

Social Psychology

A Test of Racial Disparities in Principals' Disciplinary Decisions in a National Sample

Shoshana N. Jarvis¹, Z. E. Ferguson², Jason A. Okonofua³

¹ Haas School of Business, University of California, Berkeley, CA, US, ² Psychology Department, University of Washington, Seattle, WA, US,

³ Psychology Department, University of California, Berkeley, CA, US

Keywords: principals, racial disparities, school discipline, stereotyping

<https://doi.org/10.1525/collabra.88326>

Collabra: Psychology

Vol. 9, Issue 1, 2023

Access to education is important for success as an adult. Exclusionary discipline (e.g., suspensions) reduces opportunities for students to complete their education and be strong candidates for future jobs. Black students face a disproportionately high risk of disciplinary action. Thus, it is important to understand when and how racial disparities in suspensions emerge in order to reduce their disproportionate negative impacts on Black students. Past research found racial disparities emerge after two misbehaviors among teachers and just a single misbehavior among assistant principals. The current research tests the generalizability of racial disparities in discipline from principals across the United States and a psychological process that potentially contributes to the racial disparities: their perception of their professional role relative to that of teachers. In this procedure and with a diverse sample, principals did not endorse significantly different amounts of discipline for Black and White students. We explore potential explanations of these null results in the discussion.

Racial disparities in discipline meted out to Black and White students are pervasive from pre-kindergarten through 12th grade (R. J. Skiba et al., 2002, 2011; US Department of Education Office for Civil Rights, 2016). According to records by the US Department of Education Office for Civil Rights, in the 2015–2016 school year, Black students were 3.6 times more likely to receive out-of-school suspensions in comparison to their White peers. Studies with teachers and principals that experimentally manipulate the race of the student while holding the misbehavior constant find convergent evidence for racial disparities in discipline (Jarvis & Okonofua, 2020; Okonofua & Eberhardt, 2015). The culmination of this body of research gives credence to a causal relationship for why Black students are overrepresented among students who are disciplined. The overrepresentation of Black students in suspensions cannot fully be explained by their behavior but also because they are more frequently referred to principals' offices and are ultimately given more extreme punishment for the same misbehavior as other students.

Suspensions and expulsions have consequential impacts on the educational trajectory for students and play a key role in the school-to-prison pipeline (Fabelo et al., 2011; Rocque & Paternoster, 2011). The school-to-prison pipeline refers to the systematic ways in which Black students are removed from educational environments and

pushed towards incarceration. After students are suspended, they fall further behind in their coursework, which makes it more difficult to stay in school (Arcia, 2006). With less time in school, students are more likely to become involved with activities that lead to arrest and a greater likelihood of becoming incarcerated as an adult (Bacher-Hicks et al., 2019; Kim et al., 2010). It is important to understand how each role in the education system contributes to disproportionate discipline in order to best assess how to address the school-to-prison pipeline.

In addition to bias in teachers' disciplinary decisions (Okonofua & Eberhardt, 2015), bias in principals' decisions could contribute to the process by which racial disparities in discipline occur and could amplify the racial disparities in discipline by compounding the impacts of racial disparities in teacher referrals with their own biases (Jarvis & Okonofua, 2020). While teachers make disciplinary decisions at the classroom level, principals manage the entire school. Through their power to determine school-wide policies, principals set decisions about class size, resources, and monitoring over teachers that, in turn, influence the school-wide culture (Hornig et al., 2010; Leithwood et al., 2004). Effects of principals' decisions can be felt across many domains of their school including math scores, reading scores, and graduation rates (Coelli & Green, 2012; Dhuey & Smith, 2014). As the final decision makers for dis-

disciplinary decisions, principals could have school-wide impacts on how discipline is arbitrated across students. In a correlational study with principals, their attitudes toward discipline was related to the extent to which there were racial disparities in discipline at their schools (R. J. Skiba et al., 2014).

There are aspects of principals' job responsibilities that could influence how they consider disciplinary decisions. Notably, principals have to divide their attention across many other administrative tasks which could lead to limited bandwidth for disciplinary decisions (Goldring et al., 2008). The principal is responsible for the wellbeing of the entire school community as opposed to a classroom, which could lead to different considerations as to what is considered disruptive. Additionally, principals spend less time than teachers with individual students on a daily basis which could impact what is seen as appropriate discipline for students. While there is reason to suspect principals can contribute to the school-to-prison pipeline, there is little experimental evidence to investigate the ways stereotypes could affect their perception and decision-making in the discipline process and thus their impact on racial disparities in discipline.

In a previous study with school administrators, assistant principals in one large school district rated how they would discipline a Black or White male student for the same series of two misbehaviors (Jarvis & Okonofua, 2020). Consistent with findings from teacher samples, assistant principals rated the Black student as more likely to be a troublemaker and endorsed more days of detention when the student was Black as opposed to White. Further, the greater likelihood to view the student as a troublemaker explained the relationship between student race and more severe discipline. However, an inconsistency was observed. Assistant principals endorsed greater discipline severity for the Black student across both misbehaviors. This differs from research with teachers which found a significant "escalation effect" such that racial disparities in discipline only reached significance after the second misbehavior, indicating a Black student, as compared to a White student, faced a significantly sharper escalation in negative discipline responses (Okonofua & Eberhardt, 2015).

Why might the emergence of this racial disparity in the discipline process differ for principals and teachers? The present study hypothesizes that the difference is due to principals' perception of their authority in the school relative to that of teachers. The procedure in previous research is adapted to ask principals for their beliefs about their role in school relative to teachers before they begin the study. This was measured immediately before principals read about the series of misbehaviors, because it was predicted that it would explain the escalation effect – or a lack thereof. That is, principals could view their disciplinary role as fundamentally distinct from that of teachers, and this differing perspective could explain the quick escalation. Principals are responsible for the safety and wellbeing of an entire school as opposed to a classroom, and such structural aspects of their job could influence how they

make disciplinary decisions (i.e., being prone to more severe decisions).

The present research also seeks to address key limitations in previous research: diversity and size of the participant sample. The previous sample only included assistant principals, not principals, from a single large and diverse school district in one Southeastern state. It also was relatively small and potentially underpowered to test the interactions between race and time (from the first to second misbehavior).

The current research investigates how principals may contribute to racial disparities in discipline across two misbehaviors using an adapted procedure from Jarvis & Okonofua (2020). To address the limitations of the previous study, the present study used a larger, randomly-selected sample of principals and assistant principals from across the United States. At the outset, principals answered questions about how they viewed their role as principals compared to the role of teachers. This was aimed to test a mechanism for why disciplinary disparities arose more quickly for principals than for teachers: that principals could view their disciplinary role as fundamentally distinct from that of teachers, and this differing perspective could explain the quick escalation. Principals and assistant principals read teacher referrals about a male student whose race was manipulated to be either White or Black by their name (Greenwald et al., 1998). They rated how severely they would discipline the student across two misbehaviors and responded to questions about the characteristics and future of the student (e.g., the extent to which the student is a troublemaker and the likelihood they would be suspended in the future).

There were two key pre-registered hypotheses (<https://osf.io/2z8u3>). 1) We hypothesized that we would replicate the results from Jarvis and Okonofua (2020). Namely, principals would discipline Black students more severely, rate them as more likely to be a troublemaker, and give them more days of detention than White students for the same misbehavior, and troublemaker labeling would mediate the effect of race on discipline. Our larger sample size increased power to detect if racial disparities emerge over time (like previous research with teachers), or if Black students are consistently disciplined more than White students across misbehaviors (like previous research with principals). 2) We hypothesized that the difference in the extent to which principals view their role, as compared to teachers, as one who gets to know individual students and considers the classroom environment and school climate when making disciplinary decisions would moderate the racial disparities in discipline severity and troublemaker ratings. Specifically, as differences between principals' and teachers' roles increased, racial disparities in discipline severity and troublemaker ratings would also increase. All materials, syntax, and data are available on OSF (<https://osf.io/j87zk/>). Other hypotheses were pre-registered and are reported in another paper (Ferguson et al., 2023).

Methods

Participants and Recruitment

Principals and assistant principals (henceforth simply referred to as principals) were recruited from schools randomly sampled from a list of all the schools in the United States from the 2013-2014 school year. We recruited both principals and assistant principals, because duties – namely the discipline decision-making role – differ from district to district. Publicly available contact information for each principal from the first 4000 schools of the randomized list was compiled. Samples of principals were sent an initial email and two follow up emails one week apart until our target sample size of 300 principals finished the survey. The target sample size was preregistered and selected for being slightly greater than 2.5 times more than the sample size of Jarvis & Okonofua (2020) as recommended by Simonsohn (2015) for selecting sample sizes of replications. Our preregistered recruitment procedure indicated that we would initially email 300 schools worth of principals and subsequently email 100 new schools each week until reaching our target sample size. However, participation was slower than anticipated. In order to successfully recruit our target sample before the end of the academic school year, 500-800 schools worth of principals were emailed at a time until the target sample size was achieved. The survey closed one week after it achieved the target sample size to allow all contacted principals the opportunity to complete the survey. The recruitment period lasted six weeks in April and May 2019.

We contacted 6,348 principals from 3,953 schools. Of these, 327 principals completed the online survey from 221 schools. We excluded those who met any of the following pre-registered criteria: they took less than 8 minutes or more than 50 minutes to complete the survey ($N = 73$), failed the attention check ($N = 6$), failed the suspicion check ($N = 7$), or failed multiple exclusion criteria ($N = 6$), leaving 234 principals (131 principals and 103 assistant principals) in the sample. Principals were almost evenly split by gender (46% men, 54% women), and were mostly White (79%) and 1% Asian, 11% Black, 5% Latinx, 2% Multiracial, and 2% other or did not report. The racial demographics were comparable to the national averages for principals: 77.7% White, 10.5% Black, 8.9% Latinx, and 2.9% other (*National Teacher and Principal Survey (NTPS)*, n.d.). On average, principals were 46 years old ($SD = 8$ years) and had 12 years of experience as a teacher ($SD = 6$ years) and 7 years of experience as a principal ($SD = 6$ years). Principals taught at schools across grade levels (48% elementary school, 39% middle school, 41% high school). School-level was inclusive, that is a principal could be counted in multiple categories. Any principal who worked at a school with K-5 was counted as elementary school, 6-8 as middle school, and 9-12 as high school. Principals also represented schools from 38 different states and the District of Columbia: Northeast: 16%, Southeast: 23%, Midwest: 33%, West: 18%, Southwest: 10%. Average school demographics reflected national averages: White: 60%, Latinx: 17%, Black: 12%, Asian: 5%, Other: 5%.

Procedure

This procedure was approved by the University of California, Berkeley's Committee for Protection of Human Subjects, and written consent was obtained. Principals first answered questions measuring differences between their roles as principals and the roles of teachers (principal-teacher role items), which were counterbalanced such that they either rated their role first or teachers' role first. They rated the extent to which principals and teachers get to know individual students. Principals also rated the extent to which principals and teachers consider the following aspects when making decisions about suspensions and expulsions: a student's character, a student's development, a student's academic performance, a student's behavior, the school climate, and the classroom environment. Lastly, principals rated how much they trust teachers' judgments and how much teachers trust their judgment.

Next, principals completed the same materials as Jarvis and Okonofua (2020) in which they read about a student who was either Black (Darnell or Deshawn) or White (Greg or Jake) who misbehaved twice. Participants read a teacher's office referral describing each time the student misbehaved in a classroom (referral order counterbalanced). Referrals were taken verbatim from a California school district's office referral records. All typos were present in the original referrals (Okonofua & Eberhardt, 2015).

Darnell/Greg is constantly disrupting the class environment by strolling around the classroom at random intervals, getting tissues from the tissue box multiple times during a 50 minute class, throwing items away constantly; in general, Darnell/Greg circulates around the room and up and down the rows to see what other students are doing, have eyes on him, and disrupt the flow of the lecture or activity the class was participating in.

They then rated the extent to which the behavior was severe, hindered the teacher's ability to maintain order in the classroom, how irritated they (the principal) felt by the student, and how severely they would discipline the student. The narrative continued with the student misbehaving again three days later and was referred to the principal.

Darnell/Greg is sleeping in class. I told him to pick his head up and get to work. He only picked his head up. He chose to rest it on his hand and continue to sleep. So I asked him one more time and again, Darnell/Greg refused to do work. I asked him to leave class and go to the office to tell you that he won't do his work and chose to sleep instead. He refused to do this as well.

The principals gave the same ratings as with the previous misbehavior as well as the following dependent variables on Likert scales: 1) the likelihood the student was a troublemaker, 2) the likelihood the misbehavior was indicative of a pattern, 3) the extent the principal would worry that this would become a pattern across students, 4) how many days they would send the student to detention, 5) the likelihood the student would be suspended in the future, 6) the difficulty to set the student on the right course,

7) the extent the student would benefit from extracurricular activities, 8) the likelihood the student's parents would be involved to improve the his behavior at school, and 9) the extent to which the student would benefit from seeing a counselor. They then responded to suspicion, attention, and manipulation check questions.

Additional measures were collected unrelated to this investigation and were not analyzed: a shortened discipline perspectives scale (R. Skiba et al., 2007), colorblindness/multiculturalism scale (Ryan et al., 2007), motivations to control prejudice (Plant & Devine, 1998), and a respect scale measuring attitudes on where respect is due.

Analytic Approach

All analyses were conducted using the R software program and all code and anonymized data are available on OSF (<https://osf.io/kw69a/>). Within-subjects tests of the impact of race on discipline severity and feeling troubled over time were tested using repeated measures ANOVAs, and no post hoc tests were performed. Effects of race on all other outcome variables were tested using two-tailed *t*-tests and moderation analyses using linear regression. The threshold of significance was defined as .05. As reflected in the preregistration, responses were identified as outliers if they were ± 2.5 *SD* from the mean within their condition and were excluded from analyses. A post-hoc sensitivity analysis found that for the sample with exclusions ($N = 234$), the study was powered to detect $t = 1.65$, $d = 0.43$, and $t = 1.65$, $d = 0.36$ for the full sample ($N = 327$).

Results

Preliminary Analyses

As a manipulation check, principals rated how likely they thought the student they read about was Black. Principals who read about a student named Darnell or Deshawn were more likely to think the student was Black than the principals who read about a student named Greg or Jake, $t(198) = 9.34$, $p < .001$, $d = 1.22$. To rule out potential associations between race and economic class, principals also rated the likelihood the student was from a low-income neighborhood. This was to test if principals extrapolated class from the racial information and to isolate the effects of race from those of perceived social class. Principals who read about a student who was Black thought the student was from a school in a lower socio-economic district more so than principals who read about a student who was White, $t(213) = 3.04$, $p = .003$, $d = 0.40$. That said, the effect of the race of the student predicted the extent to which the principal thought the student was Black while controlling for the perceived socio-economic status of the school district, $b = 1.05$, $t(228) = 8.68$, $p < .001$, 95% *CI* [0.81, 1.29].

The extent to which principals rated the behavior as severe, hindered the teacher's ability to maintain order in the classroom, and how irritated they felt were all highly correlated for each misbehavior (time 1: $r_s > .54$; time 2: $r_s > .83$). Consistent with past studies (Okonofua & Eberhardt, 2015), these three responses were averaged at each time point to create an aggregated score of the principal's affective re-

sponse called "feeling troubled" (time 1: $\alpha = .85$; time 2: $\alpha = .95$).

Replication of Jarvis and Okonofua (2020)

Hypothesis 1 predicted the key findings from Jarvis & Okonofua (2020) would replicate, including that principals would endorse greater discipline severity for Black as compared to White students. For discipline severity, there was no main effect of student race, $F(1, 228) = 1.71$, $p = .193$, $d = 0.14$, indicating no significant difference in how principals endorsed disciplining Black and White students. Similar to previous findings, there was a main effect of time, $F(1, 228) = 43.63$, $p < .001$, $d = 0.61$ indicating, from the first misbehavior to the second, discipline severity increased for all students (Black time 1: $M = 2.50$, $SD = 1.08$; Black time 2: $M = 3.21$, $SD = 1.37$; White time 1: $M = 2.63$, $SD = 1.19$; White time 2: $M = 3.44$, $SD = 1.30$). We did not observe a race by time interaction, $F(1, 228) = 0.28$, $p = .597$ (see Figure 1).

We next hypothesized that principals would rate Black students as more likely to be considered a troublemaker and be given more days of detention than White students. Black students ($M = 2.04$, $SD = 0.84$) and White students ($M = 2.19$, $SD = 0.84$) were not rated significantly differently in likelihood of being a troublemaker, $t(229) = 1.39$, $p = .167$, $d = 0.18$. There also were not significant differences between Black students ($M = 2.09$, $SD = 0.71$) and White students ($M = 2.12$, $SD = 0.85$) in days of detention given, $t(226) = 0.24$, $p = .810$, $d = 0.03$ (see Fig. 1). Lastly, we hypothesized that the mediation pathway from Jarvis and Okonofua (2020) would replicate: the troublemaker label would mediate the effect of race on discipline. This pathway was not tested because there was no effect of race on discipline severity or troublemaker ratings.

We also tested for race effects on all of the other collected items. Consistent with discipline severity, there was no main effect of race on "feeling troubled," $F(1, 231) = 0.96$, $p = .328$, $d = 0.15$, indicating there were not significant differences in how troubled principals felt by the Black and White students' misbehaviors. There was a main effect of time, $F(1, 231) = 39.59$, $p < .001$, $d = 0.58$, indicating, from the first incident to the second, troubled feelings increased for principals towards all students. We did not observe a race by time interaction, $F(1, 231) = 1.09$, $p = .298$.

Principals then responded to questions about how the student misbehavior impacted their view of the student and projected how the student would act in the future. Principals rated the extent to which they thought the student's behavior was reflective of a pattern. There was no significant difference between Black students ($M = 3.04$, $SD = 1.07$) and White students ($M = 3.11$, $SD = 1.00$) in how likely their behavior was interpreted as being a pattern, $t(227) = 0.46$, $p = .649$, $d = 0.06$. Nor were there significant differences between Black students ($M = 2.11$, $SD = 0.95$) and White students ($M = 2.13$, $SD = 0.88$) in the extent to which they worried that the student's misbehavior would become a pattern across other students in the classroom, $t(223) = 0.15$, $p = .882$, $d = 0.02$. There were no significant differences between Black students ($M = 1.88$, $SD = 0.69$) and White students ($M = 1.99$, $SD = 0.70$) in the likelihood the

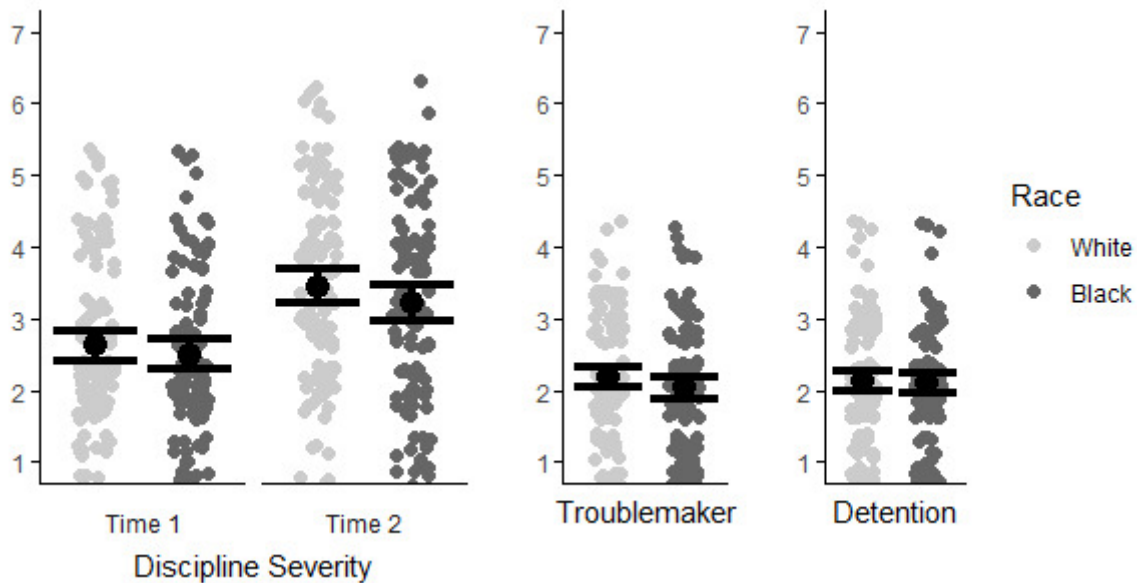


Figure 1. Mean ratings by race of the three predicted outcomes: discipline severity, troublemaker ratings and days of detention. The full detention scale is from 1-10. Error bars represent 95% confidence intervals.

student could be put on the right course, $t(230) = 1.28, p = .203, d = 0.17$ or to be suspended in the future, $t(228) = 0.82, p = .414, d = 0.11$ (Black students: $M = 1.61, SD = 0.64$; White students: $M = 1.69, SD = 0.68$).

Lastly, principals responded to questions about the extent to which students would benefit from non-punitive responses to the misbehavior. Parents of White students ($M = 3.28, SD = 0.80$) were seen as more likely than parents of Black students ($M = 3.01, SD = 0.96$) to be involved with the disciplinary process, $t(216) = 2.35, p = .020, d = 0.31$. There were no significant differences in how Black students ($M = 4.02, SD = 0.81$) and White students ($M = 3.99, SD = 0.91$) were expected to benefit from extracurriculars, $t(227) = 0.23, p = .815, d = 0.03$, or to benefit from talking to a counselor, $t(220) = 0.09, p = .930, d = 0.01$ (Black students: $M = 4.19, SD = 0.69$; White students: $M = 4.19, SD = 0.68$).

Of note, this sample is different from Jarvis and Okonofua (2020) in that it included principals in addition to assistant principals and sampled from schools of all age groups as opposed to limiting schools to middle and high schools. Analyses were also conducted with the full sample, only assistant principals, and only middle and high school principals. Analyses were universally null (see Tables S4 & S5 <https://osf.io/txrf7/>).

Moderation by Differences Between Principals' and Teachers' Roles

Hypothesis 2 predicted that differences in the extent to which principals and teachers get to know students, consider the classroom environment in disciplinary decisions, and consider school climate in disciplinary decisions would explain racial disparities in discipline. As a preliminary analysis, we tested if, on average, principals viewed their role as different than teachers on these items. Principals rated themselves less likely to get to know individual stu-

dents and more likely to consider the school climate and classroom environment when making disciplinary decisions (see Table 1).

To test our hypothesis, differences in perceptions of principals' and teachers' roles were calculated by subtracting the score a principal gave for teachers from the score they gave for themselves. This value was added as an interaction term to the manipulation of race of the student to test for moderation on discipline severity and troublemaker ratings. Our preregistration specified we would use discipline severity at time 2 if there was an interaction or an average of discipline severity if there was a main effect of race in the test of hypothesis 1. Since neither of these forecasted options were observed, we used the average of discipline severity.

The extent to which principals and teachers made connections with individual students did not moderate racial disparities in discipline severity, $b = 0.01, t(226) = 0.06, p = .952, 95\% CI [-0.31, 0.33]$, or troublemaker ratings, $b = -0.04, t(229) = -0.30, p = .761, 95\% CI [-0.28, 0.21]$ (see Table 2 for full regression tables). Differences between the extent to which principals and teachers take into account school climate when administering punishment did not moderate racial disparities in discipline severity, $b = 0.10, t(220) = 1.24, p = .218, 95\% CI [-0.06, 0.26]$, or troublemaker ratings, $b = 0.03, t(223) = 0.48, p = .633, 95\% CI [-0.10, 0.16]$. Lastly, differences between the extent to which principals and teachers take into account the classroom environment when administering punishment did not moderate racial disparities in discipline severity, $b = 0.07, t(220) = 1.01, p = .316, 95\% CI [-0.07, 0.20]$, or troublemaker ratings, $b = -0.02, t(223) = -0.40, p = .694, 95\% CI [-0.13, 0.09]$. As measured, differences in considerations between principals and teachers did not moderate racial disparities in discipline in this sample.

Table 1. Paired samples *t*-tests between the ratings of principals and perceptions of teachers on each factor.

	Principals		Teachers		<i>t</i> (<i>df</i>)	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Getting to Know Students	3.48	0.85	4.02	0.74	9.24 (233)	<.001	0.68
School Climate	4.93	1.67	3.96	1.79	8.54 (227)	<.001	0.56
Classroom Environment	5.17	1.50	4.11	1.94	8.06 (227)	<.001	0.61

Table 2. Regressions predicting discipline severity and troublemaker ratings from the interaction of student race and differences between principals' perceptions of their roles and teachers' roles.

	Discipline Severity				Troublemaker			
	<i>b</i>	<i>t</i> (<i>df</i>)	<i>p</i>	95% <i>CI</i>	<i>b</i>	<i>t</i> (<i>df</i>)	<i>p</i>	95% <i>CI</i>
Getting to Know Student								
intercept	2.95	35.49 (226)	<.001	2.78, 3.11	2.12	32.41 (229)	<.001	1.99, 2.25
race	-0.17	-1.05 (226)	.296	-0.50, 0.15	-0.17	-1.32 (229)	.189	-0.43, 0.09
principal-teacher difference	0.00	0.06 (226)	.954	-0.15, 0.16	0.01	0.19 (229)	.853	-0.11, 0.13
race*principal-teacher difference	0.01	0.06 (226)	.952	-0.31, 0.33	-0.04	-0.30 (229)	.761	-0.28, 0.21
School Climate								
intercept	2.95	36.90 (220)	<.001	2.79, 3.11	2.10	32.78 (223)	<.001	1.98, 2.23
race	-0.24	-1.51 (220)	.134	-0.56, 0.07	-0.16	-1.24 (223)	.217	-0.41, 0.09
principal-teacher difference	0.00	0.09 (220)	.929	-0.08, 0.08	0.02	0.50 (223)	.615	-0.05, 0.08
race*principal-teacher difference	0.10	1.24 (220)	.218	-0.06, 0.26	0.03	0.48 (223)	.633	-0.10, 0.16
Classroom Environment								
intercept	3.06	39.13 (220)	<.001	2.90, 3.21	2.19	34.93 (223)	<.001	2.07, 2.31
race	-0.22	-1.39 (220)	.165	-0.53, 0.09	-0.10	-0.82 (223)	.414	-0.35, 0.14
principal-teacher difference	-0.09	-2.70 (220)	.008	-0.16, -0.02	-0.07	-2.38 (223)	.018	-0.12, -0.01
race*principal-teacher difference	0.07	1.01 (220)	.316	-0.07, 0.20	-0.02	-0.40 (223)	.694	-0.13, 0.09

Differences between the other principal-teacher role items were tested as moderators for discipline severity and troublemaker ratings. There was no significant moderation for any of the other discipline consideration items: student character, student development, student academics or student behavior. Principal-teacher trust also did not moderate the effects (all *ps* > .107). We also checked if the raw principal role items moderated effects to test if simply differences in how principals view their role impact racial disparities with mixed results (see Tables S7-S10 <https://osf.io/3ezpt/>).

Exploratory Analyses with Perceived Socio-Economic Status

Due to relationship between race of the student and perceived socio-economic status, analyses were also run controlling for perceived socio-economic status of the school. As the likelihood the student was from a low-income neighborhood increased, principals rated students as more likely to be a troublemaker, *b* = 0.21, *t*(229) = 3.62, *p* < .001, 95% *CI* [0.10, 0.32], their behavior as more likely to be a pattern for the student, *b* = 0.18, *t*(230) = 2.44, *p* = .015, 95% *CI* [0.03,

0.32], and for their behavior to spread to students throughout the school, *b* = 0.15, *t*(226) = 2.25, *p* = .026, 95% *CI* [0.02, 0.27]. Greater perceptions that the student was from a low-income neighborhood also predicted more difficulty setting the student on the right course, *b* = 0.13, *t*(228) = 2.72, *p* = .007, 95% *CI* [0.04, 0.23], a greater likelihood of suspension in the future, *b* = 0.11, *t*(226) = 2.42, *p* = .016, 95% *CI* [0.02, 0.20], less parental involvement, *b* = -0.19, *t*(228) = -3.14, *p* = .002, 95% *CI* [-0.32, -0.07], and a greater likelihood that a school counselor would be beneficial, *b* = 0.12, *t*(221) = 2.48, *p* = .014, 95% *CI* [0.03, 0.22]. Perceived socio-economic status of the neighborhood did not moderate race effects except for parental involvement, *b* = -0.29, *t*(227) = -2.36, *p* = .019, 95% *CI* [-0.54, -0.05], in which to the extent to which the student was perceived to come from a neighborhood lower in socio-economic status and was Black, less parental involvement was expected (see Tables S2 & S3 for full results <https://osf.io/pwxed/>).

Exploratory Demographic Moderation

Given the pervasive null effects observed, we capitalized on our diverse sample of principals to test if racial dispar-

Downloaded from http://online.ucpress.edu/collabra/article-pdf/9/1/88326/827720/collabra_2023_9_1_88326.pdf by guest on 10 November 2024

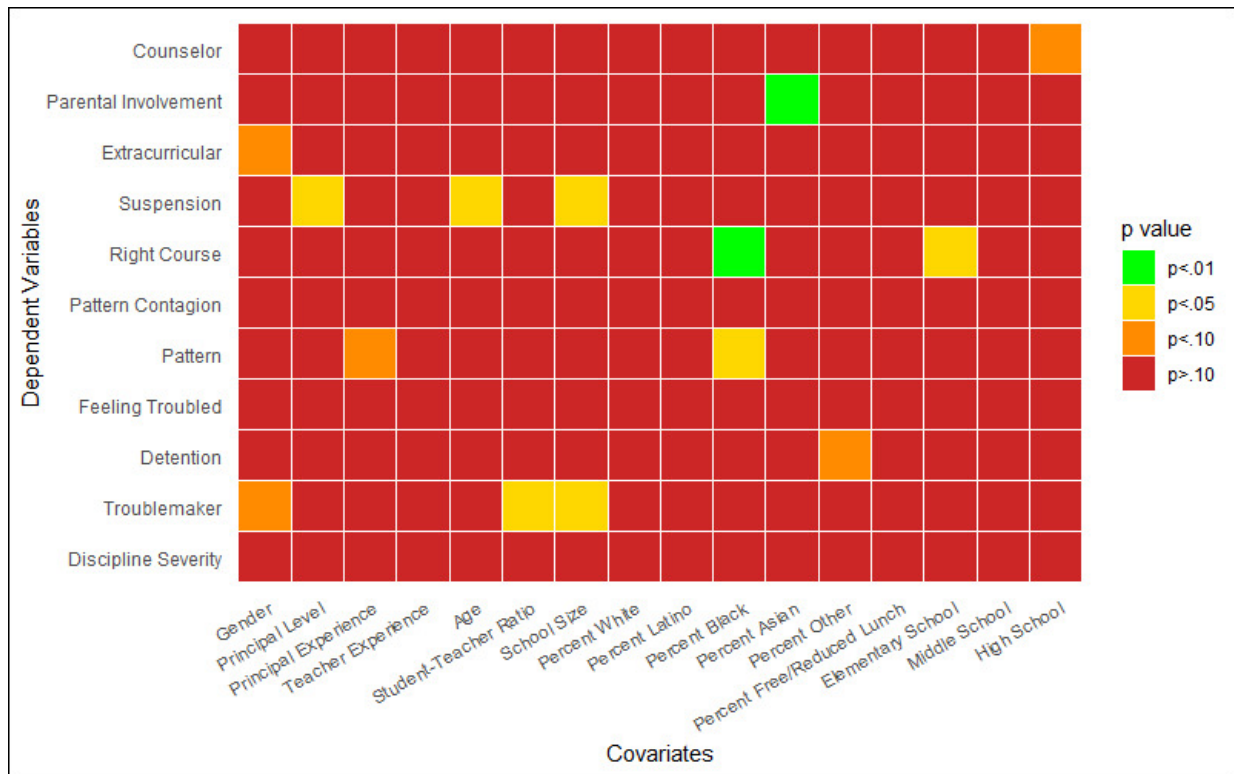


Figure 2. Heatmap of the results from the exploratory data analysis checking moderation of demographic factors on all outcomes.

ities would emerge systematically for particular principals or particular types of schools. We tested all of our collected demographic variables as moderators for all of the dependent variables, resulting in 176 tests. These demographic variables included principal gender, assistant principal vs principal, years of experience as a principal, years of experience as a teacher, principal age, student-teacher ratio, size of school, percent of student population White, percent of student population Latino, percent of student population Black, percent of student population Asian, percent of student population other, percent of student population on free or reduced lunch, and level of instruction: elementary school, middle school, and high school. Only 9 of these tests identified moderators that were significant at an alpha level of .05 with no discernable pattern. Thus, it seems likely that many of the significant interactions are due to chance. Figure 2 displays the multiverse of analyses between dependent variables and demographic moderators with a heat map.

We also did exploratory testing for if racial disparities in discipline could be observed via regional differences. For each dependent variable, a 2 (race: Black, White) by 5 (region: Northeast, Southeast, Midwest, West Coast, Southwest) ANOVA was run testing the effects of race, region, and their interaction on each outcome. There were no significant interactions of race and region for any outcome variable. There were some instances where there were effects of region that indicated one region had higher or lower ratings on a particular outcome variable than one other region. These differences were not systematic. In

Table 3, means and standard deviation for each outcome variable by race and region are reported for transparency.

Discussion

In a test of the pervasiveness of racial disparities in discipline by school administrators, principals and assistant principals from across the United States rated how they would discipline Black or White students after the same misbehaviors. Little support was observed for our confirmatory hypotheses. Principals and assistant principals did not significantly endorse more severe levels of discipline, rate students as more likely to be troublemakers, or endorse more days of detention for Black as compared to White students. There was also no moderation by the extent to which principals formed relationships with individual students or the extent to which they considered the safety and wellbeing of the classroom or school when allocating discipline.

Outside the bounds of our hypotheses, principals assumed socio-economic status from the race of the students. While the experimental manipulation of race did not predict any of our main outcomes, principals' projections of the social class of the student's neighborhood did. The extent to which principals thought the students were from lower class neighborhoods, the more they negatively perceived the students, e.g., the more likely they were to think the student was a troublemaker and to endorse more days of detention, over and above the experimental manipulation. Effects of socio-economic status on discipline have mixed evidence using data from school districts (R. J. Skiba et al., 1997, 2002, 2014). In this experimental paradigm, as-

Table 3. Means (and standard deviations) by race of student and region of the United States the principal was from.

	Northeast		Southeast		Midwest		West Coast		Southwest	
	White	Black	White	Black	White	Black	White	Black	White	Black
Discipline Severity	3.38 (1.14)	3.29 (1.17)	3.36 (1.00)	2.55 (0.93)	2.95 (0.96)	3.07 (0.93)	2.43 (0.88)	2.44 (0.89)	2.97 (1.25)	2.94 (1.05)
Troublemaker	2.38 (0.92)	2.24 (1.09)	2.38 (0.74)	1.87 (0.85)	2.11 (0.81)	2.23 (0.73)	1.83 (0.78)	1.82 (0.81)	2.13 (0.74)	1.89 (0.78)
Detention	2.10 (0.89)	2.06 (0.75)	2.20 (0.89)	2.34 (0.81)	2.03 (0.87)	1.97 (0.51)	1.87 (0.63)	1.88 (0.70)	2.43 (0.85)	2.22 (0.97)
Feeling Trouble	3.91 (1.34)	3.49 (1.41)	3.47 (0.99)	2.98 (0.99)	3.48 (1.05)	3.47 (1.11)	2.95 (1.03)	2.89 (1.07)	3.34 (1.18)	3.48 (1.45)
Pattern	3.29 (0.85)	3.67 (1.03)	3.24 (0.94)	3.03 (0.84)	3.16 (0.96)	2.91 (1.09)	2.91 (1.08)	3.06 (1.30)	2.73 (1.22)	2.56 (1.01)
Pattern Contagion	2.14 (0.91)	2.18 (1.24)	2.50 (0.89)	2.13 (0.92)	1.86 (0.83)	2.20 (0.87)	1.91 (0.75)	1.88 (0.86)	2.33 (0.90)	1.89 (1.05)
Right Course	2.10 (0.70)	2.11 (0.76)	2.10 (0.77)	1.68 (0.54)	1.97 (0.74)	1.89 (0.72)	1.91 (0.67)	2.18 (0.73)	1.87 (0.64)	1.44 (0.53)
Suspension	1.76 (0.70)	1.72 (0.67)	2.00 (0.77)	1.79 (0.56)	1.57 (0.69)	1.59 (0.70)	1.50 (0.51)	1.41 (0.62)	1.60 (0.63)	1.44 (0.53)
Extracurricular	4.19 (0.87)	4.00 (0.61)	3.90 (0.89)	4.03 (0.84)	3.97 (1.03)	4.09 (0.82)	3.91 (0.90)	4.00 (0.94)	4.00 (0.85)	3.62 (0.92)
Parent Involvement	3.30 (0.73)	2.72 (0.67)	3.52 (0.81)	3.10 (1.08)	3.22 (0.80)	2.97 (0.98)	3.13 (0.81)	3.29 (0.99)	3.53 (0.74)	2.78 (0.97)
Counselor	4.48 (0.60)	4.47 (0.62)	4.19 (0.75)	4.21 (0.62)	4.11 (0.63)	3.97 (0.77)	4.22 (0.67)	4.41 (0.51)	3.93 (0.73)	4.11 (0.78)

assumptions of socio-economic status may have been an alternative avenue by which participants could demonstrate more comfortably their biases as opposed to the race manipulation or at least was a more central factor in determining the pervasiveness of misbehavior and subsequent discipline allocations.

The lack of replication is noteworthy. It is also important to note key differences between the present study and previous research. A primary difference is in the adaptation of the procedure. The major thrust of the present study was to explain the difference in the discipline process for principals due to the race of the student. For this reason, a new task – series of targeted questions – was inserted directly before the procedure from Jarvis and Okonofua (2020). Subsequent research on wise-interventions and policy suggest that when an educator is asked a series of questions about their role in building meaningful relationships with students, that procedure can serve as an intervention that affirms the participant, reduces punitive discipline for Black students, and mitigates racial disparities in discipline decisions (R. J. Skiba et al., 1997, see also Okonofua et al., 2020 for effect of reflecting on beliefs reducing bias in a different domain). Taken together, the new task may have served as an intervention that mitigated the effects of racial bias on discipline decisions. Future research should explore if asking principals to reflect on their role reduces racial disparities in discipline.

An additional key difference between the present study and Jarvis and Okonofua (2020) is in the sample used between the two studies. While the present sample included principals and assistant principals from across school levels and districts throughout the United States, the sample in Jarvis and Okonofua (2020) only included middle and high

school assistant principals from one large and diverse school district. However, when excluding elementary school principals from the sample, null effects persisted. Similarly, administrative position, school-level of instruction, or any of the other demographic variables collected did not moderate any of our key outcome variables, indicating it is not likely that the differences in sample composition solely explain the failure to replicate. Past studies found relationships between regional bias and racial disparities in discipline (Chin et al., 2020; Riddle & Sinclair, 2019). We did not observe the same regional differences, though there were key differences between the designs of the studies. The past studies had finer measurements of region (county vs. several states) and did not use scores from individuals but rather benefited from a “wisdom of crowds” approach (Payne et al., 2017).

Differences in recruitment procedures and participation rates could have impacted our results. In this sample, principals were cold-emailed with a 5% participation rate whereas in (Jarvis & Okonofua, 2020) the entire population of assistant principals in the school district were given the study materials as part of a professional development workshop. The present sample may be more likely to have been impacted by selective participation and that principals who were less biased, were more conscientious, or worked at higher achieving schools with fewer behavioral problems were more likely to participate.

Though the findings from Jarvis and Okonofua (2020) illustrate that there are places in which principals propagate racial disparities in discipline, this geographically diverse sample was unable to provide direction as to where or how racial disparities may emerge. It is possible that principals do show racial bias when allocating discipline, but

the present study did not collect the factors that predict such disparities. While this experiment provides inconsistent results for the extent to which principals contribute to disproportionate discipline for Black students, experimental effects with teachers are more consistent (Okonofua & Eberhardt, 2015; Perez & Okonofua, 2022; Rucinski et al., 2023). It could be possible that on average principals do not display racial bias in discipline. However, because teachers disproportionately refer Black students to the principal's office, Black students are more likely to receive exclusionary discipline (R. J. Skiba et al., 2002, 2014). Rather than principals adding to the effects of systemic bias, they may perpetuate disparities present earlier in the school-to-prison pipeline.

Research with principals is necessary to establish a better understanding of how the school-to-prison pipeline might be perpetuated across positions in schools and over time. Principals serve as important gatekeepers, providing or limiting access to school, and have a great influence over procedures for large groups of students. Studying where and how principals perpetuate racial disparities will be an important step in creating steps to eliminate the school-to-prison pipeline. While Jarvis and Okonofua (2020) demonstrated that there are schools and areas in which principals perpetuate racial disparities, the present research suggests more research is necessary in order to determine how pervasive these biases are and where these biases are more likely to occur.

.....

Contributions

Contributed to conception and design: SNJ, ZEF, JAO
 Contributed to acquisition of data: SNJ, ZEF
 Contributed to analysis and interpretation of data: SNJ
 Drafted and/or revised the article: SNJ, ZEF, JAO
 Approved submitted version for publication: SNJ, ZEF, JAO

Acknowledgments

The authors would like to thank Serena Chen, Iris Mauss, Laura Kray and Don Moore for feedback on earlier versions of this manuscript and Michael Ruiz, Shayna Howlett, My Dao, Vikktor Abat, and May Xue for assistance with data collection.

Funding Information

The first author is supported by a National Science Foundation Graduate Research Fellowship DGE 1752814.

Competing Interests

The authors declare no conflicts of interest.

Data Accessibility Statement

All the stimuli, materials, participant data, and analysis scripts can be found on this paper's project page on OSF (<https://osf.io/j87zk/>).

Submitted: March 20, 2023 PDT, Accepted: September 15, 2023 PDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Arcia, E. (2006). Achievement and Enrollment Status of Suspended Students: Outcomes in a Large, Multicultural School District. *Education and Urban Society*, 38(3), 359–369. <https://doi.org/10.1177/0013124506286947>
- Bacher-Hicks, A., Billings, S. B., & Deming, D. J. (2019). *The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime* (Working Paper No. 26257; Working Paper Series). National Bureau of Economic Research. <https://doi.org/10.3386/w26257>
- Chin, M. J., Quinn, D. M., Dhaliwal, T. K., & Lovison, V. S. (2020). Bias in the air: A nationwide exploration of teachers' implicit racial attitudes, aggregate bias, and student outcomes. *Educational Researcher*, 49(8), 566–578. <https://doi.org/10.3102/0013189x20937240>
- Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, 31(1), 92–109. <https://doi.org/10.1016/j.econedurev.2011.09.001>
- Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement? *Canadian Journal of Economics/Revue Canadienne d'économique*, 47(2), 634–663. <https://doi.org/10.1111/caje.12086>
- Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., & Booth, E. A. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement*. Council of State Governments Justice Center. <https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=266653>
- Ferguson, Z. E., Jarvis, S. N., Antonopolis, S., & Okonofua, J. A. (2023). Principal beliefs predict responses to individual students' misbehavior. *Educational Researcher*, 52(5), 315–319. <https://doi.org/10.3102/0013189x231158389>
- Goldring, E., Huff, J., May, H., & Camburn, E. (2008). School context and individual characteristics: What influences principal practice? *Journal of Educational Administration*, 46(3), 332–352. <https://doi.org/10.1108/09578230810869275>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Hornig, E. L., Klasik, D., & Loeb, S. (2010). Principal's Time Use and School Effectiveness. *American Journal of Education*, 116(4), 491–523. <https://doi.org/10.1086/653625>
- Jarvis, S. N., & Okonofua, J. A. (2020). School Deferred: When Bias Affects School Leaders. *Social Psychological and Personality Science*, 11(4), 492–498. <https://doi.org/10.1177/1948550619875150>
- Kim, C. Y., Losen, D. J., & Hewitt, D. T. (2010). *The School-to-Prison Pipeline: Structuring Legal Reform*. NYU Press.
- Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). How Leadership Influences Student Learning. Review of Research. In *Wallace Foundation, The*. The Wallace Foundation, Five Penn Plaza, 7th Floor, New York, NY 10001.
- National Teacher and Principal Survey (NTPS)*. (n.d.). National Center for Education Statistics. Retrieved February 26, 2021, from https://nces.ed.gov/surveys/ntps/tables/ntps1718_19110501_a1s.asp
- Okonofua, J. A., & Eberhardt, J. L. (2015). Two Strikes Race and the Disciplining of Young Students. *Psychological Science*, 26(5), 617–624. <https://doi.org/10.1177/0956797615570365>
- Okonofua, J. A., Perez, A. D., & Darling-Hammond, S. (2020). When policy and psychology meet: Mitigating the consequences of bias in schools. *Science Advances*, 6(42), eaba9479. <https://doi.org/10.1126/sciadv.aba9479>
- Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, 28(4), 233–248. <https://doi.org/10.1080/1047840x.2017.1335568>
- Perez, A. D., & Okonofua, J. A. (2022). The good and bad of a reputation: Race and punishment in K-12 schools. *Journal of Experimental Social Psychology*, 100, 104287. <https://doi.org/10.1016/j.jesp.2022.104287>
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75(3), 811–832. <https://doi.org/10.1037/0022-3514.75.3.811>
- Riddle, T., & Sinclair, S. (2019). 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. <https://doi.org/10.1073/pnas.1808307116>
- Rocque, M., & Paternoster, R. (2011). Understanding the antecedents of the “School-to-Jail” link: The relationship between race and school discipline. *Journal of Criminal Law and Criminology*, 101(2), 633–666.
- Rucinski, C. L., Mandalaywala, T. M., & Tropp, L. R. (2023). Escalation effects in teacher perceptions of classroom behavior in a US context: The intersecting roles of student race, gender, and behavior severity. *Social Psychology of Education*, 1–20.
- Ryan, C. S., Hunt, J. S., Weible, J. A., Peterson, C. R., & Casas, J. F. (2007). Multicultural and Colorblind Ideology, Stereotypes, and Ethnocentrism among Black and White Americans. *Group Processes & Intergroup Relations*, 10(4), 617–637. <https://doi.org/10.1177/1368430207084105>
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, 26(5), 559–569. <https://doi.org/10.1177/0956797614567341>

- Skiba, R., Edl, H., & Rausch, M. (2007). How do principals feel about discipline? The Disciplinary Practices Survey. In *Annual Convention of the American Educational Research Association, Chicago, IL*.
- Skiba, R. J., Chung, C.-G., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014). Parsing Disciplinary Disproportionality: Contributions of Infraction, Student, and School Characteristics to Out-of-School Suspension and Expulsion. *American Educational Research Journal*, *51*(4), 640–670. <http://doi.org/10.3102/0002831214541670>
- Skiba, R. J., Horner, R. H., Chung, C.-G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race Is Not Neutral: A National Investigation of African American and Latino Disproportionality in School Discipline. *School Psychology Review*, *40*(1), 85–107. <https://doi.org/10.1080/02796015.2011.12087730>
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The Color of Discipline: Sources of Racial and Gender Disproportionality in School Punishment. *The Urban Review*, *34*(4), 317–342. <http://doi.org/10.1023/a:1021320817372>
- Skiba, R. J., Peterson, R. L., & Williams, T. (1997). Office Referrals and Suspension: Disciplinary Intervention in Middle Schools. *Education and Treatment of Children*, *20*(3), 295–315.
- US Department of Education Office for Civil Rights. (2016). *2013-14 Civil Rights Data Collection: A First Look*. 13.

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/88326-a-test-of-racial-disparities-in-principals-disciplinary-decisions-in-a-national-sample/attachment/182111.docx?auth_token=5PSQoXK0jwiFuKqCc4Pp

Cover Letter

Download: https://collabra.scholasticahq.com/article/88326-a-test-of-racial-disparities-in-principals-disciplinary-decisions-in-a-national-sample/attachment/182112.pdf?auth_token=5PSQoXK0jwiFuKqCc4Pp
