PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

# The dynamic nature of student discipline and discipline disparities

Sean Darling-Hammond[a,1], Michael Ruiz[b], Jennifer L. Eberhardt[c,1], and Jason A. Okonofua[b,1]

Contributed by Jennifer L. Eberhardt; received November 8, 2021; accepted January 3, 2023; reviewed by Kenneth Shores and David S. Yeager

**Researchers have long used end-of-year discipline rates to identify punitive schools, explore sources of inequitable treatment, and evaluate interventions designed to stem both discipline and racial disparities in discipline. Yet, this approach leaves us with a "static view"—with no sense of how disciplinary responses fluctuate throughout the year. What if daily discipline rates, and daily discipline disparities, shift over the school year in ways that could inform when and where to intervene? This research takes a "dynamic view" of discipline. It leverages 4 years of atypically detailed data regarding the daily disciplinary experiences of 46,964 students from 61 middle schools in one of the nation's largest school districts. Reviewing these data, we find that discipline rates are indeed dynamic. For all student groups, the daily discipline rate grows from the beginning of the school year to the weeks leading up to the Thanksgiving break, falls before major breaks, and grows following major breaks. During periods of escalation, the daily discipline rate for Black students grows significantly faster than the rate for White students—widening racial disparities. Given this, districts hoping to stem discipline and disparities may benefit from timing interventions to precede these disciplinary spikes. In addition, early-year Black–White disparities can be used to identify the schools in which Black–White disparities are most likely to emerge by the end of the school year. Thus, the results reported here provide insights regarding not only when to intervene, but where to intervene to reduce discipline rates and disparities.**

## Significance

The present research uncovers the dynamic nature of student discipline. The findings show that discipline escalates during the school year, that discipline escalates more severely for Black students, and that racial disparities in discipline escalate most in schools that have a high degree of racial disparity early in the year. The findings thus provide insights regarding when and where to intervene to reduce discipline and discipline disparities.

school discipline | racial disparities | longitudinal data

Imagine two middle school students that have the same background, the same temperament, and the same class schedule. Now imagine they both have precisely one very "bad day" at school, but that the first student's bad day happens early in the year (on the first day of the Fall), while the second student's bad day happens later on (during the first week of November). Would they be treated the same way, or might the first student receive an informal warning while the second is saddled with a formal suspension? Now imagine two students, one Black and one White, who both misbehave on the same day. Might the Black student experience markedly worse treatment than the White student at certain points in the year? In other words, does the punitiveness of a school, or the inequity in how a school treats students of different backgrounds, vary over the year, or is it stable? Similarly, do students' levels of misbehavior, and disparities in misbehavior, stay constant, or are they dynamic? If punitiveness, misbehavior, and related disparities are not stable—if they are dynamic—how should school policy adapt?

Exclusionary discipline can harm students and society. Suspended students show a heightened risk of depression and civic disengagement; and causal evidence indicates that suspensions exacerbate misbehavior, harm academic performance, and lead to both juvenile and adult incarceration (1–5). Economic analyses indicate that suspensions cost taxpayers tens of millions of dollars each year due to lifelong government expenditures and losses in tax revenue (6). Black students face the brunt of these consequences. Black students have experienced higher discipline rates than White students since schools were first integrated (7) and are now at least two times more likely than White students to receive out-of-school suspensions among all student subpopulations (e.g., female students, poor students) and across all school contexts (e.g., in preschools, in charter schools) (8). Recent policy investments (9, 10) have achieved declines in overall suspension rates, but high discipline rates remain in certain school contexts, and racial disparities are both enduring and pervasive (11). Why might current approaches fall short? Are we missing something fundamental about where and when harmful disciplinary actions and disparities emerge, and how they can be ameliorated?

Perhaps. Research regarding discipline has implicitly adopted what could be termed a "static view." One version of the static view maintains that school features that are relatively stable over a typical school year (e.g., principal leadership style, district policy, school

Author affiliations: [a]Fielding School of Public Health, University of California, Los Angeles, CA 90095; [b]Department of Psychology, University of California, Berkeley, CA 94704; and [c]Graduate School of Business, Department of Psychology, Stanford University, Stanford, CA 94305

Author contributions: S.D.-H., M.R., J.L.E., and J.A.O. designed research; S.D.-H. and M.R. performed research; J.A.O. collected data; S.D.-H., J.L.E., and J.A.O. analyzed data; and S.D.-H., M.R., J.L.E., and J.A.O. wrote the paper.

Reviewers: K.S., University of Delaware; and D.S.Y., The University of Texas at Austin.

The authors declare no competing interest.



[1]To whom correspondence may be addressed. Email: seandh@ucla.edu, jleberhardt@stanford.edu, or okonofua@berkeley.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2120417120/-/DCSupplemental.

Published April 17, 2023.

policy, teacher composition, and student composition) drive discipline rates for Black and White students, and therefore drive discipline disparities (12–14). As such, the static view imagines that discipline rates may vary between districts and schools, and may vary by student characteristics within a school, but do not vary over short periods of time. From this vantage point, discipline rates (overall, and for any given racial group) are no different in August than they are in February. Put another way, the static view assumes that the punitiveness of a school towards all students or subgroups of students does not change over time in ways that could inform educational policies or practices.

The static view is also implied by the way discipline data are collected and disaggregated—data are collected at the end of the year, and disaggregated by school and racial group, but not by various points in the school year (7). The static view is implied by how we evaluate common policy responses to discipline and disparities, such as Positive Behavioral Interventions and Supports (PBIS) and Restorative Practices (RP), which are typically evaluated based on their ability to drive down end-of-year discipline rates and discipline disparities (12, 15–21). While it holds broad implicit support, the static view offers little in the way of guidance regarding where, when, or how to respond to high discipline rates, or to large discipline disparities. This is unfortunate as thousands of schools across the country are currently implementing strategies to reduce racial disparities in discipline (e.g., conducting widespread professional development) with no sense of when the best time to implement these strategies might be. In a key respect, enduring adherence to the static view has forced many school leaders to fly blind.

Here, we introduce a "dynamic view" of discipline and discipline disparities. The dynamic view argues that phenomena that vary over time can drive temporal variation in discipline rates within a school and over a school year. Research suggests at least two reasons that discipline rates might increase at the beginning of the year: teachers' responses to misbehavior may grow more severe as misbehavior persists, and student misbehavior may grow more frequent as a response to successive instances of punishment. A series of experiments have demonstrated that teachers escalate their disciplinary responses with successive instances of misbehavior, even over very short time frames (22, 23). This teacher-side escalation is more severe when teachers interact with Black students, suggesting that daily discipline disparities might also escalate over time. Research evaluating students' reactions to discipline also suggests that students who are disciplined can become defiant (24, 25), which could lead to increases in acts of misbehavior and discipline escalation. Indeed, a detailed propensity score matching analysis suggested that discipline led students to increase their rates of misbehavior (5). If, over time, students escalate in their misbehavior and/or teachers escalate in their punitiveness, discipline rates may grow as the school year progresses. In addition, if these escalatory phenomena are uniquely pernicious for Black students, then the Black–White discipline disparity may also grow over the year.

Discipline rates, and levels of disparity, may vary dynamically for other reasons. The school year features many breaks (e.g., Thanksgiving break, Winter break, Spring break, and Summer break) that may impact student and teacher moods in ways that could shift rates of misbehavior and discipline. Research (26) has demonstrated that youth experience lower levels of depression and anxiety when school breaks are proximate. In addition, research reviewing the experiences of on working adults has found similarly that stress and aggression are diminished near periods of vacation from work (27). Might proximity to school breaks drive lower levels of anxiety among students and teachers, and subsequently reduce misbehavior and discipline?

In public schools, the school year is also punctuated by many periods of high-stakes testing. Research has shown that students experience higher levels of sleep fragmentation during exam periods (28), that students experience higher levels of anxiety during periods of high-stakes testing (29), and that testing anxiety is especially high for Black students (30). Research has also found that teachers often express feeling stressed and worried about the prospect of preparing students for high-stakes testing and about how students will perform on these tests (which are used for teacher assessment) (31). Might the anxiety experienced during periods of intense examination lead to more student misbehavior and harsher disciplinary responses from teachers, and might particularly high anxiety among Black students drive racial disparities in misbehavior and subsequent discipline?

While there are many reasons to take a dynamic view of discipline, perhaps the strongest reason is that researchers have already learned much by taking a dynamic view of academic achievement. By taking a temporally granular view of student learning via reviews of in-class activities, homework assignments, quizzes, projects, and exams, scientists have pinpointed factors that contribute to racial disparities in performance—such as fearing confirming negative stereotypes (32), feeling socially excluded (33), feeling their teachers lack confidence in their abilities (34), and low levels of parental engagement (35). Mapping academic disparities over time has also allowed researchers to pinpoint moments in the school year where racial disparities in achievement tend to accelerate and when interventions might therefore be more effective at reducing disparities before they grow (36, 37).

Accordingly, our inquiry regarding the dynamics of discipline is not merely motivated by curiosity. If discipline is indeed dynamic, then mapping related dynamics could provide potent insights regarding how, when, and where to intervene to alleviate discipline, and combat discipline disparities. For example, should we find that discipline rates grow or fall at certain times of year, it could help identify temporally specific phenomena (such as holidays or statewide tests) that might impact student–teacher relationships either positively or negatively. Moreover, knowing when discipline escalates could empower researchers and school leaders to ascertain whether implementing interventions before or during these periods of discipline escalation could help stem them.

No matter how worthwhile the endeavor, to evaluate the dynamics of discipline, one needs atypically granular data that steadily track students' disciplinary experiences over time—data that most school districts are not currently equipped to collect. One can imagine that prior efforts to explore discipline dynamics have been stymied by issues regarding data availability or quality. Here, we leverage one district's remarkably precise and detailed data which it generated using a consistent process—teachers throughout the district completed real-time, brief reports regarding each incident of discipline immediately after it occurred, including the date of the incident, the student behavior that elicited the discipline, and the precise disciplinary response. This process was used for all disciplinary incidents, including suspensions (which accounted for about half of incidents) and minor incidents such as verbal warnings, calls home, and detentions (*SI Appendix*). That the district's data include minor infractions is notable given recent research demonstrating that minor infractions often precede more serious ones, and that Black students receive more minor infractions than their White peers (38). Given this, to accurately capture the extent to which schools grow more or less punitive over time, both minor and severe disciplinary responses should be reviewed. These data allow such a precise and broad review. The data reviewed in this article capture daily disciplinary experiences for 46,964 middle school students who attended school between the 2015 to 2016 and 2018 to 2019 school years (see *SI Appendix* for more information about the sample). Given its breadth and temporal granularity, these data empower us to take a detailed, dynamic view of discipline.

Equipped with these unique data, we first seek to ascertain how the daily discipline rate shifts over time. Is it stable across the year, as implied by the static view? Does it increase over time, as might be expected given research on the escalation of student misbehavior and of teachers' responses to successive misbehaviors? Is it lower after breaks? Or higher during periods of examination? We model the daily discipline rate over time to answer these questions.

Next, leveraging student demographic information, we seek to determine whether the Black–White disparity in the daily discipline rate shifts over time. Specifically, does the daily discipline disparity grow early in the year, as might be expected given that teachers' disciplinary responses escalate more quickly when interacting with Black students?

Finally, if the daily discipline disparity does indeed increase over time, can we predict the kinds of schools where the daily discipline disparity escalates the most aggressively? Research suggests that simply witnessing racial inequity can ironically lead to behavior that increases racial disparities (39–41). Given this, we might expect that the daily discipline disparity will escalate most sharply in schools that have a higher degree of discipline disparity early in the year. Finally, using these data, we can ascertain if the beginning of the year provides sufficient information to predict the end of the year—whether early-year discipline rates will predict end-of-year discipline rates. If this is so, early-year discipline disparities could be used as a "warning sign," empowering districts to target resources toward schools where the disparity is expected to grow.

## Results

**Daily Discipline Rates Are Dynamic.** In each school year, in each school, and for each type of student analyzed, the daily discipline rate (or proportion of students who experienced discipline on a given day) followed the same general pattern (Fig. 1, see *SI Appendix* for subanalyses based on school year, school, and student demographics). The daily discipline rate at the beginning of the year (mid-August) was very low. However, it increased precipitously through Labor Day (early September) then continued to increase, albeit somewhat more slowly, in the weeks leading up to Thanksgiving (mid-November). Thereafter, it declined substantially just before major school breaks (Thanksgiving, Winter Break, and Spring Break), and grew substantially immediately after those breaks (Fig. 2, see *SI Appendix* for subanalyses of prebreak and postbreak trends). Notably, we do not see evidence that the daily discipline rate is responsive to periods of high-stakes testing (*SI Appendix*).

As presented in *SI Appendix*, these trends (steep escalation early in the year, deescalation just before breaks, and escalation directly after breaks) appeared in data subdivided by: school (for 20 schools with 100 or more students in each year), school year (2015 to 2016, 2016 to 2017, 2017 to 2018, and 2018 to 2019), discipline incident type (suspensions incidents or nonsuspension incidents), total discipline experienced in the current year (students who experienced 1 incident, 2 to 4 incidents, or 5 or more incidents), total discipline experienced in the prior year (0 incidents, 1 incident, 2 to 4 incidents, or 5 or more incidents), student grade (6th, 7th, or 8th), student sex (female or male), student English Language Learner (ELL) status (received or did not receive ELL services), student Free-or-Reduced-Price Lunch (FRPL) status (received or did not receive FRPL services), and student race (White, Black, Hispanic, or Asian). To the extent that these data indicate that schools grow more or less punitive at specific times

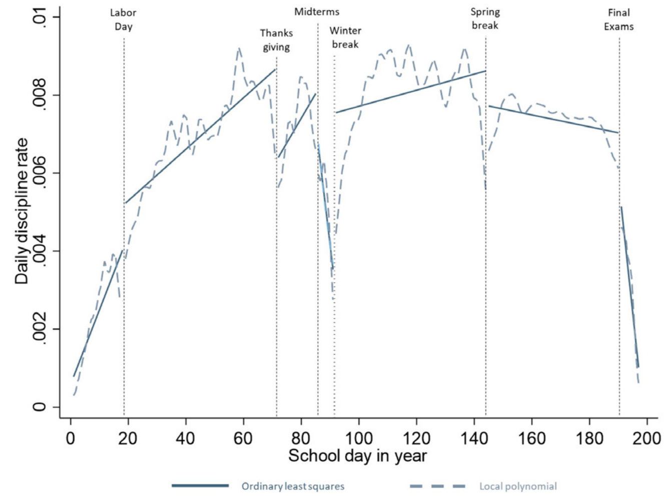


**Fig. 1.** Estimates of the daily discipline rate over time via linear and local polynomial regressions. Note. Solid lines depict linear models. Dashed lines depict local polynomial models. Linear models were executed by regressing a dichotomous indicator of whether a discipline incident occurred on a given student-day (0 – "no discipline on student-day"; 1 – "discipline occurred on student-day") against a running variable expressing the day of the year on which a given student-day fell (1 = "first day of school year," 2 = "second day of school year," and so on). Resulting regressions depict the daily discipline rate. So, for example, they indicate that on the first day of the Fall term, close to 0.1% of students experienced discipline; but at the end of the first period (on the day before Labor Day), 0.4% of students were disciplined. The data are broken out into seven periods: 1) first day of the Fall term through day before Labor Day, 2) Day after Labor Day through day before Thanksgiving break, 3) Day after Thanksgiving break through day before midterm exams, 4) First day of midterm exams through day before Winter break, 5) Day after Winter break through day before Spring break, 6) Day after Spring break through day before final exams, and 7) First day of final exams through the day before Summer break.

in the school year, these dynamic trends impact students of all backgrounds and in varying contexts. Even among students who only experienced one disciplinary incident over the course of the entire year, we see the same dynamic pattern, indicating that these

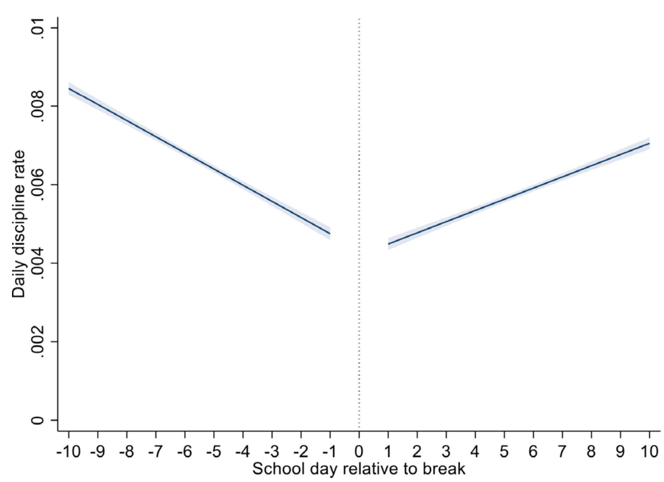


**Fig. 2.** Linear estimates of the daily discipline rate before and after school breaks. Note. Figure depicts results of regressions in which we regressed a dichotomous indicator of whether a discipline incident occurred on a given student-day (0 – "no discipline on student-day"; 1 – "discipline occurred on student-day") against a running variable expressing the day of the year on which a given student-day fell relative to school breaks (−10 = "10 days before a school break," 2 = "the second day back after a school break," and so on). The school breaks reviewed included Thanksgiving, Winter break, Spring break, and Summer break. Resulting regressions depict the daily discipline rate over time, and demonstrate that the daily discipline rate generally falls as school breaks approach, and grow after school breaks. The light blue band around the regression line represents the 95% confidence interval around the linear estimate at any given point.

dynamics are not merely driven by students who frequently experience discipline, but by students throughout the school, in all grade levels, across demographic groups, and even across different schools within the district.

To put these results in context, let us review data from the 2015 to 2016 school year. Here, we report results from ordinary least squares models, but results are functionally identical when we employ other modeling strategies (logistic regression, random coefficient models, local polynomial models—see *SI Appendix*). In 2015 to 2016, on the first day of the school year (8/24/15), our linear models estimate that 0.008% of students experienced discipline. However, after the daily discipline rate steadily increased, on the day before Labor Day (9/4/15), the estimated percentage who experienced discipline was 0.03%—approximately 40 times higher. After the daily discipline rate continued to increase, on the day before Thanksgiving break (10/16/15), 1% of students experienced discipline—over 120 times more than experienced discipline on the first day. While this may seem like a small percentage, bear in mind that we are reporting here the daily discipline rate, meaning that on October 16, 2015, we estimated that one student out of every one hundred students experienced a recorded discipline incident (as compared with a markedly lower rate earlier in the year).

This pattern of discipline escalation early in the year was remarkably similar across years, schools, and student characteristics (*SI Appendix*) and provides strong evidence that schools become more punitive over time in the early weeks of the school year. In data pooled across years and schools, regression models indicate that the "discipline escalation" effect is statistically significant in both the first period (between the first day of fall and Labor Day) and the second period (between Labor Day and Thanksgiving) (first period, $b$(1,283,484 student-days) = 0.00019, student clustered SE = 0.0000082, $t$ = 23.40, $P$ < 0.001; second period, $b$(4,349,163 student-days) = 0.000066, student clustered SE = 0.0000029, $t$ = 23.09, $P$ < 0.001).

The pattern of deescalation before school breaks, and escalation after school breaks, was also similar and consistent across years, schools, and student groups (*SI Appendix*). In pooled data, the discipline deescalation effect just before school breaks is statistically significant [$b$(3,386,883 student-days) = −0.00041, student clustered SE = 0.000016, $t$ = 26.33, $P$ < 0.001], as is the discipline escalation effect after school breaks [$b$(3,386,880 student-days) = 0.00028, student clustered SE = 0.000014, $t$ = 20.87, $P$ < 0.001].

**Racial Disparities Are Dynamic.** Like students overall, students from each racial group (White, Asian, Hispanic, or Black) see their daily discipline rate increase between the first day of the Fall and Labor Day, and again (slightly less quickly) between Labor Day and Thanksgiving. However, Black students see by far the steepest escalation in their daily discipline rate during these time periods. As such, the Black–White daily discipline disparity (or the difference between the daily discipline rate for Black students and White students) grows precipitously between the first day of the year and Labor Day, and continues to grow at a slower, albeit still alarming, rate between Labor Day and Thanksgiving (Fig. 3). The disparity escalation effect appeared when we reviewed data subdivided by school year (2015 to 2016, 2016 to 2017, 2017 to 2018, and 2018 to 2019), discipline incident type (suspensions incidents or nonsuspension incidents), student grade (6th, 7th, or 8th), student sex (female or male), student ELL status (received or did not receive ELL services), and student FRPL status (received or did not receive FRPL services), and even appeared in each of the 20 schools that served 100 or more students in each school year (*SI Appendix*).

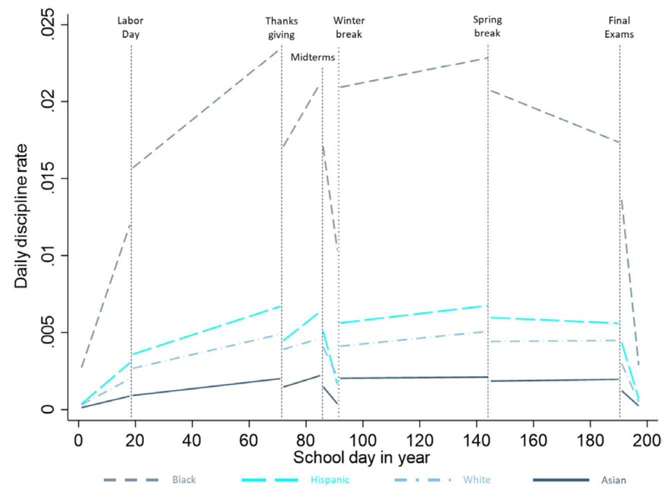


**Fig. 3.** Linear estimates of the daily discipline rate over time, by student race. Note. Figure depicts regressions in which we regressed a dichotomous indicator of whether a discipline incident occurred on a given student-day (0 – "no discipline on student-day"; 1 – "discipline occurred on student-day") against a running variable expressing the day of the year on which a given student-day fell (1 = "first day of school year," 2 = "second day of school year," and so on). Resulting regressions depict the daily discipline rate. So, for example, they indicate that in the 2015 to 2016 school year, on the first day of the Fall term, 0.3% of Black students experienced discipline; but on the 19th day of the 2015 to 2016 school year (the day before Labor Day), 1.2% of Black students were disciplined. The data are broken into the same seven periods explained in Fig. 1. Results are functionally identical when using logistic regression models and random coefficient models (*SI Appendix*).

Looking at data across years and schools, the estimated Black–White daily discipline disparity on the first day of the year is 0.3 percentage points, but increases steadily such that on the day before Labor Day, the Black–White daily discipline disparity is 1.2 percentage points (approximately four times higher). The Black–White daily discipline disparity increases steadily again between Labor Day and Thanksgiving such that on the day before Thanksgiving, it is 1.9 percentage points (more than seven times higher than it was at the beginning of the year). An interacted regression echoes the point, showing that over the period between the first day of Fall and Thanksgiving, the escalation in the daily discipline rate is significantly higher for Black students than for White students ($b_3$(4,121,797 student-days) = 0.0002, student clustered SE = 0.0000086, $t$ = 23.43, $P$ < 0.001). Overall, our models provide strong evidence of disparity escalation.

As noted above, the daily discipline rate diminishes for students of all backgrounds as school breaks approach. This decline is faster for Black students than it is for White students ($b_3$(2,479,040 student-days) = −0.00081, student clustered SE = 0.000062, $t$ = 13.24, $P$ < 0.001). As such, the daily discipline disparity diminishes as school breaks approach. In fact, as depicted in Supporting Information, the estimated daily discipline disparity 10 days A before the typical school break is approximately 0.017 while the estimated daily discipline disparity one day prior to the typical school break is 0.010 (approximately 41% smaller).

**Early-Year Discipline Disparities Predict Both Disparity Escalation over the Fall and End-of-Year Discipline Disparities.** Schools with the largest early-term racial disparities saw the steepest growth in the Black–White discipline disparity (i.e., disparity escalation). Schools varied meaningfully in their comparative treatment of Black and White students in the first 10 school days of the year. We grouped schools into quartiles based on the size of the Black–White discipline gap during these first 10 days. These initial discipline gaps were related to eventual growth in the Black/White discipline gap (Fig. 4). Schools in the fourth quartile (those with the highest Black–White disparity early in the year) saw the largest growth in the Black–White discipline disparity between the first day of the year and Thanksgiving. While schools varied somewhat in their treatment of White students, the dynamic growth of discipline disparities was largely a function of variation in how schools responded to Black students. Fourth quartile schools saw markedly more escalation in the Black discipline rate, and thus saw more escalation in the Black–White disparity. These "high-early disparity," or fourth quartile, schools saw more growth in the Black–White disparity than those in other quartiles: 59% more than first quartile schools; 34% more than second quartile schools; and 83% more than third quartile schools. Interacted regressions echo the point, showing that the Black discipline rate grew more quickly in high-early disparity schools than in high-mid disparity schools, low-mid disparity schools, or low disparity schools (high versus high-mid: $b_3$(496,518 Black student-days) = 0.00012, Black student clustered SE = 0.000024, $t$ = 4.96, $P$ < 0.001; high versus low-mid: $b_3$(380,946 Black student-days) = 0.000059, Black student clustered SE = 0.000027, $t$ = 2.21, $P$ = 0.027); high versus low: $b_3$(390,485 Black student-days) = 0.00012, Black student clustered SE = 0.000025, $t$ = 4.68, $P$ < 0.001).

Schools with larger early-term racial disparities also saw larger end-of-year discipline disparities (Fig. 5). Regression results [$b_1$(70 schools) = 1.43, SE = 0.21, $t$ = 6.93, $P$ < 0.001] indicated that early-year discipline disparities (calculated over the first 10 days of the school year) were predictive of eventual discipline disparities. Indeed, simply knowing the level of discipline disparity that occurred during the first 10 days of the school year was sufficient to predict 41% of the variation in end-of-year discipline disparities [$b_1$(70 schools) = 1.43, SE = 0.21, $t$ = 6.93, $P$ < 0.001]. The predictive power jumps considerably when one reviews the first 20 days of the school year [$b_1$(70 schools) = 1.55, SE = 0.14, $t$ = 11.18, $P$ < 0.001], as knowledge of schools' comparative treatment of Black and White students during just these first 20 days of the school year was sufficient to predict 65% of the variation in schools' end-of-year discipline disparities.

## Discussion

Those hoping to stem racial disparities in discipline have long relied on static, end-of-year discipline rates to identify (7, 42), interpret (19, 43), and evaluate interventions to stem disparities (17, 18, 44, 45). Yet the present research presents convincing evidence that discipline is dynamic. Schools tend to grow more punitive between the beginning of the year and Thanksgiving, and tend to grow increasingly more punitive towards Black students (relative to White students) during this time. They grow less punitive as school breaks approach, and grow more punitive after these breaks. These trends appear across years; they appear across school contexts; they appear across student subpopulations; and they appear both among students who are frequently disciplined and among students who are not frequently disciplined. That these trends appear across years is particularly notable given that the district we studied here implemented Restorative Practices (RP) and Positive Behavioral Interventions and Supports (PBIS) during the 2015 to 2016 school year, in the middle of the study period. Knowledge of the dynamics of discipline and discipline disparities can help us better understand what students and teachers are feeling throughout the school year; when to provide student and teacher-facing support and professional development; and where to target limited resources.

### *What Students and Teachers Feel*

As daily discipline rates and daily discipline disparities escalate, what must this period of disparity escalation feel like for a Black student? On the first day, they may not notice much—as they are treated similarly to their peers of other races. However, inequities in treatment appear quickly and grow precipitously. By Thanksgiving, they experience and may become aware of stark disparities. Research demonstrates that as early as Fall of the 6th grade, Black middle school students are more aware of racial bias in discipline than their peers, which predicts a loss of trust in their schools, and ultimately predicts yearlong discipline infractions

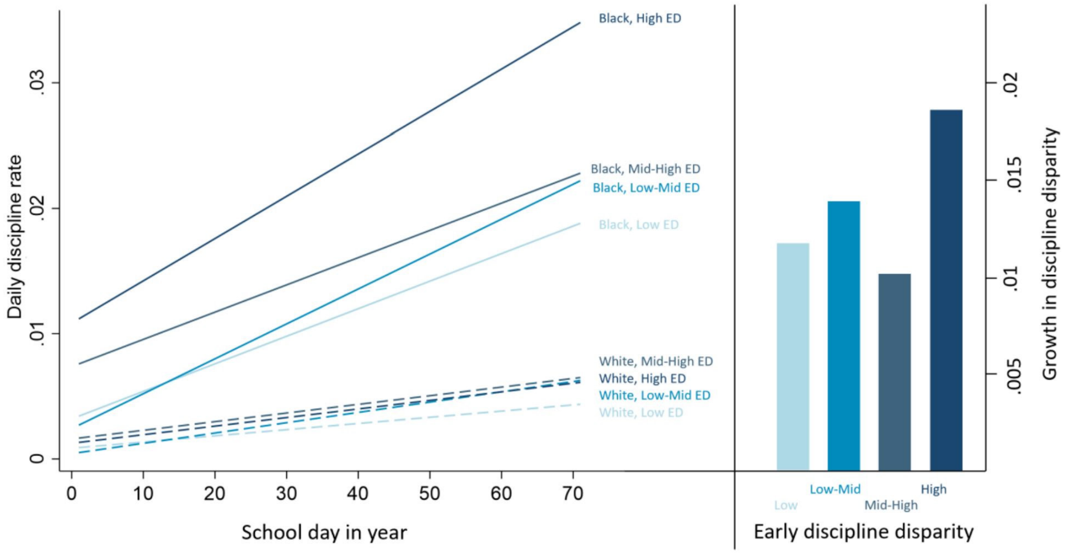


**Fig. 4.** Daily discipline rate for Black and White students, by early-year discipline disparity in school and year; and growth in discipline disparity by early-year discipline disparity. Note. Early-year discipline disparity (ED) was calculated by determining the proportion of Black students that were disciplined in the first 10 d of a given year and subtracting from it the proportion of White students that were discipline during the same time period. Using this measure, schools were subdivided based on whether they had a high ED score (fourth quartile), a high-mid ED score (third quartile), a low-mid ED score (second quartile), or a low ED score (first quartile). Analyses were limited to the 16 schools for which there was sufficient data to precisely estimate the early-year discipline rate (i.e., those that had 50 or more Black and White students in each school year). The growth in discipline disparity was measured by measuring how much the Black–White discipline disparity grew between the first day of the Fall and the day before Thanksgiving.

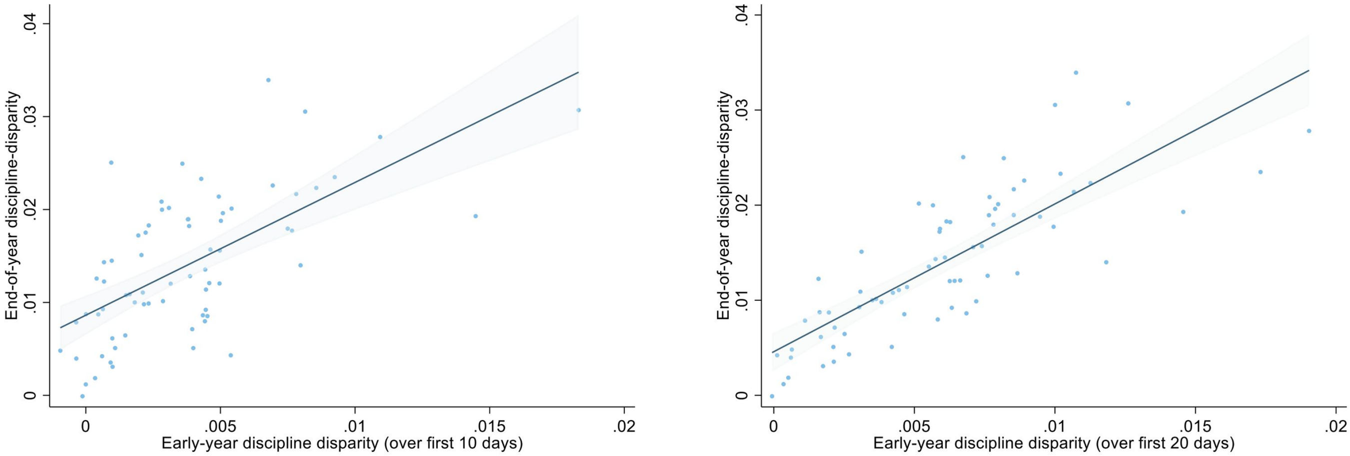


**Fig. 5.** Relationship between early-year discipline disparities and end-of-year discipline disparities using 10 (left) or 20 (right) d of data. Note. Early-year discipline disparities were computed by calculating schools' mean daily discipline rates for Black students during the first 10 (or 20) days, and subtracting schools' mean daily discipline rates for White students during the first 10 (or 20) days. End-of-year disciplinary disparities were computed using the same approach but looking at data across the entire school-year. All calculations were conducted at the level of school and year combinations (e.g., school A in the 2015 to 2016 school year had a unique score from school A in the 2016 to 2017 school year, as school A may have evidenced distinct Black–White disparities in those 2 years) and feature the n = 70 school-years with sufficient data to generate precise estimates of discipline rates for Black and White students (50 or more Black students, and 50 or more White students).



through 8th grade (45). The present research may help explain how this process begins. One possibility is that as racial disparities in daily discipline rates escalate early in the school year, Black students may conclude (rightly or wrongly) that their teachers harbor racial biases. This, in turn, could sour student–teacher relationships, lead to negative relational spirals, and accelerate disparity escalation. Alternatively, the present research could help explain how teacher biases are engendered, activated, sustained, or amplified. As racial disparities in daily discipline rates grow, teachers may conclude (even subconsciously) that Black students are more unruly, and may punish them more often and more aggressively as a result (43). Future research could explore whether students and teachers are aware of escalation in discipline and discipline disparities; whether Black students who perceive growing unevenness in discipline become more distrusting and defiant; and whether teachers who perceive growing unevenness in discipline grow more inequitable in their reactions to students of different racial backgrounds.

Just as discipline grows early in the year, it falls as school breaks approach. Across contexts and student populations, school breaks precipitate short-lived but prominent declines in both daily discipline rates and daily discipline disparities. Future research should seek to identify the cognitive processes that engender these reductions, and ascertain whether the attendant psychological states can be reproduced and leveraged to reduce discipline rates and disparities at other points in the year. For example, researchers might explore whether interventions that make the school year feel less imposing and reduce student and teacher anxiety (such as morning meditation or sporadic in-school celebrations) could help allay discipline and disparities. Finally, and contrary to our expectations, the period after school breaks evidences increases in punitiveness. Research (27) demonstrates that employees see reductions in stress and aggression after vacations, but only if they work in low-stress jobs. Might the stressfulness of the job of teaching play a role in the failure of school vacations to drive sustained reductions in punitiveness? Or might the stress of the job cause teachers to experience spikes in stress when they return that could drive the post-vacation increases in punitiveness we observe in these data? If this is so, researchers should explore whether interventions that help teachers avoid post-vacation stress spikes could yield reductions in discipline escalation and disparity escalation.

Another possible explanation for the decreases in discipline rates before breaks and the increases in discipline rates after school breaks that we observe in these data relates to how students and teachers might accrue trust and understanding. During the sustained periods of successive interactions that occur before breaks, students may grow to trust teachers more, and teachers may come to better understand students. Research indicates that the evolution of student trust (17, 18) and teacher understanding (17, 23) can drive reductions in misbehavior and discipline. However, what of the increases in discipline after school breaks? Perhaps during lengthy breaks, students and teachers lose sight of previously accumulated trust and understanding, leading to downward spirals when they reengage. Future research could explore how students' sense of trust, and teachers' levels of patience and understanding, might shift after school breaks; and whether interventions that help students and teachers reconnect in a positive way after school breaks could help alleviate discipline and discipline disparities.

### *When to Intervene*

The first days of the year evidence huge growth in both the daily discipline rate and the Black–White daily discipline disparity. Perhaps, then, interventions will be more successful if they precede or align with these periods of escalation. For example, students may benefit more from interventions that boost their sense of belonging or empower them to improve student-teacher relationships if these interventions occur on the first day of the school year. This accords with research (17, 18) finding that psychological interventions launched early in the Fall semester (when student–teacher relationships form, and initial conflicts emerge) can reduce suspensions and related racial disparities. These interventions target mechanisms (e.g., students' sense of belonging) that empower students to form positive initial impressions and navigate relationships in a healthy way. Students who experience these interventions early in middle school may avoid relational poisoning that might otherwise occur.

As Thanksgiving approaches, discipline and discipline disparities grow rapidly and reach a zenith. Given research linking exclusionary discipline to student depression (1), schools may want to ensure adequate psychological services are available for students before, during, and after Thanksgiving break. What of the timing of high-stakes testing? Perhaps the most surprising finding from this research is that the daily discipline rate does not appear to be responsive to the timing of high-stakes testing. Notably, our analyses feature data from a single district, and it is possible that this district exhibits an atypically high degree of patience and inclusion during periods of high-stakes testing. If this is so, then perhaps the daily discipline rate might increase prior to periods of high-stakes testing in other contexts. However, it may also be the case that these analyses signal that schools and districts have reduced their reliance on student push-out tactics—or the practice of suspending or expelling low-performing students immediately before high-stakes testing occurs to improve school-wide scores. Alternatively, push-out may persist, but may not be a major driver of temporal variation in discipline rates. Whatever the case may be, these findings cut against earned wisdom surrounding discipline and high-stakes testing, and instead point to other periods in the year (e.g., the periods after school breaks) as moments where discipline tends to grow more severe (and when interventions might be targeted).

This research also provides insights regarding when to implement teacher-facing professional development. Research suggests that teachers' racial biases can indeed drive racial disparities in discipline (22). Thus, another, and perhaps a more direct, means of reducing discipline (and alleviating student anxieties about teacher racial bias) is to empower equitable responses by teachers. Research (17, 23) indicates that encouraging teachers to be more empathetic toward students, to adopt a growth mindset about student–teacher relationships, and to hear and appreciate student perspectives can empower teachers to treat Black and White students more equally. Particularly given how quickly discipline disparities grow in the first days of the school year, the present research suggests that teacher-facing interventions designed to empower teachers to treat students equitably should be implemented before the school year begins, and should perhaps be revisited during or immediately after school breaks.

*Where to Intervene*

This research identified the kinds of schools where discipline disparities tend to grow rapidly over time, and where disparities tend to be most severe at the end of the year—namely schools that see high racial discipline gaps in the first days of the school year. Measuring early discipline disparities may prove a powerful means of predicting eventual discipline disparities. Just as researchers and policymakers leverage early academic indicators to identify students who are not "on track" for high school graduation and thus need targeted intervention to avoid later challenges (46–48), stakeholders could leverage schools' early-year discipline disparity scores as a marker of disparity escalation to come, and of eventual end-of-year racial disparities in discipline. Leveraging as few as 10 days of data, district and other leaders could target resources toward schools where disparities are likely to emerge, empowering these schools to mitigate disparities before escalation takes hold and huge disparities result. Notably, research (49) has provided evidence that student composition is related to differential treatment. Future research could therefore explore whether early-year discipline rates might be used in conjunction with student composition (and other school variables) to predict where racial disparities might emerge even more accurately. This combined approach might empower more effective intervention to avoid racial disparities.

As we deepen our understanding of the dynamic nature of discipline and discipline disparities, we can pinpoint levers for change and sweet spots for intervention. In doing so, we can empower educators to mitigate the lifelong negative consequences of discipline.

**Analytic Plan.** We recruit data from a large school district serving over 45,000 middle school students in over 60 middle schools. The Institutional Review Board (CPHS) at University of California, Berkeley, approved the research we conducted using the data from this district (PI: Okonofua). Like many urban and suburban districts that serve millions of our nation's students, this district is demographically similar to the broader United States (50). For example, similar percentages of students in the district and United States are Asian (5% versus 5%), Black (18% versus 15%), Hispanic (17% versus 27%), and White (55% versus 48%). In terms of economic diversity, 58% of the students receive free and reduced-priced lunch, as compared with 52% of students nationally. Like many other districts, this district serves a mix of urban and suburban neighborhoods and has a median income of around $54,000 (compared with the national median income of $62,000) (51). Schools in the district are also similar in size to schools nationally. On average, schools in the district educated about 590 middle school students, as compared with the national average of 575 middle school students (52); and, like the nation, the district features schools of varying sizes, with 38% of schools serving under 100 students, and 38% of schools serving over 1,000 students. Finally, like many districts (20), this district recently provided its staff with professional development in RJ and PBIS. Thus, in terms of student demographics, neighborhood characteristics, school characteristics, and school practices, this district appears similar to many other urban and suburban districts, and therefore the foregoing findings may be generalizable to many middle school contexts.

We review daily administrative discipline data from this district for the 46,964 middle school students who attended during the 2015 to 2016, 2016 to 2017, 2017 to 2018, and 2018 to 2019 school years. As noted in the introduction, these data were cultivated via a systematic process: Teachers were required to input information about each disciplinary incident in a consistent manner and immediately after it occurred, including the behavior that led to the incident, the disciplinary response, and the date of the incident. These data can be used to ascertain whether any student experienced discipline on any day. We use these data to produce a student-day dataset with each row indicating whether a given student experienced a recorded disciplinary incident on a given day (a student-day dataset with 23,363,457 rows). The most common sanctions were in-school suspensions; out-of-school suspensions; verbal reprimands and warnings; and detentions (*SI Appendix*). In addition, the data provide insight into the school each student attended, their racial and gender demographics, whether they were designated as receiving FRPL, and whether they were designated as receiving ELL instruction ("ELL"). Finally, the data also provide temporal information about each day in the data: the semester it occurred in (e.g., Fall, 2015 to 2016); the day of the week it fell on (Monday–Sunday); whether the day was a school holiday; and whether the day fell during midterms, finals, or high-stakes testing periods. While we omit weekends and holidays from our data (netting a dataset with 15,128,958 student-days), results are similar when they are retained.

We leverage these data to predict the daily discipline rate over time. To achieve, this, we first broke the data into successive, nonoverlapping periods based on the timing of holiday and exam periods. Thus, for example, we broke the 2015 to 2016 school year into 12, nonoverlapping periods, with, for example, the first period encompassing the duration between the first day of the Fall term (8/24/15) and the day before the Labor day holiday (9/4/15); the fifth period encompassing the duration between the first day of midterms (12/15/15) and the last day of the Fall term (12/18/15);

and the 10th period encompassing the duration between the first day back after Spring break (3/28/16) and the day before high-stakes testing began (4/8/16). We used the same approach to break up the other three school years, then used four approaches to model the daily discipline rate over time (*SI Appendix*). Findings were similar across approaches. The four modeling approaches were 1) linear regression, 2) local polynomial regression, 3) logistic regression, and 4) student-level multi-level random coefficient models. Findings were also similar across years; so we pooled data to create a "representative year." Findings were also similar whether we leveraged our more detailed approach to breaking up the school year (into 12 nonoverlapping periods) or used a simpler approach (breaking the school year into seven periods demarcated by the First day of the Fall term, Labor Day, Thanksgiving break, midterm exams, Winter break, Spring break, final exams, and the last day of the Spring term). This approach was deemed appropriate in part because we did not find any evidence to suggest that the discipline rate was responsive to the timing of high-stakes testing (*SI Appendix*).

Above, we present results generated using linear regression to estimate discipline rates over time. We opt for linear models primarily because our findings do not change if we leverage other approaches and because linear models produce more easily comprehensible visuals and analyses. Formally, we regress a dichotomous indicator of whether a discipline incident occurred on a given student-day (0 – "no discipline on student-day"; 1 – "discipline occurred on student-day") against a running variable expressing the day of the year on which a given student-day fell (1 = first day of school year," 2 = "second day of school year," and on). Our linear regression models are:

$$\text{Model 1: Discipline on Student} - \text{Day} = \alpha + \beta_1 (\text{Day of Year}) + \varepsilon.$$

It may not be immediately intuitive why this model estimates the discipline rate over time. Note, however, that on the left side of our regression is our indicator of whether a given student-day involved a disciplinary incident. Regression will attempt to predict the mean of this value contingent on whatever is on the right side of the regression. Now, imagine that the discipline rate does indeed increase steadily as a function of the day of the term such that the proportion of students experiencing discipline per day increases by 0.001 (or 0.1%) each day. Given (as is the case) that each student-day encompasses data from a similar number of students, this model therefore indicates the proportion of student-days on a given day that involved a disciplinary incident or, put another day, the proportion of students who experienced discipline on that day—in short, the daily discipline rate.

Our first research question concerned how the daily discipline rate might evolve over time—is it static or dynamic? To ascertain how the daily discipline rate changes over time within each of our nonoverlapping periods, we look at $\beta_1$. If the daily discipline rate does not change over time in a given period, $\beta_1$ will be zero or very near zero. If the daily discipline rate grows over time in the period, $\beta_1$ will be positive. We term this phenomenon discipline escalation. In addition, if the daily discipline rate falls over time in that the period, $\beta_1$ will be negative—a phenomenon we term "discipline deescalation." While we leverage hypothesis tests to evaluate whether $\beta_1$ is significantly distinct from zero, given the huge amount of data in our models, these tests may not be an effective means of ruling out false positives. We thus also leverage visual review processes that demonstrate how daily discipline rates evolve over time within periods. Our first research question also involved how the daily discipline rate change early in the year, how it might be responsive to school breaks, and how it might react during periods of high-stakes testing. We leverage both hypothesis testing and visual examination approaches to answer related questions: is there discipline escalation at the beginning of the year, is there discipline deescalation around school breaks, is there discipline escalation before or during periods of high-stakes testing, and is there discipline deescalation after periods of high-stakes testing? Questions regarding school breaks and high-stakes testing periods required additional data formatting. For the former, we generated a variable that indicated how close to a school break a given student-day was (e.g., if it was 10 days before a break, it was given a value of "−10," and if it occurred on the third day after students had returned from a given break, it was given a value of "3"). For the latter, we generated a similar variable, however the variable also captured whether a given student-day fell during the high-stakes testing period and, if so, at what point.

Our second research question is whether discipline escalation is differential based on student characteristics such that the daily discipline rate for Black students grows more quickly than the daily discipline rate for White students—a phenomenon we term "disparity escalation." To ascertain whether there is evidence of disparity escalation, we use the same model depicted above, but run on student subgroups. We then compare $\beta_1$ estimates and related visual representations of these estimates for Black students and White students. We also use an interaction regression model in which we formally regress the dichotomous indicator of whether a discipline incident occurred on a given student-day against the running variable expressing the day of the year on which a given student-day fell interacted with the race of the student attached to a given student-day (0 = White, 1 = Black). In the model below, if Black students see steeper discipline escalation than White students in a given period (i.e. if there is evidence of disparity escalation), then $\beta_3$ will be positive and significantly distinct from zero.

$$\begin{aligned}\text{Model 2: Discipline on Student} - \text{Day} &= \alpha + \beta_1 (\text{Day of Year}) + \beta_2 (\text{Student Race}) \\ &\quad + \beta_3 (\text{Day of Year} \times \text{Student Race}) + \varepsilon.\end{aligned}$$

As noted above, we observe that the daily discipline rate diminishes for students of all backgrounds as school breaks approach, and appears to diminish more quickly for Black students than for White students. To empirically ascertain if the daily discipline disparity reduces over time as school breaks approach, we leverage an interacted regression in which we regress our discipline on student-day marker against a variable that marks proximity to school breaks (in which a student-day would have a score of "−10" if it occurs 10 days prior to a school break, and a score of −1 if it occurs the day before a school break), a dummy variable indicating if the student is Black (1) or White (0), and the interaction of the prior two variables. Whereas in our visuals we look both at student-days that occur before school breaks and at student-days that appear after school breaks, in this analysis, we limit ourselves to student-days that appear before school breaks to ascertain whether the deescalation in the daily discipline rate that occurs before school breaks is differential based on race. Formally our model is:

$$\begin{aligned}\text{Model 3: Discipline on Student} - \text{Day} &= \alpha + \beta_1 (\text{Day Relative to School Break}) + \beta_2 (\text{Student Race}) \\ &\quad + \beta_3 (\text{Day Relative to School Break} \times \text{Student Race}) + \varepsilon.\end{aligned}$$

Our last set of analyses reviews whether the disparity escalation is more extreme in schools with a high initial level of discipline disparity. To conduct these analyses, we first predict the proportion of

Black students who experienced discipline on any day in the first 10 d of the school year in each school. To do this, within each school and in each year, we calculate the mean of the discipline-incident marker for student-days of Black students and occurring in the first 10 days of the year. We repeat the process for White students to calculate the school-specific initial discipline rate for White students. We then subtract the White initial discipline rate from the Black initial discipline rate to generate the school-and-year-specific measure of initial discipline rate disparity. Importantly, to ensure a precise measure of discipline disparity, we restrict this particular analysis to students who attended schools that had at least 50 White students and at least 50 Black students in each of the four school years (16 schools). Having generated a school-and-year-specific measure of discipline disparity, we identify the early-year disparity quartile that each student belongs to—for example, in that year, was that student in a school that had one of the largest early-year disparities in discipline (quartile four), or was that student in a school that had one of the smallest early-year disparities in discipline (quartile one). Having determined the kind of school that each student attended in each year, we then ascertained whether students saw faster discipline disparity escalation when they attended schools that had higher rates of discipline disparity early in the year. As with our approach to evaluating the second research question, here again we use a visual representation of linear regression results to ascertain if disparities grow more quickly in certain schools. We also formally compare the first derivative of the Black daily discipline rate in schools in the fourth quartile (high initial disparity) to that of schools in the other three quartiles (e.g., low initial disparity) via interacted regression:

$$
\begin{aligned}
&\text{Model 4: Discipline on Student} - \text{Day (Black Students)}\\
&= \alpha + \beta_1\,(\text{Day of Year}) + \beta_2\,(\text{Discipline Quartile})\\
&+ \beta_3\,(\text{Day of Year} \times \text{Discipline Quartile}) + \varepsilon.
\end{aligned}
$$

We also review whether end-of-year Black–White discipline disparities are more extreme in schools with a high initial level of discipline disparity. Put another way, we examine whether early-year discipline disparities might be used as an early warning sign to predict end-of-year discipline disparities. To conduct these analyses, we first group the data into school and year combinations (e.g., school A in 2015, or school C in 2018). Then, within each school-year combination, we predict the mean daily discipline rate for Black students during the first 10 days of the year, and repeat the process for White students. Finally, we calculate the difference between the Black mean early-year discipline rate and the White early-year mean discipline rate to glean the early-year Black–White discipline disparity. This can be understood as how much more discipline Black students typically received than White students during the first 10 d of the school year. We repeat this process for a 20-days time period in part because discipline escalation and disparity escalation are both most severe during the first 20 or so days of the school year. We then calculate the end-of-year Black–White discipline disparity by repeating the same process, but over the year (rather than simply over the first 10 or 20 days). To ensure precision, we restrict our analyses to school-year combinations that have at least 50 Black students and at least 50 White students. Having generated school-and-year specific measures of initial discipline disparity and eventual discipline disparity, we regress the end-of-year discipline disparity against the initial discipline disparity via the following two models:

$$
\begin{aligned}
&\text{Model 5: End} - \text{of} - \text{Year Discipline Disparity}\\
&\quad = \alpha + \beta_1(10 - \text{Day Discipline Disparity}) + \varepsilon,\\
&\text{Model 6: End} - \text{of} - \text{Year Discipline Disparity}\\
&\quad = \alpha + \beta_1(20 - \text{Day Discipline Disparity}) + \varepsilon.
\end{aligned}
$$

**Data, Materials, and Software Availability.** Data are not publicly available at present due to the potentially identifiable nature of students' discipline records. However, all data created and or analyzed as part of this study and all R and STATA code used to analyze the data are available from the corresponding authors on reasonable request.

**ACKNOWLEDGMENTS.** We would like to thank the school district for collecting and sharing such detailed data with our research team; Rucker Johnson and Avi Feller for graciously discussing and reviewing our modeling strategies; our two PNAS reviewers for providing thoughtful and helpful feedback, Amrita Maitreyi for assisting in the preparation of the manuscript; and the Tides Foundation (SPO ID 042589-001, TFR 16-04025), Google's Computer Science Education Research team, and the Character Lab for financially supporting this work. We would also like to thank E. Adaora for her support.

1. J. L. Rushton, M. Forcier, R. M. Schectman, Epidemiology of depressive symptoms in the National Longitudinal Survey of Adolescent Health. *J. Am. Acad. Child Adolesc. Psychiatry* **41**, 199–205 (2002).
2. A. Kupchik, T. J. Catlaw, Discipline and participation: The long-term effects of suspension and school security on the political and civic engagement of youth. *Youth Soc.* **47**, 95–124 (2015).
3. A. L. Noltemeyer, R. M. Ward, C. Mcloughlin, Relationship between school suspension and student outcomes: A meta-analysis. *School Psych. Rev.* **44**, 224–240 (2015).
4. A. Bacher-Hicks, S. Billings, D. Deming, The school to prison pipeline: Long-run impacts of school suspensions on adult crime. https://www.nber.org/system/files/working_papers/w26257/w26257.pdf (September 2019).
5. C. LiCalsi, D. Osher, P. Bailey, An empirical examination of the effects of suspension and suspension severity on behavioral and academic outcomes. *American Institutes for Research* (Aug. 2021). https://www.air.org/sites/default/files/2021-08/NYC-Suspension-Effects-Behavioral-Academic-Outcomes-August-2021.pdf. Accessed 7 June, 2022.
6. R. W. Rumberger, D. J. Losen, The high cost of harsh discipline and its disparate impact. The Center for Civil Rights Remedies at The Civil Rights Project. https://www.civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/the-high-cost-of-harsh-discipline-and-its-disparate-impact/UCLA_HighCost_6-2_948.pdf (June, 2016).
7. U. S. Department of Education, 2013-2014 Civil Rights Data Collection: A First Look. US Department of Education Office for Civil Rights, October 2016. https://ocrdata.ed.gov/assets/downloads/2013-14-first-look.pdf. Accessed 6 October 2021.
8. J. M. Nowicki, K-12 education: Discipline disparities for black students, boys, and students with disabilities. *U.S. Government Accountability Office* (Mar. 22, 2018). https://www.gao.gov/assets/700/692095.pdf.
9. The United States Department of Justice: Office of Public Affairs, Attorney General Holder, Secretary Duncan announces effort to respond to school-to-prison pipeline by supporting good discipline practices, https://www.justice.gov/opa/pr/attorney-general-holder-secretary-duncan-announce-effort-respond-school-prison-pipeline (July, 2011).
10. Department of Education and Department of Justice, US Departments of Education and Justice release school discipline guidance package to enhance school climate and improve school discipline policies/practices. https://www.ed.gov/news/press-releases/us-departments-education-and-justice-release-school-discipline-guidance-package/ (March 2014).
11. A. Kamenetz, Suspensions are down In U.S. schools but large racial gaps remain (Dec. 17, 2018). https://www.npr.org/2018/12/17/677508707/suspensions-are-down-in-u-s-schools-but-large-racial-gaps-remain.
12. R. O. Welsh, S. Little, The school discipline dilemma: A comprehensive review of disparities and alternative approaches. *Rev. Educ. Res.* **88**, 752–794 (2018).
13. M. Gopalan, A. A. Nelson, Understanding the racial discipline gap in schools. *AERA Open.* **5**, 1–26 (2019).
14. C. Redding, A teacher like me: A review of the effect of student-teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Rev. Educ. Res.* **89**, 499–535 (2019).
15. R. J. Skiba *et al.*, Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psych. Rev.* **40**, 85–107 (2011).
16. R. J. Skiba *et al.*, Parsing disciplinary disproportionality: Contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion. *Am. Educ. Res. J.* **51**, 640–670 (2014).
17. J. A. Okonofua, D. Paunesku, G. M. Walton, Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5221–5226 (2016).
18. J. P. Goyer *et al.*, Targeted identity-safety interventions cause lasting reductions in discipline citations among negatively stereotyped boys. *J. Pers. Soc. Psychol.* **117**, 229–259 (2019).
19. T. Riddle, S. Sinclair, Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8255–8260 (2019).
20. S. Darling-Hammond *et al.*, Effectiveness of restorative justice in U.S. k-12 schools: A review of quantitative research. *Contemp. Sch. Psychol.* **24**, 295–308 (2020).
21. K. Mcintosh *et al.*, Effects of an equity-focused PBIS approach to school improvement on exclusionary discipline and school climate. *Prev. Sch. Fail.* **64**, 354–361 (2021).
22. J. A. Okonofua, J. L. Eberhardt, Two strikes: Race and the disciplining of young students. *Psychol. Sci.* **26**, 617–624 (2015).
23. J. A. Okonofua, A. D. Perez, S. Darling-Hammond, When policy and psychology meet: Mitigating the consequences of bias in schools. *Sci. Adv.* **6**, eaba9479 (2020).

24. S. M. Way, School discipline and disruptive classroom behavior: The moderating effects of student perceptions. *Sociol. Q.* **52**, 346–375 (2011).
25. R. Pesta, School punishment, deterrence, and race: A partial test of defiance theory. *Crime Delinq.* **68**, 463–494 (2022).
26. S. Verma, N. B. Allen, B. Bei, Highs and lows: Naturalistic changes in mood and everyday hassles over school and vacation periods in adolescents. *J. Adolesc.* **61**, 17–21 (2017).
27. B. B. Gump, B. Hruska, S. D. Pressman, A. Park, K. G. Bendinskas, Vacation's lingering benefits, but only for those with low stress jobs. *Psychol. Health.* **36**, 895–912 (2021).
28. J. F. Dewald, A. M. Meijer, F. J. Oort, G. A. Kerkhof, S. M. Bogels, Adolescents' sleep in low-stress and high-stress (Exam) times: A prospective quasi-experiment. *Behav. Sleep Med.* **12**, 493–506 (2014).
29. N. K. Segool, J. S. Carlson, A. N. Goforth, N. von der Embse, J. A. Barterian, Heightened test anxiety among young children: Elementary school students' anxious responses to high-stakes testing. *Psychol. Sch.* **50**, 489–499 (2013).
30. B. G. Turner, D. C. Beidel, S. Hughes, M. W. Turner, Test anxiety in African American school children. *Sch. Psychol. Q.* **8**, 140–152 (1993).
31. S. L. Nichols, S. Brewington, Preservice teachers' beliefs about high-stakes testing and their working environments. *Educ. Policy Anal. Arch.* **28**, 1–38 (2020).
32. C. M. Steele, J. Aronson, Stereotype threat and the intellectual test performance of African Americans. *J. Pers. Soc. Psychol.* **69**, 797–811 (1995).
33. G. M. Walton, G. L. Cohen, A brief social-belonging intervention improves academic and health outcomes of minority students. *Science* **331**, 1447–1451 (2011).
34. D. S. Yaeger *et al.*, Break the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide. *J. Exp. Psych.* **143**, 804–824 (2014).
35. A. H. Houri, A. J. Thayer, C. R. Cook, Targeting parent trust ot enhance engagement in a school-home communication system: A double-blind experiment of a parental wise feedback intervention. *School Psych.* **34**, 421–432 (2019).
36. G. L. Cohen, J. Garcia, N. Apfel, A. Master, Reducing the racial achievement gap: A social-psychological intervention. *Science* **313**, 1307–1310 (2006).
37. G. L. Cohen, J. Garcia, V. Purdie-Vaughns, N. Apfel, P. Brzustoski, Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science* **324**, 400–403 (2009).
38. J. Amemiya, E. Mortenson, M. T. Wang, Minor infractions are not minor: School infractions for minor misconduct may increase adolescents' defiant behavior and contribute to racial disparities in school discipline. *Am. Psychol.* **75**, 23–26 (2020).
39. M. Peffley, J. Hurwitz, Persuasion and resistance: Race and the death penalty in America. *Am. J. Pol. Sci.* **51**, 996–1012 (2007).
40. R. C. Hetey, J. L. Eberhardt, Racial disparities in incarceration increase acceptance of punitive policies. *Psychol. Sci.* **25**, 1949–1954 (2014).
41. R. C. Hetey, J. L. Eberhardt, The numbers don't speak for themselves: Racial disparities and the persistence of inequality in the criminal justice system. *Curr. Dir. Psychol. Science* **27**, 183–187 (2018).
42. U. S. Department of Education, Data from "Civil Rights Data Collection". *Office for Civil Rights* http://ocrdata.ed.gov/ (2012).
43. J. A. Okonofua, G. M. Walton, J. L. Eberhardt, A vicious cycle: A social-psychological account of extreme racial disparities in school discipline. *Perspect. Psychol. Sci.* **11**, 381–398 (2016).
44. A. Gregory *et al.*, My teaching partner-secondary: A video-based coaching model. *Theory Pract.* **56**, 38–45 (2017).
45. 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: Erratum. *Child Dev.* **88**, 1033 (2017).
46. E. Allensworth, J. Q. Easton, *The on-track indicator as a predictor of high school graduation* (Consortium of Chicago School Research at the University of Chicago, https://consortium.uchicago.edu/sites/default/files/2018-10/p78.pdf Accessed 6 October 2021 June2005).
47. E. Allensworth, The use of ninth-grade early warning indicators to improve Chicago schools. *J. Educ. Stud. Placed Risk* **18**, 68–83 (2013).
48. J. J. Kemple, M. D. Segeritz, N. Stepheson, Building on-track indicators for high school graduation and college readiness: Evidence from New York City. *J. Educ. Stud. Placed Risk* **18**, 7–28 (2013).
49. R. E. Fish, Standing out and sorting in: Exploring the role of racial composition in racial disparities in special education. *Am. Educ. Res. J.* **56**, 2573–2608 (2019).
50. Institute of Education Sciences. Data retrieved from "Racial/Ethnic Enrollment in Public Schools" (National Center for Education Statistics, 2017, https://nces.ed.gov/programs/coe/indicator_cge.asp Accessed 6 October 2021).
51. United States Census Bureau, Data from "QuickFacts." Census Bureau. https://www.census.gov/quickfacts/. Accessed 6 October 2021.
52. Institute of Education Sciences. Data retrieved from "Numbers and Types of Public Elementary and Secondary Schools From the Common Core of Data: School Year 2009–10" (National Center for Education Statistics, https://nces.ed.gov/pubs2011/pesschools09/tables/table_05.asp Accessed 6 October 2021 2011).