indeed, empathy is harder in general across group lines or when another person's behavior or status is seen as more difficult or distressing (42-43). In these contexts, an intentional exercise that encourages teachers to take an empathic stance toward students when they misbehave may help.

We also explored effects by gender, as boys are more likely than girls to get suspended (3). However, boys are not broadly stigmatized in school or less likely to receive empathy and thus may not show greater reductions in suspensions with treatment.

An additional contribution of this paper is the availability of school-level administrative data that allows for the exploration of effects for students of teachers in the same schools not assigned to either condition in the launch period. Thus, complementing experimental analyses, we include quasi-experimental analyses using these teachers as a second, non-randomized comparison group, increasing statistical power and testing the robustness of the effects. The district discipline records also include information on the types of suspensions students might receive: in-school suspension (student is prohibited from going to class) and out-of-school suspension (student is prohibited from coming on campus). In-school suspensions tend to be for minor offenses and ones more subjective and open to interpretation (e.g., willful defiance). Out-of-school suspensions tend to be for more objective and major offenses (e.g., possession of a firearm). In encouraging a more empathic response for misbehaviors that are open to interpretation, the empathic-discipline intervention may have more robust effects on in-school suspensions (34).

Current Research

The present research tested whether the empathic-mindset intervention could (1) reduce suspension rates in a large ($n_{students}=5,822$) and diverse (16.5% Black; 14.7% Hispanic; 57.8% White; 6.2% Asian; 0.003% American Indian; 4.5% two or more races; 49.4% female; average age 13.05; $SD=0.66$) student sample (see Tables S1-S4), (2) mitigate racial disparities in suspension rates, and (3) produce effects that persist to a subsequent year. Following past research (36), math teachers ($n=66$; $n_{control}=36$, $n_{treatment}=30$) were randomized to condition early

in the school year. They were predominantly White (control=84.8%; treatment=85.7%) and female (control=75.8%; treatment=82.1%), as is the case nationally (80% and 77%, respectively) (33), and experienced ($M_{\text{control}}=11.42$, $SD=9.39$; $M_{\text{treatment}}=10.57$, $SD=7.38$ years teaching in the district) (Tables S5-S8). We focus on middle schools for two reasons. First, this is when suspension rates and disparities in suspensions spike for students (44). Second, the intervention content focuses specifically on this context and developmental stage, including students' experiences of early adolescence, puberty, identity development, and learning to navigate multiple relationships with adults. Teachers and students were disbursed across 20 middle schools in 17 cities (Table S9), all in the same large public district in the southeastern region of the United States, which thus had similar curricular offerings and discipline policies.

There were no significant differences in race-ethnicity, gender, age, years teaching, average prior-year suspended students, childhood SES, or baseline psychometrics between teachers assigned to treatment ($N=30$) versus control ($N=36$) in semester 1, all $ps \geq .08$, indicating successful randomization. There also were no treatment versus control differences in their students' gender, previous-year suspension status, intervention-year special education status, or probability of being American Indian, two or more races, or Asian versus White, all $ps \geq .22$. However, the probability of being Black or Hispanic versus White was somewhat higher in the control condition, $P_C=32.3\%$ vs. $P_T=29.8\%$, $b=0.11$, $p=.047$, $OR=1.12$, and student age was somewhat lower, $M_C=13.03$, $M_T=13.08$, $b=-0.047$, $p=.007$ (SOM, Tables S3-7, Figure S1).

Students were 7th and 8th grade students, nearly all of whom ($N=5,533$; 95%) had school record discipline data from the district from middle school (6th and 7th grade) in the prior year; 6th grade students were excluded because they did not have prior middle school suspension records (both the nature of discipline records and rates of discipline incidents differ meaningfully in

elementary school); moreover, 5th grade suspension records were not available. Nearly all students of teachers assigned to condition during the designated launch of the intervention had only one math teacher assigned to condition ($N=5,600$; 96.2%).

Specific pre-registered hypotheses are indicated below in section headers (45). Complete model details are provided in SOM (p. 1). The pre-registration prioritized the main effect of condition and the student race-ethnicity $\times$ condition interaction. The scope of the research expanded after initial analyses and additional effects were explored. Which analyses were not pre-registered are also marked in the Results section. Additional survey measures administered before teachers engaged with randomized materials are not the focus of this project and thus not reported.

A total of 175 math teachers met eligibility for participation in the implementation period, Fall semester 2017. Teachers had to have been assigned to condition during that semester to be considered assigned to condition. While 66 were randomly assigned to condition (“assigned teachers”), 109 did not participate during the implementation period and thus were not assigned to condition (“unassigned teachers”). The main effect was tested with both the full sample (all teachers eligible for participation) and only teachers randomized to condition (excluding unassigned teachers). The former increases statistical power and provides a second, non-randomized control group; the latter provides a pure experimental test. In all models, we control for students’ race-ethnicity, gender, previous year suspension status, previous year math course suspension rate, and previous year teacher suspension rate. For details see SOM (pp. 5-8).

Results

Main effect (pre-registered). Did the empathic-mindset intervention reduce the probability of students spending at least one day on a suspension? Yes, marginally, when the full sample of teachers eligible for participation ($N=173$) and their students ($N=13,210$) is included in

the model. Students with a teacher randomized to the treatment were less likely to be suspended over the course of the academic year ($M=17.3\%$), as compared to students whose math teacher(s) was randomized to the active control condition ($M=20.4\%$), $b=0.031$, 95% $CI[-0.000, 0.061]$, $SE=.016$, $z=1.95$, $p=.051$, and significantly compared to otherwise similar students with unassigned teachers ($M=22.6\%$), $b=0.053$, 95% $CI[0.025, 0.080]$, $SE=.014$, $z=3.76$, $p<.001$ (one-tailed: $p<.001$). The effects trended in the same hypothesized direction with the smaller sample including only those students with assigned teacher(s) ($N_{\text{teachers}}=66$; $N_{\text{students}}=5,822$), $b=0.024$, 95% $CI[-0.007, 0.055]$, $SE=.016$, $z=1.51$, $p=.130$ (one-tailed: $p=.065$). For brevity and clarity, the remaining results focus solely on the sample of students with at least one fall-semester math teacher assigned to condition. Results for the model including the non-randomized sample is described in SOM, Tables S10-S11).

We also explored two key types of suspension. During the year prior to the intervention, among students with available data ($N=5,533$) whose math teachers were assigned to condition during the launch period, 11.6% of students spent at least one day on in-school suspension while 6.9% spent at least one day on out-of-school suspension. During the intervention year, students with a teacher randomized to the treatment were significantly less likely to receive an in-school suspension over the course of the academic year ($M=9.4\%$), as compared to students with a teacher(s) randomized to the active control condition ($M=14.2\%$), $b=0.048$, 95% $CI[0.019, 0.076]$, $SE=.014$, $z=3.29$, $p=.001$. The condition effect on out-of-school suspensions was not significant ($M_T=9.3\%$; $M_C: M=10.1\%$), $b=0.008$, 95% $CI[-0.018, 0.033]$, $SE=.013$, $z=0.59$, $p=.553$.

Students' race-ethnicity (pre-registered). Was the intervention particularly effective for students from racial groups that face negative behavioral stereotypes? Yes. Analyses focused on

Black and Hispanic students versus any other students, because these groups (a) are highlighted for national intervention efforts as they tend to experience the highest rates of suspensions across districts (3) and (b) had the highest suspensions rates in the previous year among all racial-ethnic groups in the present research's sample (Figures S2-3; for analyses of race-ethnicity coded differently and previous year suspension rates see SOM, Tables S12-18). In this group, the empathic-mindset treatment reduced the suspension rate by 5.6 percentage-points ($M_T=20.8\%$ vs. $M_C=26.5\%$), $b=0.056$, 95% $CI[0.017, 0.096]$, $SE=0.020$, $z=2.79$, $p=.005$. Other students had lower probability of suspensions, main effect of student race-ethnicity in control condition, $b=0.106$, 95% $CI[0.082, 0.130]$, $SE=0.012$, $z=8.64$, $p<.001$, and their probability did not differ by condition ($M_T=14.9\%$ vs. $M_C=15.9\%$), $b=0.009$, $p=.584$). The race-ethnicity $\times$ condition interaction was significant, $b=0.047$, 95% $CI[0.011, 0.083]$, $SE=0.019$, $z=2.55$, $p=.011$, reflecting a 45% reduction in the racial disparity with treatment. See Figure 1.

We also explored condition effects among Black and Hispanic students separately. The reduction was significant and similar in size for both Black students, ($M_T=30.7\%$ vs. $M_C=36.4\%$), $b=0.057$, 95% $CI[0.007, 0.106]$, $SE=0.025$, $z=2.23$, $p=.025$, and Hispanic students, even though probability of suspension was lower than for Black students, ($M_T=11.4\%$ vs. $M_C=16.7\%$), $b=0.054$, 95% $CI[0.005, 0.103]$, $SE=0.025$, $z=2.14$, $p=.032$.

Students' gender (not pre-registered). We also examined the role of gender. Neither the gender $\times$ condition interaction, $b=0.012$, 95% $CI[-0.019, 0.043]$, $SE=0.016$, $z=0.75$, $p=.454$, nor the race-ethnicity $\times$ gender $\times$ condition interaction was significant, $b=-0.001$, 95% $CI[-0.067, 0.066]$, $SE=0.034$, $z=-0.02$, $p=.988$, showing that the intervention was not differentially effective for either boys or girls (See Tables S19-21).

Fig. 1. Intervention year effects. Interaction between condition (empathic-mindset treatment v. active randomized control assignment among students' grade 7 or 8 middle school math teachers) and (1) students' race-ethnicity ($n_{teachers}=66$; $n_{students}=5,822$), (2) history of suspension ($n_{teachers}=65$; $n_{students}=5,533$), and (3) disability (special-education assignment) status ($n_{teachers}=66$; $n_{students}=5,822$) on student suspension rates (probability of one or more days on suspension) during the academic year of the intervention. Differences in sample sizes are due to missing data for the relevant predictor. Students who are not Black or Hispanic are White, Asian, American Indian, or students who reported two or more races. Error bars reflect ± 1 SE. The significance of simple effects is denoted by * significant simple effect, $p \leq .050$; ** significant simple effect, $p \leq .005$.

Additional student subgroups. Next, we explored intervention effects for students with a history of suspensions and students with disabilities. There was some overlap between student race-ethnicity and members in these subgroups, as Black and Hispanic students were overrepresented among students with a history of suspensions and among students with disabilities (see SOM, Table S1-2). However, the following models control for student race-ethnicity (and also for gender, as in the primary models).

Prior suspensions (not pre-registered). Was the intervention effective for students with history of suspension? Yes. Students with one or more suspensions the previous year were less

likely to be suspended if they had a teacher randomized to the treatment condition as compared to the control condition ($M_T=49.6\%$ vs. $M_C=56.1\%$) $b=.065$, 95% $CI[0.017, 0.113]$, $SE=0.024$, $z=2.67$, $p=.008$. Students without a prior suspension had far lower probability of suspension relative to previously-suspended students, main effect of prior suspension in control condition, $b=0.440$, 95% $CI[0.409, 0.471]$, $SE=0.016$, $z=27.46$, $p<.001$, and this probability of suspension did not differ significantly by condition ($M_T=10.6\%$ vs. $M_C=12.1\%$), $b=.015$, $p=.352$). The prior suspension $\times$ condition interaction was significant, $b=.050$, 95% $CI[0.005, 0.095]$, $SE=0.023$, $z=2.20$, $p=.028$.

Further exploratory analyses found, interestingly, that the intervention was notably effective for students with one prior suspension, ($M_T=30.7\%$ vs. $M_C=41.0\%$), $b=0.089$, 95% $CI[0.029, 0.149]$ $SE=0.031$, $z=2.92$, $p=.003$, but not significantly effective for students with two or more prior suspensions, ($M_T=67.6\%$ vs. $M_C=69.2\%$), $b=0.003$, $p=.930$. See Figure 1 and Tables S22-S25.

Students with special education status (not pre-registered). Was the intervention effective for students with special education status? Yes. These students were less likely to be suspended if they had a teacher randomized to the treatment condition (13.8%), as compared to the control condition (23.8%), $b=0.100$, 95% $CI[0.032, 0.168]$, $SE=0.035$, $z=2.87$, $p=.004$. Students without special education status had directionally lower probability of suspension, main effect of special education status in the control condition, $b=0.048$, 95% $CI[0.002, 0.094]$, $SE=0.023$, $z=2.06$, $p=.039$, and did not show a significant reduction in the treatment versus control condition ($M_T=17.0\%$ vs. $M_C=18.9\%$), $b=0.019$, 95% $CI[-0.012, 0.049]$, $SE=0.016$, $z=1.18$, $p=.238$). The disability $\times$ condition interaction was significant, $b=0.082$, 95% $CI[0.032, 0.168]$, $SE=0.033$, $z=2.45$, $p=.014$. See Figure 1, Tables S26-S27.

Subsequent year suspensions for students (not pre-registered). If students' math teacher was randomized to the treatment condition, were they less likely to receive a suspension the next academic year, as they interacted with new teachers? Yes.

To test this question, we examined those students who could be tracked at district middle schools through the school year following the year of randomization, $N=2,819$ students (46% of intervention year sample), with 56 different intervention-year teachers. Nearly all (99%) of these students were 7th graders in the intervention year and 8th graders the subsequent year. Almost all (94%) did not have the same math teacher in the subsequent year as the one randomized to condition during the intervention year. Thus, the small percentage who did are unlikely to drive the condition effects. However, we retained the full sample ($N=2,819$ students) to preserve the integrity of random assignment.

During the intervention year, this subsample showed a slightly larger (3.5 vs. 2.4 percentage points) treatment reduction in probability of suspension as the full sample ($N=5,822$ students) ($M_T=16.8\%$ vs. $M_C=20.3\%$), $b=0.035$, 95% $CI[-0.005, 0.075]$, $p=.085$. Strikingly, this reduction in probability of suspensions persisted through the next academic year, and in fact grew slightly, when nearly all students had new teachers ($M_T=16.7\%$ vs. $M_C=20.9\%$), $b=0.042$, 95% $CI[0.002, 0.081]$, $SE=0.020$, $z=2.06$, $p=.040$ (See Figure 2, Tables S28-26).

We also explored the role of race-ethnicity in this subsample. As in the full sample, intervention-year treatment effects were greater for Black and Hispanic students (7.6 percentage points, $p=.006$) than for other students (1.9 percentage points, $p=.380$), as reflected by a significant race-ethnicity $\times$ condition interaction, $b=0.057$, 95% $CI[0.004, 0.111]$, $SE=0.027$, $z=2.11$, $p=.035$. In the subsequent year, the simple effects of condition for these racial-ethnic groups persisted (Black and Hispanic: 8.2 percentage points, $p=.003$; Other: 2.5 percentage

points, $p=.242$), as did the race-ethnicity $\times$ condition interaction, $b=0.057$, 95% $CI[0.002, 0.112]$, $SE=0.028$, $z=2.03$, $p=.042$, (See Tables S29-30), a 56% reduction in the racial disparity with treatment.

Figure 2. Subsequent year effects. Condition effects (active randomized control v. empathic-mindset treatment) on suspension rates (one or more suspension days) in the year *after* teachers had been randomized to condition for all students, neither Black or Hispanic students, and Black or Hispanic students. Data reflect the subset of students who could be tracked for two years ($n=56$ teachers, $n=2,819$ students). Error bars reflect ± 1 SE. * significant simple effect, $p \leq .050$; ** significant simple effect, $p \leq .005$.

Discussion

The present research shows that a brief empathic-mindset intervention can meaningfully reduce suspension rates and mitigate racial-ethnic disparities in suspensions, replicating and extending past research.

The results implicate teachers' interactions with students in students' risk of suspensions. While teachers were the participants in the experiment and the subject of randomization, the focus was on their mindsets about interacting and maintaining positive relationships with students when they misbehave and the outcome—suspensions—was a student outcome. The findings thus point to the importance of interactions and relationships between teachers and students relative to teachers or students and their skills, biases, or behaviors alone. Moreover, they isolate teachers' mindsets as an opportunity for improvement above and beyond other factors.

The pre-registered reduction in suspension was greatest for Black and Hispanic students gives greater confidence in this effect. This finding is important both for theory and for application. It suggests that teachers' ability to empathize and maintain positive relationships with students in times of misbehavior—the focus of the intervention—may be most challenged with racial-minoritized students. Indeed, we have theorized that this challenge is made more severe by a context that includes racial stereotypes about misbehavior for both teachers and students (33). Thus, focusing on empathy and relationships at these times can have differential benefits for minoritized students.

In mitigating racial inequality in suspension rates, the intervention addresses a major policy priority (3). The implications for policy are deepened in two ways. First, this reduction was achieved in a ~70-minute online exercise with teachers. It was not necessary, for instance, to reach students individually (32; 39) or to implement an intensive professional development program for teachers (37-38). The intervention could be brief precisely because it did not attempt to build teachers' skills or to systematically change their practices. It focused only on mindsets. While we believe there are important nuances in the effective delivery of the empathic-mindset

intervention (e.g., that the exercise is presented to teachers as honorific, not remedial) (46), and while there are likely to be important boundary conditions around its effects (47), the empathic-discipline intervention is certainly easier and cheaper for schools and districts to administer than alternatives. An important direction for future research is identifying these boundary conditions, and thus where the intervention can be used effectively (48). More broadly, the results illustrate the potential to target the mindsets of “gatekeepers”—people in positions of power—to improve the experiences of many (49). Here, a shift in dozens of teachers’ mindsets improved outcomes for thousands of students.

Second, even as the empathic-mindset intervention benefitted Black and Hispanic students most, the direction of the effect was positive for all groups and significant for other policy-relevant groups susceptible to biased perception and a lack of empathy, namely students who had been suspended previously and those with disabilities (40-41). Yet it is notable that the intervention was not differentially effective for boys as compared to girls, even as boys are more likely to get suspended than girls (3). Previous research suggests that boys tend to engage in a broader range of misbehaviors than girls, including offenses that might be less malleable to intervention (i.e. drugs, weapons). (50). Future research may further explore how teachers respond to these misbehaviors and whether there are other opportunities to reduce the elevated risk of suspensions boys face.

The results also provide further insight into for whom and how the intervention works. First, it is notable that the empathic-mindset intervention focused on teachers’ approach to relationships with students in general when behavioral issues arise. It did not name or provide individualized guidance for relationships with any one of their students. This quality may inform the efficacy of the intervention and its limitations. Even as the treatment was highly effective in

reducing suspensions for students with one prior suspension (8.3% of students with available data), consistent with its goal in preventing the escalation of disciplinary responses following a prior misbehavior (9-10; 23), it did not significantly reduce suspensions for students for whom this escalation process had not yet begun (no suspensions the previous year; 80.8% of students with available data). Nor did it reduce suspensions for students for whom this process was better established (two or more prior suspensions; 10.9%). Students with one prior suspension at risk for more may represent a “sweet spot” for the intervention. Indeed, laboratory experimentation has shown that the intervention’s content can cause teachers to be less likely to view a student with a prior misbehavior as a troublemaker and to want to respond with less severe discipline (34). Meanwhile, students with a more extensive history of misbehavior may require a more individuated approach that focuses on the educator’s specific relationship with that student (36).

The focus on students in general may similarly inform the differential effect of the treatment on reducing in-school as compared to out-of-school suspensions. While both kinds of suspensions remove students from the learning environment, in-school suspensions are often for less severe offenses, for more directly relational offenses (e.g., “insubordination”), and for offenses that entail greater subjective judgment. This relative ambiguity may provide more room for a shift in teachers’ mindset and an improvement in patterns of interaction to reduce removals from the learning environment. Out-of-school suspensions, by contrast, typically require documented rationale and remanding of the student to the custody of a legal guardian. They may require a more individuated approach that focuses on particular patterns of previous misbehaviors (e.g., multiple in-school suspensions), more severe and objective offense (e.g., fighting), and may require involvement of other actors (e.g., principals) (51).

Second, a striking finding involved the persistence of the reduction in suspensions for students into the next academic year. To persist even as students interact with new teachers, it seems likely that the treatment shifted adolescents' developing beliefs about the kinds of relationships they have or can have with teachers (32). The intervention's strategic focus on teachers' approach to relationships with students may help teachers empathize with students and offer students what they crave (e.g., respect) during key interactions. Previous laboratory experimentation has found that students feel greater respect for teachers and more motivation to behave well and follow teachers' instructions when a teacher has responded in an empathic manner to their misbehavior (36). Also consistent with this theorizing, past direct-to-student interventions, when delivered at developmentally important junctures, can have powerful, lasting benefits for adolescents by sustaining their trust of teachers and sense of belonging in school (52; 37). A limitation of the present study is that we did not assess such student beliefs and trust directly. Understanding how empathic treatment from one teacher may contribute to this process is an important direction for future research.

The persistence of the effect for students also raises the question of whether the effect would also persist for teachers into a new year, as they interact with new students. While this is possible, teachers, unlike middle school students, are not developing new beliefs about the kinds of interactions and relationships they can have in general with students; they are, instead, developing beliefs about specific cohorts of students. If so, it may be that persistent benefits at the teacher level will require reactivation of empathic-mindset themes with each new cohort of students. The design of this experiment did not allow for exploration of this question, because some 6th grade students were exposed to a treated teacher during the intervention, reducing

power to detect the effect of 7th grade teachers treated in the prior year on new 7th graders whose teachers had not previously been treated. Yet future research may explore this question.

It will also be exciting explore effects in adjacent age groups, both among 6th grade students, ideally in K-8 schools that have the same context, policies, and procedures for discipline in both elementary and middle school, and among high school students.

With these open questions in mind, the results highlight the importance of empathy in pivotal teacher-student relationships and its capacity to benefit children at the most risk of inequity in education and beyond.

Materials and Methods

Empathic-Mindset: A Focus on Teacher-Student Relationships

We focus on teachers and students together and conflict that can arise between them. The “empathic-mindset” intervention is a brief online reading and reflection exercise. It invites teachers to reflect on the pivotal opportunity they have to help students grow and learn when they misbehave, including to listen to students and understand their perspective, and to sustain positive relationships in circumstances of misbehavior. Past research found initial evidence that this empathic-mindset can sideline—render immaterial—an otherwise punitive response in teachers to student misbehavior. Moreover, in a field trial with a predominantly Hispanic sample, the intervention reduced suspension rates among students a teacher taught by 4.8 percentage points over an academic year (36).

The empathic-mindset intervention encourages an empathic approach to understanding and responding to student misbehavior with targeted articles, narratives, and thought exercises. The materials focus on students’ social and emotional development, teachers’ capacity to help

students develop prosocial skills, and the value of prioritizing students' perspectives. The first page of the module introduced teachers to the program by saying,

In this web module, we will share with you some... research. Then, we will ask you for your input as a professional educator. We are especially interest in your thoughts about how teachers like you can and do use these ideas to have better interactions with students and to improve their lives (36).

Teachers are told that their responses will be incorporated in trainings for future teachers and thus allowing teachers to assume the role as experts rather than treating them as recipients of an intervention. Teachers then read how “everyone has a personal story about a great teacher who influenced his or her life” and these stories often involve the teacher effectively communicating that they care for and support students, including by listening to and valuing student's perspectives. The materials emphasize that doing so can also better-equip teachers to interact with students in ways that nurture their growth into responsible, motivated young adults.

These ideas are reinforced in student and teacher narratives, relatable stories that illustrate how understanding students' perspectives can be enlightening and useful. For example, a Black 7th grader shares:

In middle school, I didn't feel like I belonged. It seemed like the teachers always called on the other students. So I didn't pay attention in class and sometimes I got in trouble. One day I got detention and, instead of just sitting there, my teacher talked with me about what happened. He really listened to me. And then he told me that he had trouble sometimes in middle school but that it gets better. It felt good to know I had someone I could trust in school (36).

For these narratives and throughout the content, students' perspectives were presented as paradigmatic examples, not teachers' specific students. Thus, their purpose was to encourage teachers to learn more about their own students, and give them tools to do so, rather than to use the content as a substitute or to basis for assumptions about their own students.

Throughout, the materials are interactive. For instance, teachers are asked how they maintain positive relationships with their own students even when they misbehave. Last, teachers are asked to reflect on the materials they had reviewed and their own experiences with students to write a letter to a new teacher in their district to help them navigate relationships with students. Such “saying-is-believing” tasks can help embed a psychological message in teachers' understanding of past experiences and their use of it going forward.

The control condition module was similar in length and exercises. However, the content addressed how technology can enhance teachers' capacity to engage their students. While this content also involves teacher-student relationships and supporting students' learning, it does not prioritize address teachers' responses to misbehavior (34).

All analyses control for individual differences (e.g., student gender, student race-ethnicity, suspension status in the previous year) (see Supplemental Materials, pp. 5-6). While most 96.2% of students only had one math teacher, some students had multiple math teachers which could confound randomization. However, there was demographic balance among students with multiple math teachers across conditions (see Table S3-S4). Methodology to account for this is included in the Supplemental Materials (p. 6). Before any experiment began, approval from the Institutional Review Board at the University of California at Berkeley was given.

References

1. J. Owens, S. S. McLanahan, Unpacking the Drivers of Racial Disparities in School Suspension and Expulsion. *Social Forces*. **98**, 1548–1577 (2019), doi:10.1093/sf/soz095.
2. M. Gopalan, Understanding the linkages between racial/ethnic discipline gaps and racial/ethnic achievement gaps in the United States. *Education Policy Analysis Archives*. **27**, 154 (2019).
3. US Department of Education Office for Civil Rights. 2013-2014 civil rights data collection: A first look. *US Department of Education Office for Civil Rights* (2016).
4. W. C. Jacobsen, School punishment and interpersonal exclusion: Rejection, withdrawal, and separation from friends. *Criminology*, **58**(1), 35-69 (2020).
5. W. C. Jacobsen, G. T. Pace, N. G. Ramirez, Punishment and inequality at an early age: Exclusionary discipline in elementary school. *Social forces*. **97**(3), 973-998 (2019).
6. A. L. Noltemeyer, R. M. Ward, C. McLoughlin, Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, **44**, 224-240 (2015).
7. Northeastern University Center for Labor Market Studies. *Left behind in America: The nation's dropout crisis*. Center for Labor Market Studies, Northeastern University (2009).
8. T. González, A. Etow, C. De La Vega, Health equity, school discipline reform, and restorative justice. *The Journal of Law, Medicine & Ethics*, **47**(2 suppl), 47-50 (2019).
9. T. J. Mowen, J. J. Brent, J. H. Boman IV, The effect of school discipline on offending across time. *Justice Quarterly*, **37**(4), 739-760 (2020).
10. P. Hemez, J. J. Brent, T. J., Mowen, Exploring the school-to-prison pipeline: How school suspensions influence incarceration during young adulthood. *Youth Violence and Juvenile Justice*, **18**(3), 235-255 (2020).
11. M. G. Yudof, Suspension and expulsion of black students from the public schools: Academic capital punishment and the Constitution. *Law and Contemporary Problems*, **39**(2), 374-411 (1975).
12. R. O. Welsh, S. Little, The school discipline dilemma: A comprehensive review of disparities and alternative approaches. *Review of Educational Research*, **88**(5), 752-794 (2018).
13. Bill and Melinda Gates Foundation. Learning about teaching: Initial findings from the Measures of Effective Teaching project (2013). Retrieved from http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf
14. K. Ferguson, L. Frost, D. Hall, Predicting teacher anxiety, depression, and job satisfaction. *Journal of teaching and learning*, **8**(1) (2012).
15. A. L. Yarrow, State of Mind. *Education Week*, **29**, 21-23 (2009).
16. L. Sutchter, L. Darling-Hammond, D. Carver-Thomas, Understanding teacher shortages: An analysis of teacher supply and demand in the United States. *Education Policy Analysis archives*, **27**(35) (2019).
17. R. W. Rumberger, D. J. Losen, The high cost of harsh discipline and its disparate impact (2016).
18. J. A. Okonofua, J. L. Eberhardt, Two strikes: Race and the disciplining of young students. *Psychological science*, **26**, 617-624 (2015).
19. T. Riddle, S. Sinclair, Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias. *Proceedings of the National Academy of Sciences*, **116**(17), 8255-8260 (2019).

20. C. A. Lindsay, C. M. Hart, Exposure to same-race teachers and student disciplinary outcomes for Black students in North Carolina. *Educational Evaluation and Policy Analysis*, **39**(3), 485-510 (2017).
21. R. J. Skiba, R. S. Michael, A. C. Nardo, R. L. Peterson, The color of discipline: Sources of racial and gender disproportionality in school punishment. *The urban review*, **34**(4), 317-342 (2002).
22. C. K. Lai, M. Marini, S. A. Lehr, C. Cerruti, J. E. L. Shin, J. A. Joy-Gaba... B. A. Nosek, Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, **143**(4), 1765 (2014).
23. C. K. Lai, A. L. Skinner, E. Cooley, S. Murrar, M. Brauer, T. Devos... B. A. Nosek, Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, **145**(8), 1001 (2016).
24. P. S. Forscher, C. K. Lai, J. R. Axt, C. R. Ebersole, M. Herman, P. G. Devine, B. A. Nosek, Author notes (2017).
25. Pre-K Student Discipline Amendment Act of 2015. D. C. Assemb. B, 21-0001 (2014-2015).
26. Public Counsel, California enacts first-in-the-nation law to eliminate student suspensions for minor misbehavior [Press Release] (2014). Retrieved from http://www.publiccounsel.org/press_releases?id=0088
27. Pupil discipline: Suspensions and expulsions: Willful defiance. Cal. Assemb. B, 420 (2013-2014), Chapter 660 (Cal. Stat. 2014).
28. W. M. Reinke, K. C. Herman, M. Stormont, Classroom-level positive behavior supports in schools implementing SW-PBIS: Identifying areas for enhancement. *Journal of Positive Behavior Interventions* **15**, 39–50 (2013).
29. W. Sailor, G. Dunlap, G. Sugai, R. Horner, (Eds). *Handbook of positive behavior support*. New York, NY: Springer (2009).
30. C. G., Vincent, J. R. Sprague, M. Pavel, T. J. Tobin, J. M. Gau, Effectiveness of schoolwide positive behavior interventions and supports in reducing racially inequitable disciplinary exclusion. *Closing the school discipline gap: Equitable remedies for excessive exclusion*, 207 (2015).
31. A. Porowski, R. O'Conner, A. Passa, Disproportionality in School Discipline: An Assessment of Trends in Maryland, 2009-12. REL 2014-017. *Regional Educational Laboratory Mid-Atlantic* (2014).
32. J. P. Goyer, G. L. Cohen, J. E. Cook, A. Master, N. Apfel, W. Lee, ... G. M. Walton, Targeted identity-safety interventions cause lasting reductions in discipline citations among negatively stereotyped boys. *Journal of Personality and Social Psychology*, **117**, 229 (2019).
33. J. A. Okonofua, G. M. Walton, J. L. Eberhardt, A vicious cycle: A social–psychological account of extreme racial disparities in school discipline. *Perspectives on Psychological Science*, **11**, 381-398 (2016).
34. J. A. Okonofua, A. D. Perez, S. Darling-Hammond, When policy and psychology meet: Mitigating the consequences of bias in schools. *Science advances*, **6**, (2020) eaba9479.
35. G. M. Walton, J. A. Okonofua, K. S. Remington, D. Hurst, A. Pinedo, E. Weitz, J. Ospina, H. Tate, J. L. Eberhardt, Lifting the Bar: A Relationship-Orienting Intervention Reduces Recidivism Among Children Reentering School from Juvenile Detention. *Psychological Science*, (In Press).

36. J. A. Okonofua, D. Paunesku, G. M. Walton, Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proceedings of the National Academy of Sciences*, **113**, 5221-5226 (2016).
37. A. Gregory, J. P. Allen, A. Y. Mikami, C. A. Hafen, R. C. Pianta, The promise of a teacher professional development program in reducing racial disparity in classroom exclusionary discipline. *Closing the school discipline gap: Equitable remedies for excessive exclusion*, **168** (2015).
38. A. Gregory, E. Ruzek, C. A. Hafen, A. Y. Mikami, J. P. Allen, R. C. Pianta, My Teaching Partner-Secondary: A video-based coaching model. *Theory into practice*, **56**(1), 38-45 (2017).
39. D. S. Yeager, R. Johnson, B. J. Spitzer, K. H. Trzesniewski, J. Powers, C. S. Dweck, The far-reaching effects of believing people can change: Implicit theories of personality shape stress, health, and achievement during adolescence. *Journal of personality and social psychology*, **106**, 867 (2014).
40. A. M. Aloe, S. M. Shisler, B. D. Norris, A. B. Nickerson, T. W. Rinker, A multivariate meta-analysis of misbehavior and teacher burnout. *Educational Research Review*, **12**, 30-44 (2014).
41. B. G. Cook, M. Tankersley, L. Cook, T. J. Landrum, Teachers attitudes toward their included students with disabilities. *Exceptional children*, **67**, 115-135 (2000).
42. D. B. Bugental, P. C. Ellerson, E. K. Lin, B. Rainey, A. Kokotovic, N. O'hara, A cognitive approach to child abuse prevention (2010).
43. K. Schumann, J. Zaki, C. S. Dweck, Addressing the empathy deficit: beliefs about the malleability of empathy predict effortful responses when empathy is challenging. *Journal of personality and social psychology*, **107**(3), 475 (2014).
44. R. J. Skiba, R. H. Horner, C. G. Chung, M. K. Rausch, S. L. May, T. Tobin, Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, **40**(1), 85-107 (2011).
45. J. Okonofua, J. P. Goyer, Empathic Discipline Scaled (2019, December 4). Retrieved from <https://osf.io/8fdmk/>
46. J. A. Okonofua, M. Ruiz, Title of chapter. In G. M. Walton & A. J. Crum (Eds.), *Handbook of wise interventions*. Guilford Publications (2020).
47. G. M. Walton, D. S. Yeager, Seed and soil: Psychological affordances in contexts help to explain where wise interventions succeed or fail. *Current Directions in Psychological Science*, **29**(3), 219-226 (2020).
48. C. J. Bryan, E. Tipton, D. S. Yeager, Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, 1-10 (2021).
49. K. Lewin, Group decision and social change. In E. Maccoby, T. M. Newcombe, & E. L. Hartley (Eds.), *Readings in social psychology* (pp. 197–211). New York, NY: Holt, Rinehart, & Winston (1958).
50. R. J. Skiba, R. S. Michael, A. C. Nardo, R. L. Peterson, The color of discipline: Sources of racial and gender disproportionality in school punishment. *The urban review*, **34**(4), 317-342 (2002).
51. E. Chen, G. H. Brody, T. Yu, L. C. Hoffer, A. Russak-Pribble, G. E. Miller, Disproportionate school punishment and significant life outcomes: a prospective analysis of Black youths. *Psychological science*, 0956797621998308 (2021).

52. D. S. Yeager, V. Purdie-Vaughns, S. Y. Hooper, G. L. Cohen, Loss of institutional trust among racial and ethnic minority adolescents: A consequence of procedural injustice and a cause of life-span outcomes. *Child Development*, **88**, 658-676 (2017).

Acknowledgments

The authors would like to thank the school district for partnering on the research, David Paunesku, Sarah Gripshover, Eric Smith, and Project for Education Research that Scales (PERTS) for assistance with data collection and E. Adaora Okonofua, Shanette Porter, Michael Ruiz, and Equity, Diversity, and Empathy Navigation Sciences (EDENS) Lab, and Faheemah Mustafaa for their support and assistance with implementation of this project.

Funding:

Google/Tides Foundation (JAO)

Character Lab (JAO)

Mindset Science Solutions (JAO)

Author contributions:

Conceptualization: JAO, GMW

Methodology: JAO

Investigation and Experiment Supervision: JAO

Funding acquisition: JAO

Data processing and analysis: JPG

Project administration: JAO

Writing: JAO, JPG, GMW, CAL

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data and scripts needed to evaluate the conclusions in the paper are available with the pre-registration at Open Science Framework (34).

Figures and Tables

Fig. 1. Intervention year effects. Interaction between condition (empathic-mindset treatment v. active randomized control assignment among students' grade 7 or 8 middle school math teachers) and (1) students' race-ethnicity ($n_{teachers}=66$; $n_{students}=5,822$), (2) history of suspension ($n_{teachers}=65$; $n_{students}=5,533$), and (3) disability (special-education assignment) status ($n_{teachers}=66$; $n_{students}=5,822$) on student suspension rates (probability of one or more days on suspension) during the academic year of the intervention. Differences in sample sizes are due to missing data for the relevant predictor. Students who are not Black or Hispanic are White, Asian, American Indian, or students who reported two or more races. Error bars reflect ± 1 SE. The significance of simple effects is denoted by * significant simple effect, $p \leq .050$; ** significant simple effect, $p \leq .005$.

Fig. 2. Subsequent year effects. Condition effects (active randomized control v. empathic-mindset treatment) on suspension rates (one or more suspension days) in the year *after* teachers had been randomized to condition for all students, neither Black or Hispanic students, and Black or Hispanic students. Data reflect the subset of students who could be tracked for two years ($n=56$ teachers, $n=2,819$ students). Error bars reflect ± 1 SE. * significant simple effect, $p \leq .050$; ** significant simple effect, $p \leq .005$.

Supplementary Materials for

A Scalable Empathic-Mindset Intervention Reduces Group Disparities in School Suspensions

Jason A. Okonofua, J. Parker Goyer, Constance A. Lindsay, Johnetta Haugabrook, & Gregory M. Walton

Correspondence to: okonofua@berkeley.edu

This PDF file includes:

Materials and Methods
Tables S1 to S30
Figures S1 to S3

Materials

Hypothesis and Analytic Plan for Main Text Analyses

1. **Pre-Registered Analyses.** We pre-registered (34) the following hypotheses before data analyses were performed:
 - 1.1. Consistent with Okonofua, Paunesku, and Walton, 2016, the intervention, as compared to control condition, will reduce the probability of receiving a suspension.
 - 1.2. The intervention, as compared to control condition, will reduce the probability of receiving a suspension for Black, Hispanic, American Indian, and multiple race students. Further, there will be a two-way interaction between the intervention effect and the effect of student race such that the reduction in suspensions is driven by a reduction in suspensions for Black, Hispanic, American Indian, and multiple race students. Research shows that students from racially stigmatized groups face a heightened risk of stereotypes barring them from meaningful relationships with teachers (28).
2. **Non-Pre-Registered Analytic Plan.** We also followed an analytic plan (not pre-registered) to explore other relevant outcomes:
 - 2.1. Is the intervention effective for students at the most risk for suspensions (3)?
 - 2.1.1. Is the intervention effective for students with disabilities (eligible for special education)?
 - 2.1.2. Is the intervention effective for students with a history of suspension(s) (9-11)? Also, does the intervention have a sweet spot (9-10; 23) in disciplinary interactions that present manageable difficulty for teachers?
 - 2.2. Does the intervention effect persist for students to reduce their suspensions in a subsequent academic year with new teachers?
 - 2.3. Does the intervention effect persist for teachers to reduce suspensions for their new students in a subsequent academic year with new students?
3. **Additional pre-registered exploration**
 - 3.1. We will explore a three-way interaction among the intervention effect and the effects of student race and gender. The intervention effect will be larger for either (3a) Black, Hispanic, American Indian, and multiple race girls than for other race X gender groups or (3b) Black, Hispanic, American Indian, and multiple race boys than for other race X gender groups. We did not specify a direction for the gender effect because past research was consistent with both possibilities.

Methods

Procedures

Math teachers at 20 middle schools (For school characteristics, see Table S4) in a large and diverse school district in Florida were recruited to participate in the randomized placebo-controlled field experiment. Teachers were told the purpose of the modules was to collect their perspectives on various aspects of their profession. In Fall 2017, teachers were offered an opportunity to complete a 45-minute online module embedded in a Qualtrics survey. Though they were also offered an opportunity to complete a second 25-minute booster module in the Spring semester, only six teachers did so.

Teachers who met our eligibility criteria were randomly assigned to condition immediately after consent and the randomization was stratified by school (to try to ensure 50% of teachers from each school would be assigned to treatment versus control). The intervention and control materials were identical to previous research with the

empathic-discipline intervention (22). Both conditions included narratives, stories from other teachers and from students, and personal reflection exercises. Further, both were represented as honorific rather than as remedial—as an effort to learn from experienced teachers so as to help future teachers in teacher training programs. In the treatment condition, the materials asked teachers how they valued students’ perspectives and maintained high-quality relationships with students, especially when they misbehave. In the control condition, the materials asked teachers how they used technological tools to keep students engaged with lessons.

District records identified 200 teachers who taught at least one math class to 6th, 7th, or 8th graders across these schools. Our focus was on reduction in the probability of suspension for 7th and 8th grade students, students who had previous middle-school disciplinary interactions with middle-school teachers and staff for whom we could appropriately control for previous suspension rates in middle school. Moreover 5th-grade suspension rates were unavailable. Therefore, we removed teachers who primarily or exclusively taught 6th grade math classes as well as any students whose grade level during the intervention year was 6th grade from the final student sample. We defined math teacher as a teacher who taught a math course with at least five students. A total of 175 teachers, who primarily taught 7th and 8th grade math, were eligible. **Analyses featured 173 of these 175 teachers because two such teachers were only teachers for students who had another teacher assigned to condition. Analyses removed data for non-assigned teachers for students who had at least one assigned teacher.** Collectively, these 175 teachers taught 13,210 7th and 8th grade students across the 20 middle schools. The 109 teachers began the survey containing the intervention materials, 90 did so in the appropriate Fall implementation period, and 66 reached the point of random assignment and were randomly assigned to condition ($N=36$ control; $N=30$ treatment). **19 teachers (109 – 90) used the second module opportunity offered in Spring 2018 to instead participate in the session 1 module. Because these teachers were randomly assigned after Fall 2017, they were excluded from ITT analyses and coded as “no-assignment,” along with teachers who never started either module, in analyses comparing students of treatment and control teachers to students of non-assigned teachers.** For the CONSORT diagram, see Figure S1.

In intent-to-treat (ITT) analyses that sought to compare suspension outcomes for students of treatment versus control teachers, any math courses taught by teachers who did not start the survey, started the survey but did not reach the point of random assignment, or who were randomly assigned after Fall 2017 were excluded. Thus the ITT sample contained only student – math course – teacher – school rows in which the teacher had a Fall 2017 randomized condition assignment. In analyses that sought to compare suspension outcomes for students of treatment versus control teachers to other math teachers at the participating schools ($N=107$), a third “no-assignment” condition group was created for teachers who taught math to 7th and/or 8th graders but did not receive a randomized condition assignment in Fall 2017. As they elected not to participate in the study or failed to comply with study instructions to begin it in Fall 2017, these non-assigned teachers differed in several ways from assigned teachers (see Table S8). With the exception of the main effect of condition, all analyses focus on students of the 66 teachers who were randomized to condition in Fall 2017. There were no exclusions based on engagement with or completion of materials.

The school district provided discipline records, course records, and demographic records for all students in all district middle schools for the years before (AY 2016-2017), during, and after (AY 2018-2019) the intervention, including the 13,210 seventh and eighth grade students these teachers taught. Albeit, there was occasional missing data due to clerical issues. Before arriving at the final dataset for analysis, we excluded 481 student - math course – teacher – school rows in which the number of students with that teacher-school combination was < 5 (affected teachers were otherwise retained in rows where the teacher-school group had at least five students), which was likely an irregularity in the district data.

Analytical Approach

All analyses were conducted in STATA Version 16.1 using multilevel linear probability random effects models (implemented by the mixed regression command) in which students are nested within math course title – teacher groups (e.g., classrooms) within teachers within schools.

The goal was to understand the effect of treatment on student yearlong probability of suspension when both teachers and students were exposed to condition in semester 1, the Fall semester. A student was considered to be in the treatment condition if the student had at least one math course in Fall 2017 in which the math teacher was randomly assigned to the treatment during the implementation period, Fall 2017 (both student and teacher(s) were exposed to treatment in Fall 2017). A student was considered to be in the control condition if the student had no math courses in Fall 2017 in which the math teacher was randomly assigned to the treatment in Fall 2017 and at least one math course in Fall 2017 in which the math teacher was randomly assigned to the control in Fall 2017 (both student and teacher(s) were exposed to control but not treatment in Fall 2017).

In the intent-to-treat sample, $N=5,822$; hereafter the intent-to-treat (ITT) sample, where the goal was to compare only students in treatment versus control conditions, any math course-teacher-school rows that did not include a fall-semester math course with a fall-semester assigned teacher were excluded from analysis. In a larger sample ($N=13,210$; hereafter the ALL sample), where the goal was to increase power for the focal comparison and compare grade 7 and 8 students in both randomly assigned conditions to students of all other math teachers at the participating schools, additional students were included as a “non-assigned” condition (cases where the student and none of the student’s teachers was exposed to treatment or control in Fall 2017). Non-assigned students could include 1) students whose Fall-2017 math courses had no math teachers assigned to condition in Fall 2017 (because these teachers never started the study materials, because these teachers never reached the point of random assignment, or because they started the materials in Spring) 2) students who had no Fall-2017 math courses (only Spring-2018 math courses) or 3) students whose first exposure to an assigned math teacher(s) was in the spring semester (only their spring semester math courses were taught by assigned math teachers). The manuscript text reports only the main effect of condition in the larger sample including non-assigned students. All other results reported reflect results in the ITT sample.

The treatment condition served as the reference category in all models. The effect of condition was represented by a single contrast in treatment-versus-control-only models (1=Control; 0=Treatment) and by two contrasts in treatment versus control and non-assigned models (1=Control; 0=Otherwise; 1=Non-Assigned; 0=Otherwise). All models controlled for student-level dichotomous race-ethnicity, gender, and suspension status in the previous school year (0=No – Student received no days on suspension in the 2016-2017 school-year, 1=Yes – Student received at least one day on suspension in the 2016-2017 school-year). Additionally, we controlled for continuous average course-level (district wide, averaging across schools that offered the math course and all district teachers who taught the math course) and teacher-level (averaging across all students taught by the teacher, in any course) suspension rates in the previous school-year (see below). When students had multiple math courses (because they had multiple courses with the same teacher or multiple math teachers), we computed an average previous suspension rate across the student’s courses for the course-level and teacher-level previous suspension rate, respectively. We used three separate missing value indicators to indicate whether students were missing data (0 = No, 1 = Yes), one for each of the three previous suspension covariates (student-level, course-level, and teacher-level).

In all models (models excluding or including non-assigned students), we mean-centered course-level and teacher-level prior suspensions on the course-level mean and teacher mean, respectively, for math courses and teachers in the ITT sample so the intercept in the model would represent the treatment-condition mean for a student in a course with average levels of previous suspensions relative to all ITT-sample math courses and a teacher with average levels of previous suspensions relative to all ITT-sample math teachers. Consistent with prior research (35), we mean-centered student-level prior suspensions on the mean for a student’s race-ethnicity $\times$ gender group, so the intercept in the model would represent the treatment-condition mean for the average student in a given race-ethnicity $\times$ gender group, given that suspension rates differ across such groups. When they functioned primarily as a covariate in the model, as in analyses assessing the main effect of condition, the remaining dichotomous indicators (student race-ethnicity, gender, and missing value indicators) were mean-centered on the student-level mean for that variable. When the goal was to test an interaction between a specific dichotomous variable or variables (e.g., student race-ethnicity, gender, or previous suspension status) and condition, the relevant dichotomous indicator was left as a 0/1 variable and allowed to interact with each condition contrast (while all other dichotomous indicators were student-mean-centered). The analysis assessing the interaction between condition and student special education status (0=No disability; Yes=At least one disability) added this 0/1 predictor and its interaction with condition to a model in which all other student-level predictors (race-ethnicity, gender, and previous suspension status) were student-mean-centered. We computed simple effects of condition using the margins command. These reflect the control versus treatment difference (or none versus treatment difference) in the probability of suspension when the student group of interest was defined as 0 for the moderator and had average values for all other covariates.

Measures and coding methods

Dependent variable. Consistent with previous research (23), the primary dependent variable was whether a student received at least one suspension day (=1) versus not (=0) as a result of a discipline referral from any teacher or staff member at any time during the 2017-2018 school year, the year of intervention. For students who had school records at district schools in both 2017-2018 and 2018-2019, we also assessed whether a student received at least one suspension day (=1) versus not (=0) during the school year immediately following the intervention year (2018-2019).

Student-level condition. Though the vast majority of students in the ALL (N=13,210) sample had one math teacher (86.3%) during the intervention year, some (13.7%) students had more than one. Most of these multiple math teachers represented multiple non-assigned math teachers. **Only 1.8% of students in the 13,210 sample had multiple randomly assigned math teachers and only 0.9% had math teachers assigned to both treatment and control conditions.** In the sample that focused only on math courses taken in Fall 2017 in which the math teacher was assigned to (the treatment or control) condition that same semester, 96.2% of students had one math teacher. To accommodate the few remaining students who still had multiple assigned math teachers (3.8%), we defined a student as in the treatment condition if they had at least one Fall-2017 math teacher who was randomly assigned to the empathic mindset intervention in Fall 2017. We defined a student as in the control condition if they had at least one Fall-2017 math teacher who was randomly assigned to the control condition in Fall 2017 but no Fall-2017 teachers assigned to the treatment condition that semester. We grouped all other students into a “no-assignment category”: these students did not have a Fall-2017 math teacher who received a random assignment to condition in Fall 2017.

Previous suspensions at the student-, course-, and teacher-level covariates. To control for previous student-level suspensions, we used a single contrast to distinguish students who received no suspension days in the prior academic year (=0) from students who received one or more suspension days (=1). Because this district offers a common math curriculum across schools and students in specific types of math classes may have different propensity for receiving suspensions, independent of teacher, we also controlled for the average number of suspensions associated with each math course title across schools in the prior school year. We further controlled for the average suspension rate of students in the prior school year who had the student’s same math teacher during the previous school year.

Student demographic covariates. Students belonged to one of six racial-ethnic groups: White, Asian, Black, Hispanic, American Indian, Mixed (two or more races). These racial-ethnic groups differed in both proportional representation and average previous suspension rates in our sample in ways that are consistent with national data. When models simply sought to control for race-ethnicity, we used two contrasts: Black and Hispanic (=1) versus White (=0) and all other groups (=1) versus White (=0). Here Other indicates Asian, American Indian, or Students with Two or More Races. This approach most closely reflected the proportional representation of racial-ethnic groups in our sample, while also preserving some distinctions between groups in ways that partially reflected prior group discipline patterns. **In the 13,210 sample, both Black and Hispanic students had similar proportional representation (16.8% and 17.6% respectively) while all other minority groups each had $\leq 5.3\%$ representation. This was also true in the 5,822 ITT sample (16.5% Black, 14.7% Hispanic, all other minority groups $\leq 6.2\%$ representation respectively).** This coding scheme grouped together the two largest minoritized groups in our sample that tend to face disadvantage in disciplinary contexts (Black and Hispanic). Black and Hispanic students also had significantly higher suspensions the year prior to the intervention than White students (See Figure S3). The reference category was White students, which was the most numerous racial-ethnic group in the sample (55.7% of the ALL sample and 57.8% of the ITT sample; See Figure S2). Since this model best reflects the nature of this particular sample, we use this racial coding as a mean-centered covariate for race-ethnicity. Gender was coded as female (=1) versus male (=0) and also mean-centered.

Effects by student race-ethnicity and gender. To increase power for tests of interactions between condition and student race-ethnicity, we converted these two dichotomous race-ethnicity variables into a single race-ethnicity contrast that distinguished Black or Hispanic (=1) from not Black or Hispanic students (=0). This coding scheme was more parsimonious and required only one interaction term with condition yet still grouped together the two largest minoritized groups in our sample, and in national rates, that tend to face disadvantage in disciplinary contexts (Black and Hispanic students). We also report results based on other coding schemes (including the pre-registered Black, Hispanic, American Indian, and Two or More Races vs White or Asian) in these Supplemental Materials (Tables S14-15).

Effects by student suspension history. We first tested whether condition effects varied as a function of receiving one or more 2016-2017 discipline referrals that resulted in suspension day(s) (=1) versus no (=0) 2016-2017 discipline referrals that resulted in suspension days. This required a single predictor (1=One or More Prior Suspensions; 0=No Prior Suspensions) interacted with condition. Second, we disaggregated the former category into receiving exactly one discipline referral that resulted in a suspension day(s) versus receiving two or more discipline referrals that resulted in a suspension day(s). This created three categories: students with no, one, and two or more prior suspensions. We tested these as two contrasts, with no suspensions serving as the reference category and each of the other two coded as 1 (One Prior Suspension = 1; Otherwise = 0; Two or More Prior Suspensions = 1; Otherwise = 0).

Effects by student disability. We used a dichotomous variable for whether students had any disability as listed in school records (0=No – Student had no disabilities; 1=Yes – Student had at least one disability). These disabilities included one or more of 17 disability statuses defined by the district, including physical, academic, social, or emotional behavioral disabilities.

Tables

Table S1. *Proportions representing the relationships among the three moderators for probability of suspension: students' race-ethnicity, prior-year suspension status, and special-education status.*

	Proportion of Sample (N=5822)	Proportion of Students Suspended in Prior Year (N=870)	Proportion of Students in Special Education in Intervention Year (N=349)
Black	16.5%	42.8%	30.7%
Hispanic	14.7%	14.0%	16.9%
American Indian	0.3%	0.2%	0.3%
Two or More Races	4.5%	4.1%	4.3%
Asian	6.2%	2.1%	2.3%
White	57.8%	36.8%	45.6%
Total	100.0%	100.0%	100.0%

Note. The table presents the proportion of the students with teachers assigned to condition in the ITT sample (N=5,822) in each racial-ethnic group, as well as the proportion of each racial-ethnic group who received at least one suspension in the pre-intervention school year (AY 2016-2017) and the proportion of each racial-ethnic group with any special education status in the intervention school-year (AY 2017-2018). In this sample, prior discipline status was available for 5,533 students (15.7% or N=870 received at least one suspension in the prior school year) and special education status was available for 5,821 students (6.0% or N=349 had one or more disabilities in the intervention year).

Table S2. *Regression coefficients representing the relationships among the three moderators for probability of suspension*

	Coef.	Std. Err.	t	p	[95% CI]	
Prior Suspensions → Black or Hispanic student	0.309	0.017	18.692	0.000	0.277	0.341
Special Education Status → Black or Hispanic student	0.174	0.025	6.816	0.000	0.124	0.224
Special Education Status → Prior Suspensions	0.218	0.021	10.505	0.000	0.177	0.259

Note. The table presents a comparison of linear regression coefficients for prior-year (AY 2016-2017) suspension status (1=Received at least one discipline referral resulting in suspension days; 0=otherwise) predicting student's race-ethnicity is Black or Hispanic, intervention-year (AY 2017-2018) special-education status predicting student's race-ethnicity is Black or Hispanic, and intervention-year (AY 2017-2018) special-education status predicting prior-year suspension status in the ITT sample (N=5,822).

Table S3. *Covariate balance between student condition in the ITT sample, overall and for one versus multiple ITT math teachers*

	Means			Control vs Treatment Contrast		
	ITT Sample	Treatment	Control	ITT Sample	One ITT Math Teacher	Multiple ITT Math Teachers
Student Race is Black/Hispanic vs White	0.312	0.298	0.323	0.024*	0.022	0.079
Student Race is Other vs White	0.109	0.106	0.111	0.005	0.005	0.009
Student Gender is Female	0.494	0.485	0.500	0.015	0.021	-0.150*
Student Suspended in Previous Year	0.157	0.160	0.156	-0.004	-0.010	0.144*
Student Has Special Education Status	0.060	0.064	0.057	-0.008	-0.011	0.070
Student Has Multiple Math Teachers	0.038	0.037	0.039	0.002	--	--
Sample Size	5822	2470	3352	5822	5600	222

Note. The table presents the balance for key covariates for students classified as being in the treatment or control condition in the ITT sample ($N=5,822$). In this sample, 5600 students ($n_{\text{control}}=3,222$, $n_{\text{treatment}}=2,378$) had one Fall-semester math teacher randomly assigned to condition that semester; 222 students ($n_{\text{control}}=130$, $n_{\text{treatment}}=92$) had multiple Fall-semester math teachers randomly assigned to condition that semester (though some of these were still assigned to the same condition). Contrasts are the difference in means or proportions. Significance is based on logistic regression coefficients (all measures were dichotomous). *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$.

Table S4. *For students with multiple math teacher assigned to condition ($n=222$; $n_{\text{control}}=130$, $n_{\text{treatment}}=92$), the relationships among the student demographics (students' race-ethnicity, gender, prior-year suspension status, special-education status, multiple teachers or not) by condition.*

Var Desc	Students with Multiple ITT Math Teachers			Control vs Treatment
	ALL	Treatment	Control	
Student Race is Black/Hispanic vs White	0.514	0.467	0.546	0.079
Student Race is Other vs White	0.081	0.076	0.085	0.009
Student Gender is Female	0.477	0.565	0.415	-0.150*
Student Suspended in Previous Year	0.327	0.244	0.388	0.144*
Student Has Special Education Status	0.117	0.076	0.146	0.070
Sample Size	222	92	130	

Note. Contrasts are the difference in means or proportions. Significance is based on regression coefficients (logistic for dichotomous measures). *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$.

Table S5. District-Provided Ethnicity Coded by Census Definitions, by Condition/Group, for All District Middle School Teachers

	Math Teachers				All Teachers		
	Study Teachers			Non-Eligible Teachers	All Math Teachers (N=200)	Non-Math Teachers (N=459)	All Teachers (N=659)
	Treatment (N=30)	Control (N=36)	No Assignment (N=107)	Exclude (N=27)			
Hispanic	0.0%	2.8%	6.5%	7.4%	5.0%	4.4%	4.6%
Black	6.7%	8.3%	5.6%	11.1%	7.0%	7.4%	7.3%
White	80.0%	77.8%	60.7%	70.4%	68.0%	66.2%	66.8%
Asian	6.7%	2.8%	0.9%	7.4%	3.0%	0.9%	1.5%
American Indian	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	0.3%
Unknown	6.7%	8.3%	26.2%	3.7%	17.0%	20.7%	19.6%
All Races	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Note. We coded all district-provided race-ethnicities with which a teacher identified according to census definitions. We first coded teachers who identified as Hispanic (Hispanic) for at least one ethnicity. We then coded teachers who identified as Black, not Hispanic; followed by White, not Hispanic or Black; followed by Asian, not White, Hispanic, or Black; followed by American Indian, not Black, Hispanic, White, or Asian. “Exclude” teachers taught only or primarily 6th grade classes (see Figure S1). Percentages reflect teachers with available race-ethnicity data for any of the four academic years provided by the district, prioritizing the race-ethnicity reported in the intervention year (2017-2018) if available. The sample size below each group represents all teachers in that group, whether or not (coded unknown) they had available data.

Table S6. Math Teacher Demographic Variables, by All Teachers, Math Teachers, and Study Math Teachers

	Teachers in AY 2017-2018		
	Study Teachers (N=173)	All Math Teachers (N=200)	All Teachers (N=659)
Missing in District Demographic Records	19.1%	17.0%	19.6%
Missing Student-Level Discipline Data in Previous Year	26.6%	26.5%	22.6%
Census Ethnicity is White	83.6%	81.9%	83.0%
Gender is Female	67.9%	68.7%	76.2%
Age in Intervention Year	41.85 (11.02)	42.01 (11.01)	44.01 (11.73)
Years Teaching Experience as of Intervention Year	9.14 (7.95)	9.34 (7.89)	10.19 (8.59)
Average Suspensions Received by Teacher's Students in Previous Year	0.22 (0.15)	0.22 (0.15)	0.22 (0.16)

Table S7. Math Teacher Covariate Balance, by Randomized Condition, for Study Math Teachers

	Group Means		Condition Difference
	Treatment (N=30)	Control (N=36)	Control vs Treatment
Census Ethnicity is White ^D	85.7%	84.8%	0.93
Gender is Female ^D	82.1%	75.8%	0.68
Age in AY 2017-2018	44.61	44.82	0.21
Years Teaching Experience in AY 2017-2018	10.57	11.42	0.85
Average Suspensions Received by Teacher's Students in Previous Year	0.18	0.14	-0.04
Total distinct math course titles taught	2.10	2.22	0.12
Taught any Grade 6 math course ^D	33.3%	25.0%	0.67
Taught any Grade 7 math course ^D	66.7%	69.4%	1.14
Taught any Grade 8 math course ^D	53.3%	52.8%	0.98
Taught any remedial math course ^D	40.0%	47.2%	1.34
Taught any regular math course ^D	46.7%	47.2%	1.02
Taught any advanced math course ^D	76.7%	86.1%	1.89

Note. The table reports the condition means for teachers randomly assigned to the treatment or control condition in Fall 2017 and their difference for demographic variables (available from district teacher records), suspension-related variables (computed from student discipline records), and grade levels and kinds of math courses taught (available from district course records). The average suspension rate represents the average number of suspensions received by the teacher's students (a student who had that teacher at any point during the academic year) at any time during the pre-intervention academic year (the suspension could be issued by any staff member). Linear regression (logistic regression) was used for continuous (dichotomous, indicated by superscript D) indicators. Logistic regression coefficients are represented as odds ratios, the ratio of the odds of having a value of 1 versus 0 on the indicator in the control (none) versus the treatment condition. Demographic indicators and prior suspension rates could not be computed for teachers for whom these records were unavailable. The sample size below each group represents all teachers in that group, whether or not they had available data for a given covariate. *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$

Table S8. Comparison of Assigned and Non-Assigned Study Math Teachers

	Group Means		Condition Difference
	Assigned (N=66)	Non-Assigned (N=107)	Assigned vs Non-Assigned
Census Ethnicity is White	85.2%	82.3%	1.24
Gender is Female	78.7%	59.5%	2.51*
Age in AY 2017-2018	44.72	39.63	5.09**
Years Teaching Experience in AY 2017-2018	11.03	7.67	3.36**

Average Suspensions Received by Teacher's Students in Previous Year	0.16	0.27	-0.11***
Total distinct math course titles taught	2.17	2.17	0.00
Taught any Grade 6 math course	28.8%	34.6%	0.76
Taught any Grade 7 math course	68.2%	72.9%	0.80
Taught any Grade 8 math course	53.0%	52.3%	1.03
Taught any remedial math course	43.9%	64.5%	0.43**
Taught any regular math course	47.0%	67.3%	0.43**
Taught any advanced math course	81.8%	51.4%	4.25***

Note. This table compares assigned (teachers who were randomized to condition in Fall 2017) versus non-assigned teachers (teachers who elected not to participate in the study or who were randomized to condition in Spring 2018). Survey variables (childhood hometown, childhood SES, and psychometrics) were unavailable for the vast majority of non-assigned teachers. *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$

Table S9. *Comparison of participating schools on key indicators.*

School	Total School Enrollment	Students Suspended Previous Year (%)	Students Suspended Intervention Year (%)	Racial Composition of School (%)				
				Asian	Black	Hispanic	White	Other
1	3278	19.7%	18.4%	5.2%	14.1%	19.8%	56.1%	4.8%
2	3067	21.4%	21.5%	4.0%	11.3%	16.0%	64.2%	4.6%
3	2506	20.6%	31.3%	9.1%	18.1%	17.3%	50.8%	4.7%
4	2227	21.5%	24.5%	4.4%	13.2%	15.2%	62.0%	5.3%
5	2218	20.3%	21.2%	3.7%	16.0%	12.5%	62.9%	4.9%
6	2083	27.5%	32.0%	4.5%	38.8%	11.9%	40.2%	4.6%
7	1974	24.4%	27.2%	2.4%	14.7%	28.5%	50.0%	4.5%
8	1899	17.9%	18.0%	4.7%	13.0%	15.9%	61.5%	4.8%
9	1791	25.8%	28.5%	6.0%	21.0%	15.8%	52.7%	4.5%
10	1784	24.4%	29.8%	5.3%	20.3%	21.1%	48.0%	5.1%
11	1750	30.1%	37.1%	3.2%	34.4%	15.8%	41.6%	5.1%
12	1739	22.0%	24.3%	8.1%	15.2%	22.0%	50.2%	4.4%
13	1691	15.2%	12.5%	4.4%	9.3%	15.0%	67.7%	3.5%
14	1563	14.3%	14.1%	5.7%	27.5%	10.7%	51.3%	4.8%
15	1436	25.8%	32.9%	7.0%	20.1%	17.7%	49.6%	5.7%
16	1387	31.3%	33.2%	3.5%	41.2%	12.4%	37.7%	5.2%
17	1345	23.7%	27.3%	4.6%	18.8%	24.3%	46.6%	5.6%
18	1229	14.0%	9.2%	3.2%	9.2%	17.0%	65.9%	4.7%
19	1161	24.9%	25.8%	3.5%	12.1%	16.1%	63.5%	4.7%
20	592	12.3%	12.1%	4.5%	8.3%	10.9%	72.2%	4.1%

Note. Other consists of American Indian and multiple-race students.

Main Text Analyses Tables (pre-registered)

All models predict the probability of receiving at least one suspension day in the intervention year unless otherwise indicated. All models controlled for the following covariates (not listed for simplicity): mean-centered student race-ethnicity, mean-centered student gender, whether or not the student received a suspension in the prior year (student-level), average suspension rate for students taking same math course in the prior year district-wide (course-level), average suspension rate among students who had same math teacher the prior year (teacher-level), and missing value indicators for prior suspensions data at the student-, course-, and teacher-level. All effects tables present the intercept (the treatment-condition mean for a student in the relevant reference category) and the condition difference (control or none vs treatment) from this mean from the corresponding model. When student race-ethnicity, gender, or prior suspension status is the focal moderator, it is coded as 0/1 rather than mean-centered.

Table S10. Key Regression Coefficients from Model of Main Effect of Control v. Treatment Effect in ITT Sample

At Least One Suspension Day	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.024	0.016	1.512	0.130	-0.007	0.055
Intercept (Treatment Mean)	0.168	0.016	10.270	0.000	0.136	0.200

At Least One In-School-Suspension	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.048	0.014	3.289	0.001	0.019	0.076
Intercept (Treatment Mean)	0.094	0.014	6.898	0.000	0.067	0.121

At Least One Out-of-School Suspension	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.008	0.013	0.593	0.553	-0.018	0.033
Intercept (Treatment Mean)	0.093	0.011	8.754	0.000	0.073	0.114

Note. N=5,822 students, with 66 different intervention-year teachers.

Table S11. Key Regression Coefficients from Model of Main Effect of Control or No Assignment v. Treatment Effect in ALL Sample

	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.031	0.016	1.955	0.051	0.000	0.061
No Assignment v. Treatment	0.053	0.014	3.760	0.000	0.025	0.080
Intercept (Treatment Mean)	0.173	0.016	10.865	0.000	0.142	0.205

Note. N=13,210 students, with 173 different intervention-year teachers.

Table S12. Key Regression Coefficients from Model testing the interaction between Control v. Treatment and student race-ethnicity is whether or not Black or Hispanic in the ITT Sample

	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.009	0.017	0.548	0.584	-0.024	0.042
Student Race is Black or Hispanic	0.059	0.014	4.137	0.000	0.031	0.087
Student Race x Control v. Treatment	0.047	0.019	2.549	0.011	0.011	0.083
Intercept (Non-Black-Hispanic Treatment Mean)	0.149	0.017	8.900	0.000	0.117	0.182

Note. N=5,822 students, with 66 different intervention-year teachers. Student race-ethnicity is a dichotomous variable defined as 1 if race-ethnicity is Black or Hispanic and 0 if race-ethnicity is not Black or Hispanic.

Table S13. Simple Effects of Control v. Treatment from Model testing the interaction between Control v. Treatment and student race-ethnicity is whether or not Black or Hispanic in the ITT Sample

	Coef.	Std.Err.	z	p	[95% CI]	
Not Black or Hispanic	0.009	0.017	0.548	0.584	-0.024	0.042
Black or Hispanic	0.056	0.020	2.787	0.005	0.017	0.096

Note. N=5,822 students, with 66 different intervention-year teachers. Simple effects were computed using the model shown in Table S12.

Table S14. Simple Effects of Control v. Treatment from Separate Models testing the interaction between Control v. Treatment and student race-ethnicity is Black (Hispanic) in the ITT Sample

	Coef.	Std.Err.	z	p	[95% CI]	
Not Black	0.016	0.016	1.037	0.300	-0.015	0.047
Black	0.057	0.025	2.235	0.025	0.007	0.106
Not Hispanic	0.016	0.016	0.971	0.332	-0.016	0.047
Hispanic	0.054	0.025	2.143	0.032	0.005	0.103

Note. N=5,822 students, with 66 different intervention-year teachers. In the top two rows, student race-ethnicity is a dichotomous variable defined as 1 if race-ethnicity is Black and 0 if race-ethnicity is not Black. In the bottom two rows, student race-ethnicity is a dichotomous variable defined as 1 if race-ethnicity is Hispanic and 0 if race-ethnicity is not Hispanic.

Table S15. Key Regression Coefficients from Model testing the interaction between Control v. Treatment and student race-ethnicity is whether or not Black or Hispanic or American Indian or Two or More Races in the ITT Sample

	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.010	0.017	0.560	0.575	-0.024	0.043
Student Race is Black or Hispanic or Other	0.069	0.014	5.032	0.000	0.042	0.096

Student Race x Control v. Treatment	0.039	0.018	2.163	0.031	0.004	0.074
Intercept (White or Asian Treatment Mean)	0.143	0.017	8.423	0.000	0.110	0.176

Note. $N=5,822$ students, with 66 different intervention-year teachers. Student race-ethnicity is a dichotomous variable defined as 1 if race-ethnicity is Black or Hispanic or American Indian or Two or More Races and 0 if race-ethnicity is not White or Asian.

Table S16. *Key Regression Coefficients from Model testing the interaction between Control v. Treatment and student race is Black or not in the ITT Sample*

	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.016	0.016	1.037	0.300	-0.015	0.047
Student Race is Black	0.167	0.019	8.937	0.000	0.130	0.204
Student Race x Control v. Treatment	0.040	0.024	1.666	0.096	-0.007	0.088
Intercept (Non-Black Treatment Mean)	0.140	0.017	8.454	0.000	0.108	0.173

Note. $N=5,822$ students, with 66 different intervention-year teachers. Student race is a dichotomous variable defined as 1 if race is Black and 0 if race is non-Black. The simple effects of condition for this model are presented in the top two rows in Table S14.

Table S17. *Key Regression Coefficients from Model testing the interaction between Control v. Treatment and student race is Black or White in the ITT Sample*

	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.006	0.017	0.357	0.721	-0.027	0.039
Student Race is Black v. White	0.157	0.019	8.131	0.000	0.119	0.195
Student Race x Control v. Treatment	0.051	0.025	2.043	0.041	0.002	0.100
Intercept (Non-Black Treatment Mean)	0.149	0.017	8.630	0.000	0.115	0.182

Note. $N=5,822$ students, with 66 different intervention-year teachers. Student race is a dichotomous variable defined as 1 if race is Black and 0 if race is White.

Table S18. *Key Regression Coefficients from Model testing the interaction between Control v. Treatment and student race is Hispanic or White in the ITT Sample*

	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.006	0.017	0.369	0.712	-0.027	0.040
Student Race is Hispanic v. White	-0.040	0.018	-2.175	0.030	-0.075	-0.004
Student Race x Control v. Treatment	0.048	0.024	2.037	0.042	0.002	0.095
Intercept (Non-Black Treatment Mean)	0.152	0.017	8.939	0.000	0.118	0.185

Note. $N=5,822$ students, with 66 different intervention-year teachers. Student race is a dichotomous variable defined as 1 if race is Hispanic and 0 if race is White.

Additional Gender Analyses Tables

Table S19. Key Regression Coefficients from Model testing the interaction between Control v. Treatment and student gender is Female or not in the ITT Sample (non-preregistered)

	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.018	0.018	1.020	0.308	-0.017	0.053
Student Gender is Female	-0.094	0.012	-7.828	0.000	-0.118	-0.071
Student Gender x Control v. Treatment	0.012	0.016	0.749	0.454	-0.019	0.043
Intercept (Male Mean)	0.215	0.017	12.295	0.000	0.181	0.249

Note. N=5,822 students, with 66 different intervention-year teachers. Student gender is a dichotomous variable defined as 1 if gender is Female and 0 if gender is male. This finding was not predicted; its reliability and meaning should be explored in future research.

Table S20. Key Regression Coefficients from Model testing the interactions among Control v. Treatment and student race-ethnicity is Black or Hispanic or not and student gender is Female or not in the ITT Sample (pre-registered).

	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.004	0.019	0.217	0.829	-0.034	0.042
Student Race is Black or Hispanic	0.035	0.019	1.855	0.064	-0.002	0.072
Student Gender is Female	-0.110	0.015	-7.527	0.000	-0.138	-0.081
Student Race x Gender	0.049	0.026	1.911	0.056	-0.001	0.100
Student Race x Control v. Treatment	0.047	0.025	1.908	0.056	-0.001	0.096
Student Gender x Control v. Treatment	0.011	0.019	0.581	0.561	-0.027	0.049
Student Race x Gender x Control v. Treatment	-0.001	0.034	-0.015	0.988	-0.067	0.066
Intercept (Not Black or Hispanic Male Mean)	0.203	0.018	11.108	0.000	0.167	0.239

Note. N=5,822 students, with 66 different intervention-year teachers. Student race-ethnicity is a dichotomous variable defined as 1 if race-ethnicity is Black or Hispanic and 0 if race-ethnicity is not Black or Hispanic. Student gender is a dichotomous variable defined as 1 if gender is Female and 0 if gender is male.

Table S21. Simple effects of condition from Model testing the interactions among Control v. Treatment and student race-ethnicity is Black or Hispanic or not and student gender is Female or not in the ITT Sample

	Coef.	Std.Err.	z	p	[95% CI]	
Not Black or Hispanic Males	0.004	0.019	0.217	0.829	-0.034	0.042
Black or Hispanic Males	0.052	0.024	2.109	0.035	0.004	0.099
Not Black or Hispanic Females	0.016	0.020	0.794	0.427	-0.023	0.054
Black or Hispanic Females	0.062	0.025	2.498	0.012	0.013	0.111

Note. N=5,822 students, with 66 different intervention-year teachers. Simple effects were computed using the model shown in Table S18.

Main Text Analyses Tables: Primary outcomes (not pre-registered)

Table S22. Key Regression Coefficients from Model testing the interaction between Control v. Treatment and dichotomous student history of suspension in the ITT sample

	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.015	0.016	0.930	0.352	-0.017	0.047
Student Was Previously Suspended	0.390	0.017	22.752	0.000	0.356	0.423
Student Suspension History x Control v. Treatment	0.050	0.023	2.195	0.028	0.005	0.095
Intercept (Mean for Students with Zero Prior Suspensions)	0.106	0.016	6.523	0.000	0.074	0.138

Note. N=5,533 students, with 65 different intervention-year teachers. Student history of suspension is defined as 1 (v. 0) if student received one or more (versus zero) discipline referrals in the previous year (AY 2016-2017) resulting in suspension days, for students in the ITT Sample with available data.

Table S23. Simple effects of condition from Model testing the interaction between Control v. Treatment and dichotomous student history of suspension in the ITT sample

	Coef.	Std.Err.	z	p	[95% CI]	
No Suspensions in Previous Year	0.015	0.016	0.930	0.352	-0.017	0.047
One or More Suspensions in Previous Year	0.065	0.024	2.669	0.008	0.017	0.113

Note. N=5,533 students, with 65 different intervention-year teachers. Simple effects were computed using the model shown in Table S20.

Table S24. Key Regression Coefficients from Model testing the interaction between Control v. Treatment and two types of suspension history

	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.013	0.015	0.893	0.372	-0.016	0.043
Prior Suspension (1 v. 0)	0.204	0.022	9.092	0.000	0.160	0.248
Prior Suspension (2 or more v. 0)	0.572	0.022	26.122	0.000	0.529	0.615
Prior Suspension (1 v. 0) x Control v. Treatment	0.089	0.031	2.924	0.003	0.029	0.149
Prior Suspension (2 v. 0) x Control v. Treatment	0.003	0.029	0.088	0.930	-0.054	0.059
Intercept (Mean for Students with Zero Prior Suspensions)	0.103	0.015	6.735	0.000	0.073	0.133

Note. N=5,533 students, with 65 different intervention-year teachers. Student history of suspension is defined with two dichotomous variables (1=Student received one discipline referral in previous year resulting in suspension days; 0=Otherwise; 1=Student received two or more discipline referrals in previous year resulting in suspension days; 0=Otherwise), for students in the ITT Sample with available data. Students receiving zero suspension days in the previous year serve as the reference category.

Table S25. Simple Effects of Condition from Model testing the interaction between Control v. Treatment and two types of suspension history

	Coef.	Std.Err.	z	p	[95% CI]	
No Suspensions in Previous Year	0.013	0.015	0.893	0.372	-0.016	0.043

One Suspension in Previous Year	0.103	0.031	3.269	0.001	0.041	0.164
Two or More Suspensions in Previous Year	0.016	0.030	0.537	0.591	-0.042	0.074

Note. $N=5,533$ students, with 65 different intervention-year teachers. Simple effects were computed using the model shown in Table S12.

Table S26. Key Regression Coefficients from *Model testing the interaction between Control v. Treatment and student special education status*

	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.019	0.016	1.179	0.238	-0.012	0.049
Student Has Disability Status	-0.034	0.024	-1.388	0.165	-0.081	0.014
Student Disability Status x Control v. Treatment	0.082	0.033	2.452	0.014	0.016	0.147
Intercept (Mean for Students with No Disabilities)	0.170	0.016	10.477	0.000	0.138	0.202

Note. $N=5,822$ students, with 66 different intervention-year teachers. Student special education status was defined as 1 if a student had any disability in the intervention year (AY 2017-2018) and 0 otherwise.

Table S27. Simple effects of condition from *Model testing the interaction between Control v. Treatment and student special education status*

	Coef.	Std.Err.	z	p	[95% CI]	
No Disability Status	0.019	0.016	1.179	0.238	-0.012	0.049
Has Disability Status	0.100	0.035	2.872	0.004	0.032	0.168

Note. $N=5,822$ students, with 66 different intervention-year teachers. Simple effects were computed using the model shown in Table S24.

Main Text Analyses Tables: Subsequent year outcomes (not pre-registered)

Table S28. *Key Regression Coefficients from Model of Main Effect of Control v. Treatment for students who could be tracked in district schools for two years.*

Intervention Academic Year (AY 2017-2018)	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.035	0.020	1.724	0.085	-0.005	0.075
Intercept (Treatment Mean)	0.168	0.020	8.281	0.000	0.128	0.208

Subsequent Academic Year (AY 2018-2019)	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.042	0.020	2.056	0.040	0.002	0.081
Intercept (Treatment Mean)	0.167	0.016	10.280	0.000	0.135	0.199

Note. N=2,712 students, with 56 different intervention-year teachers. The model was fit for students with district school records in the intervention year and subsequent academic year, classified according to their 7th grade math teacher(s)' condition assignment during the intervention year. Nearly all of such students were 7th graders in the intervention year and 8th graders in the subsequent year.

Table S29. *Key Regression Coefficients from Model testing the interaction between Control v. Treatment and student race-ethnicity is Black or Hispanic for students who could be tracked in district schools for two years.*

Intervention Academic Year (AY 2017-2018)	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.019	0.021	0.878	0.380	-0.023	0.061
Student Race is Black or Hispanic	0.046	0.021	2.165	0.030	0.004	0.088
Student Race x Control v. Treatment	0.057	0.027	2.108	0.035	0.004	0.111
Intercept (Non-Black-Hispanic Treatment Mean)	0.151	0.021	7.282	0.000	0.111	0.192

Subsequent Academic Year (AY 2018-2019)	Coef.	Std.Err.	z	p	[95% CI]	
Control v. Treatment	0.025	0.021	1.171	0.241	-0.017	0.066
Student Race is Black or Hispanic	0.046	0.022	2.073	0.038	0.002	0.089
Student Race x Control v. Treatment	0.057	0.028	2.031	0.042	0.002	0.112
Intercept (Non-Black-Hispanic Treatment Mean)	0.151	0.017	8.957	0.000	0.118	0.184

Note. N=2,712 students, with 56 different intervention-year teachers. Student race-ethnicity is a dichotomous variable defined as 1 if race-ethnicity is Black or Hispanic and 0 if race-ethnicity is not Black or Hispanic. The model was fit for students with district school records in the intervention year and subsequent academic year, classified according to their 7th grade math teacher(s)' condition assignment during the intervention year. Nearly all of such students were 7th graders in the intervention year and 8th graders in the subsequent year.

Table S30. *Effects of treatment by student race-ethnicity (Black or Hispanic (=1) from Not Black or Hispanic (=0)) on intervention year suspension rate for teachers.*

Intervention Academic Year (AY 2017-2018)	Coef.	Std.Err.	z	p	[95% CI]	
Not Black or Hispanic	0.019	0.021	0.878	0.380	-0.023	0.061
Black or Hispanic	0.076	0.028	2.736	0.006	0.022	0.131

Subsequent Academic Year (AY 2018-2019)	Coef.	Std.Err.	z	p	[95% CI]	
Not Black or Hispanic	0.025	0.021	1.171	0.241	-0.017	0.066
Black or Hispanic	0.082	0.028	2.929	0.003	0.027	0.137

Note. $N=2,712$ students, with 56 different intervention-year teachers. Simple effects were computed using the models shown in Table S27.

Figures

Figure S1. *Teacher Inclusion Criteria and Condition Assignments*

Figure S2. Proportion of Students by Race-Ethnicity During Intervention Year

Note. Data are from the 5,822 students of teachers assigned to condition during the implementation period.

5 **Figure S3.** *Suspensions During the Year Prior to the Intervention Year as a Function of Race-Ethnicity in the*
6 *Intervention Year Sample*

7
8 *Note.* Data are from the 870 students of teachers assigned to condition during the implementation period who had
9 received a suspension the previous year.