Preprint

The Racialized Classroom: The Influence of Classroom Composition on Teachers' Decisions to Use Disciplinary Measures

© The Author(s) 2025

Iniobong Essien[1], Birte Siem[1], Lysann Zander[2], and Anette Rohmann[3]

Abstract

Does racialized classroom composition affect how teachers deal with disruptive behavior? In two preregistered experiments with preservice and in-service teachers in Germany (total *N* = 787), we examined whether participants respond to disruptive behavior differently depending on the percentage of students perceived as Arab in a classroom. When confronted with disruptive behavior of a fictitious student, both preservice and in-service teachers suggested more severe disciplinary measures in classrooms with a higher share of students perceived as Arab (vs. only German). Preservice teachers in Study 1 also (a) reported feeling more irritated by the students' behaviors and (b) perceived the entire classroom context as more challenging when there was a higher proportion of students perceived as Arab, but these effects were not replicated among in-service teachers in Study 2. The studies provide initial evidence for an underexplored mechanism that may exacerbate educational inequalities by systematically shaping learning environments students experience.

Keywords

racialized classroom composition; context-based disparities; school discipline

Last modified: 4/11/2025

[1]Leuphana Universität Lüneburg, Germany
[2]Leibniz Universität Hannover, Germany
[3]FernUniversität in Hagen, Germany

Corresponding Author:
Iniobong Essien, Department of Social and Organisational Psychology of Social Work, Leuphana Universität Lüneburg, 21335 Lüneburg, Germany. E-mail: iniobong.essien@leuphana.de.

Classrooms are composed in myriad ways. They are comprised of students with varying ability levels, socioeconomic backgrounds, or gender ratios. Here, we focus on the racialized[1] composition of classrooms—i.e. the proportion of students from racialized groups in a classroom—and its potential impact on teachers' decision-making in response to disruptive behavior.

Patterns of segregation help explain dramatically different racialized classroom compositions from one location to another. In the United States, because of racialized segregation, approximately 69 percent of Black students attended schools where students from racialized groups were the predominant groups (García, 2020). Data from Germany similarly suggest pervasive levels of racialized segregation in that 1 in 4 adolescents with a so-called migration background were enrolled in a school where students with a migration background were in the majority (Fincke & Lange, 2012). These examples illustrate how segregation shapes the composition of educational contexts, creating classrooms with substantially different racialized compositions.

A considerable body of research on *racialized school composition* suggests that school personnel employ more severe disciplinary measures in response to disruptive behavior in schools with higher proportions of students from racialized groups (Anyon et al., 2014; Edwards, 2016; Payne & Welch, 2010; Skiba et al., 2014; Welch & Payne, 2010; Welsh & Little, 2018). Relying on a survey of US public schools, Welch and Payne (2010) observed that schools were more likely to respond punitively to disruptive behavior (e.g., suspending students; calling the police) when a higher percentage of Black students were enrolled. Schools with higher percentages of Black students were also less likely to respond mildly to disruptive behavior (e.g., sending students to counselors). Lastly, Skiba et al. (2014) found that severe disciplinary measures were more frequent in schools with higher percentages of Black students, even when controlling for student behaviors that gave rise to the disciplinary responses (see also Anyon et al., 2014). In sum, school personnel responds differently to disruptive behavior based on the racialized school composition; students in racialized school contexts face more severe disciplinary responses compared to their peers in non-racialized school contexts, leading to *context-based disparities*.

Additionally, research suggests that *racialized classroom composition* is related to systematic differences in teacher behaviors, such as varying emotional support, instructional support, and classroom management. In one study, Osei-Twumasi and Pinetta (2023) observed based on a large dataset of classroom videos that White teachers in classrooms with higher percentages of Black students provided less emotional support, lower-quality instructional support, and were less able to effectively manage their classrooms. The potential impact of such different student experiences based on the classroom composition are worth noting (e.g., Childs & Wooten, 2023). Theorizing about the consequences of teacher biases points to potential vicious cycles between teachers and students (Okonofua et al., 2016). Students who perceive their teachers as biased might feel they belong less, develop mistrust, and lose school motivation. Importantly, this erosion of trust may increase the likelihood that students act out, leading to self-reinforcing cycles of escalated discipline and more frequent disruptive behaviors (Okonofua et al., 2016). Such negative dynamics between teachers and students underscore the relevance of context-based disparities in school discipline.

To our knowledge, studies have not yet experimentally investigated effects of the racialized classroom composition on teachers' disciplinary decisions. Instead, previous experimental research has focused on teachers' decision-making regarding individual students. This research suggests that, following identical disruptive behavior, teachers suggest more severe discipline for students from racialized groups than for their non-racialized peers. For example, Okonofua and colleagues (2015) presented scenarios of students disrupting class and varied the students' names to evoke teachers' mental images of Black or White students. Teachers ascribed more negative stereotypes to Black students, leading to escalated disciplinary decisions for minor infractions. However, whether and how racialized classroom composition affects teachers'

[1] The term "racialized" refers to socially constructed, stigmatized "racial" and "ethnic" groups, such as Black or Arab, and highlights that human "races" do not exist. The term is not supposed to suggest that an umbrella term should replace the names and lived experiences of individual marginalized groups; in contrast, the term „racialized" acknowledges that „racism" is at the heart of the racialization process that shapes how different marginalized groups are constructed, perceived, and treated.

disciplinary decisions has not yet been experimentally investigated.

Because of the correlational nature of previous research, important gaps remain in our understanding of context-based disparities. Notably, it is unclear whether racialized classroom composition *causally* influences teachers' decision-making. When trying to make causal claims from observational data, researchers must carefully consider potential confounding factors that could affect both racialized classroom composition and teachers' decision-making, such as poverty levels within schools (e.g., Skiba et al., 2014). However, measuring and controlling for confounding factors is notoriously difficult and complicated by potential unknown or unobserved confounding variables (Bailey et al., 2024). Experiments offer a complementary approach, allowing researchers to isolate effects of racialized classroom composition on teachers' decision-making. The present research addresses an important gap by experimentally investigating the impact of racialized classroom composition on teachers' decisions to use disciplinary measures.

The Present Research

The present research examined whether teachers' decisions to use discipline in response to minor infractions differed depending on the racialized classroom composition. Specifically, we investigated in a German context whether students in classrooms with many students with Arabic names were treated differently than those in classrooms with only students with German names. People perceived as Arab are severely stigmatized in Germany (Stürmer et al., 2019), and face substantial disadvantages in the educational system (e.g., Henschel et al., 2019). We conducted two online experiments with preservice teachers (Study 1) and in-service teachers (Study 2). We hypothesized that when confronted with descriptions of disruptive behavior in the classroom, participants would suggest more severe disciplinary measures if classrooms had a higher proportion of students perceived as Arab. Both studies were preregistered at the Open Science Framework (Study 1: https://osf.io/kqjae/?view_only=3b0706591ee8407b95712b84c27714ba; Study 2: https://osf.io/5pr7e/?view_only=ad7ca72814df4e3d8d9d79566309956e).

Study 1

The aim of Study 1 was to investigate the effects of the racialized classroom composition on preservice teachers' decisions to use disciplinary measures in response to disruptive behaviors.

Method

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (see Simmons, Nelson, & Simonsohn, 2012). All materials, data, and codebooks are accessible at https://osf.io/xkers/?view_only=87444e8f0eaf4b58882eda1c0f23d340.

Participants. We aimed at recruiting $N = 278$ pre-service teachers from a large university in Northern Germany. The targeted sample size was estimated based on a power analysis, in which we assumed a small-to-medium effect of $d = .30$, based on differences in decisions to use discipline between students with Arabic and German names observed in our own preliminary work (reference anonymized). Given $\alpha = .05$, $1 - \beta = .80$ in an independent t-test (one-tailed), a sample size of at least $N = 278$ is needed to detect an effect of that magnitude. To incentivize participation in the study, participants could enter a lottery to win a prize (i.e., two tablet computers). In addition, we advertised the study via e-mail in teaching and education study programs at universities in 10 large German cities and their surrounding areas. Three hundred and eighty-four participants (298 female, 83 male, 2 non-binary), who provided informed consent after having been fully debriefed at the end, comprised our final sample. Participants' mean age was 24.81 ($SD = 5.71$), and a majority of the sample ($n = 322$) self-categorized as White (10 Turkish or Turkish-German, 1 Arab or Arab-German, 1 Black or Afro-German, 42 other, 7 no response). Importantly, most participants were preservice teachers ($n = 299$) or preservice teachers with a focus on special needs education ($n = 40$) , and a substantial proportion of the sample ($n = 89$) reported work experience in school contexts (e.g., as teachers or teachers in training).

Design and procedure. The study was an online experiment that followed a between-subjects design with two conditions. After being presented with a photograph of a school building, participants were asked to imagine themselves teaching at that school. Depending on the experimental condition, participants

were then presented with one of two class registers that varied in terms of the perceived classroom composition, using first names that were perceived as stereotypically Arabic and German.[2] Participants in the Arabic names condition were presented with a class register with 20 first names, of which 12 names (60%) were Arabic first names (see Table S3 in the Supplement). In the German names condition, all 20 names (100%) were German first names. In both experimental conditions, 12 names (60%) were male names and 8 names (40%) were female names. Participants were asked to memorize the students' names and could proceed to the next page after 20 sec. Next, participants were presented with a sequence of five scenarios describing a student misbehaving in school, translated and adapted from previous research (Jarvis & Okonofua, 2020; Okonofua & Eberhardt, 2015)—see Table S4 in the Supplement. After each scenario, participants completed measures assessing how they perceived and would respond to the students' behavior. After responding to additional measures, a manipulation check, and demographic questions, participants were fully debriefed, asked for their consent to using their data, and indicated whether they had participated seriously.

Measures

Perceptions of student and misbehavior. After each scenario, participants responded to two items adapted from Okonofua and Eberhardt (2015) to assess their perceptions of the student and the misbehavior (i.e., "How serious was the misbehavior of the student?"; "How irritated do you feel by the student?"). Responses were provided on scales ranging from 1 (not at all) to 7 (very). The correlations between these two items ranged from .57 to .73. To create a single index of the extent to which participants perceived the student and the misbehavior as irritating across situations, mean responses to the two items were aggregated across all five scenarios.

Suggested severity of disciplinary measures. After each scenario, participants rated how severely the student should be disciplined on an item adapted from Okonofua and Eberhardt (2015), using a scale from 1 (not at all) to 7 (very). We aggregated item responses across all five scenarios to create a single index of the suggested severity of disciplinary measures (ω = .79).

Perceptions of the classroom context. After all five scenarios, participants reported their perceptions of the entire classroom context, responding to three items (e.g., "How annoyed are you by this class?"; ω = .86). Additionally, participants responded to an item about their felt enthusiasm about teaching that class ("To what extent do you feel enthusiasm to teaching this class?").[3]

Perceived classroom composition. Participants were asked to estimate the percentage of students with Arabic names in the classroom by typing a number between 0 and 100, designed as a manipulation check.

Additional measures. At the end of the survey, participants were asked to list all names they were able to recall from the class register. Additionally, participants were asked whether they had imagined a specific school type (e.g., comprehensive school, intermediate school, grammar school) while responding to the survey.[4]

Results

We used R (Version 4.4.2; R Core Team, 2023)[5], for all our analyses. See Table 1 for the correlation matrix of study measures.

Manipulation check. First, we tested for the effect of our manipulation by examining whether the perceived percentage of students with Arabic names in the classroom differed between experimental conditions. Indeed, participants in the Arabic names condition

[2] Based on pretesting, we selected Arabic and German first names that were high or low in perceived Arab stereotypicality, respectively, as well as German first names that were associated with moderate perceived socioeconomic status.

[3] An exploratory factor analysis suggested that all four items loaded onto a single factor, but internal consistency of the four-item scale was substantially decreased (ω = .69) when the item about felt enthusiasm was included. For the sake of transparency, we conducted an additional set of analyses using an index that included all four items (see Supplement).

[4] The German educational system stratifies students into different school pathways based on teacher recommendation.

[5] We, furthermore, used the R-packages *corx* (Version 1.0.7.2; Conigrave, 2023), *cowplot* (Version 1.1.3; Wilke, 2024), *data.table* (Version 1.16.4; Dowle & Srinivasan, 2023), *here* (Version 1.0.1; Müller, 2020), *jmv* (Version 2.5.6; Selker, Love, & Dropmann, 2023), *lavaan* (Version 0.6.19; Rosseel, 2012), *lmerTest* (Version 3.1.3; Kuznetsova, Brockhoff, & Christensen, 2017), *MBESS* (Version 4.9.3; Kelley, 2023), *papaja* (Version 0.1.3; Aust & Barth, 2022) and *tidyverse* (Version 2.0.0; Wickham et al., 2019).

estimated a higher percentage of students with Arabic names in the classroom (M = 43.29, SD = 17.72) than participants in the German names condition (M = 11.54, SD = 13.45), $t(289.39) = 19.07$, $p < .001$, $d_s = 2.02$, 95% CI [1.72; 2.21].

Table 1. Zero-order correlations of study measures for the Arabic names condition (below the diagonal) and German names condition (above the diagonal) in Study 1.

Variable	1	2	3	4	5
1. Suggested Discipline	-	.83***	.55***	-.29***	.23***
2. Perceptions Student	.84***	-	.62***	-.29***	.19**
3. Perceptions Class	.63***	.69***	-	-.52***	.22**
4. Enthusiasm	-.24**	-.29***	-.60***	-	-.20**
5. Perceived Composition	.15	.16*	.16*	-.08	-

Note: * $p < .05$; ** $p < .01$; *** $p < .001$; Suggested Discipline = Suggested disciplinary measures; Perceptions Student = perceptions of student and misbehavior; Perceptions Class = perceptions of the entire classroom context; Enthusiasm = felt enthusiasm about teaching the class; Perceived Composition = perceived classroom composition.

Perceptions of student and misbehavior. Next, we examined whether participants' perceptions of the student and the misbehavior differed between experimental conditions. We observed that participants in the Arabic names condition judged the misbehavior as more serious and felt more irritated by the student (M = 4.05, SD = 0.97) than participants in the German names condition (M = 3.79, SD = 0.94), $t(347.16) = 2.63$, $p = .009$, $d_s = 0.27$, 95% CI [0.07; 0.47]—see Fig. 1, Panel A.

Suggested severity of disciplinary measures. Next, we examined whether participants' suggested severity of disciplinary measures differed between experimental conditions. Consistent with our pre-registered hypothesis, participants in the Arabic names condition suggested the use of more severe disciplinary measures (M = 3.59, SD = 1.17) than participants in the German names condition (M = 3.37, SD = 0.99), $t(317.31) = 1.96$, $p = .026$ (one-sided, as pre-registered), $d_s = 0.20$, 90% CI [-∞; 0.37], to respond to the misbehavior—see Fig. 1, Panel B.[6]

Perceptions of the classroom context. We next examined whether participants' perceptions of the classroom context differed between experimental conditions. We observed that participants in the Arabic names condition perceived the classroom context as more challenging and more negatively (M = 3.77, SD = 1.25) than participants in the German names condition (M = 3.35, SD = 1.14), $t(335.84) = 3.40$, $p < .001$, $d_s = 0.35$, 95% CI [0.15; 0.55]—see Fig. 1, Panel C.

We next examined whether the extent to which participants felt enthusiasm to teaching the class differed between experimental conditions. We only observed a numerical difference between the Arabic names condition (M = 4.06, SD = 1.33) and the German names condition (M = 4.28, SD = 1.25), which was not statistically significant, $t(340.72) = -1.66$, $p = .097$, $d_s = -0.17$, 95% CI [-0.37; 0.03]—see Fig. 1, Panel D.

Mediation analysis: classroom composition → perceptions of student and misbehavior → severity of discipline. Lastly, we explored whether the extent to which participants felt irritated by the student and the misbehavior mediated the effects of the classroom composition on the suggested severity of disciplinary measures. In other words, we examined to what extent the effects of the classroom composition on disciplinary decisions were explained by participants' perceptions of the student and the misbehavior. We used the "lavaan" package in R to conduct a mediation analysis with 1,000 bootstrap samples with classroom composition as the independent variable, perceptions of the student and the misbehavior as the mediator, and participants' suggested severity of

[6] Based on comments on an earlier version of the manuscript, we also tested this hypothesis by fitting a linear mixed-effects model, which yielded similar results to our pre-registered analysis (see Supplement).

disciplinary measures as the dependent variable. First, we tested the total effect of classroom composition on the suggested severity of disciplinary measures, b = 0.22, SE = 0.12, p = .027, 95% CI [0.00, 0.45]. To maintain consistency with our one-sided test of the total effect reported earlier, we report the adjusted (i.e., halved) p-value for the total effect. We then tested the direct effect of classroom composition on perceptions of the student and the misbehavior (the mediator), which was statistically significant, b = 0.26, SE = 0.10, p = .010, 95% CI [0.07, 0.46]. Next, we regressed the suggested severity of discipline (the dependent variable) on the perceptions of the student and the misbehavior (the mediator) and on the classroom composition (the independent variable). Perceptions of the student and the misbehavior significantly predicted the suggested severity of disciplinary measures, b = 0.94, SE = 0.03, p < .001, 95% CI [0.88, 1.00], whereas the effect of classroom composition on the suggested severity of discipline became non-significant, b = -0.02, SE = 0.06, p = .737, 95% CI [-0.14, 0.10]. Importantly, the indirect effect of classroom composition on the suggested severity of discipline via perceptions of the student and the misbehavior was statistically significant, b = 0.24, SE = 0.10, p = .011, 95% CI [0.06, 0.44]. Thus, perceptions of the student and the misbehavior fully mediated effects of the classroom composition on the suggested severity of disciplinary measures.

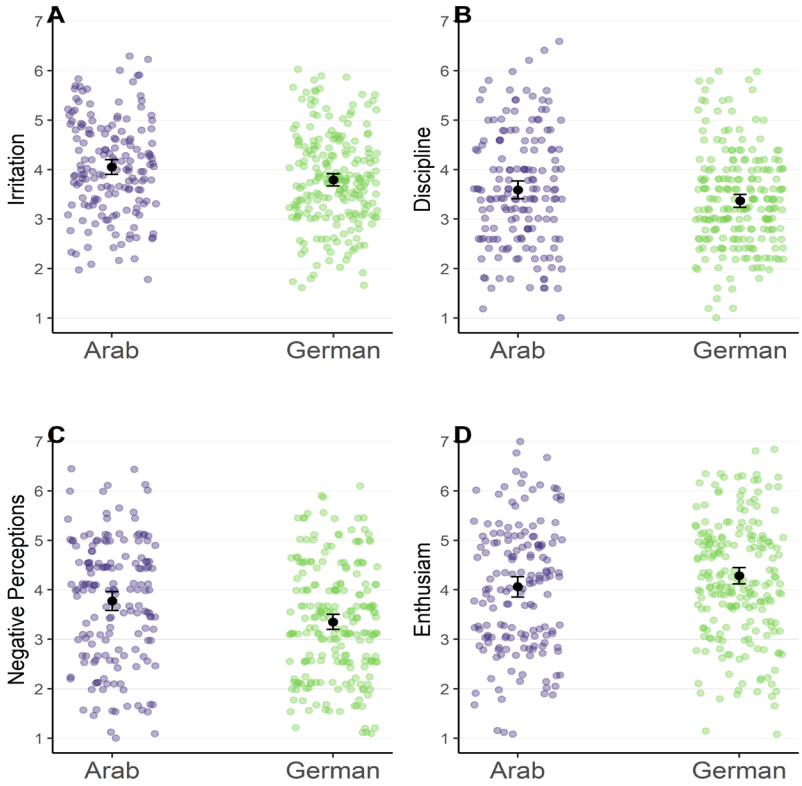

Figure 1. Jittered dot plot of preservice teachers' responses by experimental condition, with upper left panel (A) displaying aggregated perceptions of the severity of the misbehavior and feeling irritated by the student, upper right panel (B) displaying the suggested severity of disciplinary measures, lower right panel (C) displaying negative perceptions of the classroom environment, and lower right panel (D) displaying felt enthusiasm about teaching in the class. Black dots represent means and error bars represent 95% confidence intervals.

Discussion

Findings of Study 1 suggest that preservice teachers who were presented with descriptions of disruptive behaviors in a classroom with many students with Arabic names reported feeling more irritated by the behavior and suggested more severe disciplinary measures than those in a classroom with students with German names. Participants also

perceived the entire classroom context as more challenging and negatively when there were many students with Arabic names. However, self-reported enthusiasm in teaching the class did not differ between experimental conditions, suggesting that participants did not feel less enthusiastic about teaching in a classroom with many students with Arabic names. Lastly, we found that felt irritation mediated effects of the classroom composition on disciplinary decisions.

Study 2

One potential limitation of Study 1 was its reliance on a sample with limited teaching experience. Research suggests that some work experiences might be related to more pronounced intergroup biases (Kemme et al., 2020), whereas other research suggests that expertise might sometimes reduce intergroup biases (Sim et al., 2013). To address potential limitations with regard to teaching experience and to test the generalizability of our findings, we conducted a close replicaton study with a sample of in-service teachers with on-the-job experience.

Method

Participants. We aimed to recruit N = 400 teachers through an online panel provider. This sample size represented the maximum feasible number in light of resource constraints. Again, we were interested to test for the effect of classroom composition on the suggested use of disciplinary measures. Given α = .05, 1 – β = .80 in an independent t-test (one-tailed), a sample size of N = 400 provides sufficient power to detect a small-to-medium effect of d_s = .25. Four hundred and three participants (200 female, 197 male, 2 unknown), who provided informed consent after having been fully debriefed at the end, comprised our final sample. Participants' mean age was 47.10 (SD = 13.62), and a majority of the sample (n = 326) self-categorized as White (8 Turkish or Turkish-German, 6 Arab or Arab-German, 5 Black or Afro-German, 46 other, 4 no response). A majority of participants reported to be in-service teachers (n = 244) or teachers in training (n = 30), with a mean work experience of 15.83 years (SD = 12.50).

Design and procedure. The study was a close replication of Study 1 and thus identical in terms of experimental design, conditions, materials, and measures. The experiment followed a between-subjects design with two conditions (i.e., classroom with 60% Arabic names vs. classroom with 0% Arabic names).

Measures

The two items to assess perceptions of the student and the misbehavior (rs ranged from .63 to .79), suggested severity of disciplinary measures (ω = .89), perceptions of the classroom context (ω = .89), and perceived classroom composition were identical to Study 1.

Internal motivation to control prejudice. For exploratory purposes, we included three items adapted from Plant and Devine (1998) to assess internal motivation to control prejudice (e.g., "I attempt to act in nonprejudiced ways toward other ethnic groups because it is personally important to me."). Participants responded using a scale from 1 (completely disagree) to 7 (completely agree). We calculated a mean score of participants' motivation to control prejudice across the three items (ω = .92).

Additional measures. Items used to assess participants' recall of names from the class registers, as well as whether participants had imagined a specific school type while responding to the survey were identical to those used in Study 1.

Results

See Table 2 for the correlation matrix of study measures.

Manipulation check. A manipulation confirmed that participants in the Arabic names condition estimated a higher percentage of students with Arabic names in the classroom (M = 46.72, SD = 21.23) than participants in the German names condition (M = 14.99, SD = 18.44), t(391.81) = 15.95, p < .001, d_s = 1.60, 95% CI [1.36; 1.81].

Perceptions of student and misbehavior. Next, we examined whether participants' perception of the student and the misbehavior differed between experimental conditions. We only observed a numerical difference between the Arabic names condition (M = 4.65, SD = 1.18) and the German names condition (M = 4.44, SD = 1.10), which was not statistically significant, t(399.59) = 1.84, p = .067, d_s = 0.18, 95% CI [-0.01; 0.38].

Table 2. Zero-order correlations of study measures for the Arabic names condition (below the diagonal) and German names condition (above the diagonal) in Study 2.

Variable	1	2	3	4	5
1. Suggested Discipline	-	.86***	.67***	-.14	.09
2. Perceptions Student	.88***	-	.71***	-.19**	.02
3. Perceptions Class	.66***	.73***	-	-.42***	.06
4. Enthusiasm	-.05	-.12	-.27***	-	-.04
5. Perceived Composition	.12	.13	.11	-.12	-

Note: * $p < .05$; ** $p < .01$; *** $p < .001$; Suggested Discipline = Suggested disciplinary measures; Perceptions Student = perceptions of student and misbehavior; Perceptions Class = perceptions of the entire classroom context; Enthusiasm = felt enthusiasm about teaching the class; Perceived Composition = perceived classroom composition.

Suggested severity of disciplinary measures. Again, consistent with our pre-registered hypothesis, participants in the Arabic names condition suggested the use of more severe disciplinary measures (M = 4.33, SD = 1.26) than participants in the German names condition (M = 4.10, SD = 1.22), $t(400.80) = 1.92$, p = .028 (one-sided, as pre-registered), d_s = 0.19, 90% CI [-∞; 0.36].[7]

Perceptions of the classroom context. We observed only a numerical difference between the Arabic names condition (M = 4.21, SD = 1.56) and the German names condition (M = 4.05, SD = 1.37), which was very small and not statistically significant, $t(395.42) = 1.10$, p = .135, d_s = 0.11, 95% CI [-0.09; 0.31]. Participants' felt enthusiasm to teaching the class in the Arabic names condition (M = 3.85, SD = 1.53) and in the German names condition (M = 3.81, SD = 1.42) was not significantly different, $t(395.21) = 0.28$, p = .782, d_s = 0.03, 95% CI [-0.17; 0.22].

Mediation analysis: classroom composition → perceptions of student and misbehavior → severity of discipline. We again used the "lavaan" package in R to conduct a mediation analysis with 1,000 bootstrap samples with classroom composition as the independent variable, perceptions of the student and the misbehavior as the mediator and the suggested severity of disciplinary measures as the dependent variable. Again, we first tested the total effect of classroom composition on the suggested severity of disciplinary measures, b = 0.24, SE = 0.12, p = .025, 95% CI [0.00, 0.48]. Again, to maintain consistency with our one-sided test of the total effect, we report the adjusted (i.e., halved) p-value for the total effect. We then tested the direct effect of classroom composition on perceptions of the student and the misbehavior (the mediator), which was statistically non-significant, b = 0.21, SE = 0.11, p = .062, 95% CI [-0.02, 0.43]. Next, we regressed the suggested severity of disciplinary measures on perceptions of the student and the misbehavior and on the classroom composition. The perceptions of the student and the misbehavior significantly predicted the suggested severity of disciplinary measures, b = 0.95, SE = 0.03, p < .001, 95% CI [0.89, 0.99], whereas the effect of classroom composition on the suggested severity of discipline became statistically non-significant, b = 0.04, SE = 0.06, p = .518, 95% CI [-0.08, 0.16]. Importantly, the indirect effect of classroom composition on the suggested severity of disciplinary measures via perceptions of the student and the misbehavior was statistically non-significant, b = 0.20, SE = 0.11, p = .064, 95% CI [-0.02, 0.42]. Thus, among in-service teachers, perceptions of the student and the misbehavior did not mediate the effects of the classroom composition on the suggested severity of disciplinary measures.

General Discussion

The present research examined whether teachers' decision-making in response to minor infractions differed depending on the racialized classroom composition. Specifically, we investigated whether teachers would suggest harsher discipline in classrooms with a high proportion of students perceived as Arab compared to those in classrooms with only students perceived as German. In two experiments, both preservice and in-service

[7] As in the previous study, we also tested this hypothesis by fitting a linear mixed-effects model, which yielded similar results to our pre-registered analysis (see Supplement).

teachers who were presented with disruptive behaviors in a classroom with a high proportion of students perceived as Arab suggested more severe discipline than those in a classroom with students perceived as German. Preservice teachers in Study 1 also (a) reported feeling more irritated by the students' behaviors and (b) perceived the entire classroom context as more challenging and negatively when there was a high proportion of students perceived as Arab, but these effects were not replicated among in-service teachers in Study 2.

It is worth highlighting that participants who were confronted with disruptive behavior in racialized classrooms *generally* suggested the use of more severe discipline. The presented scenarios did not reveal whether the disrupting student was perceived as Arab or German. Thus, our findings indicate that the effects of the classroom composition on disciplinary decisions might affect students, potentially regardless of whether they belong to a racialized or non-racialized group. If the classroom composition affects educational outcomes for all students regardless of racialized group membership, then this has important implications for our understanding of disparities in educational outcomes. Previous research has conceptualized disparities primarily as group-based disparities, i.e., the extent to which students from racialized groups are treated in comparison to their non-racialized peers (e.g., Okonofua & Eberhardt, 2015; Skiba et al., 2002). In contrast, the present research points to *context-based disparities* with implications that might extend beyond group-based disparities. Simply being in a classroom with many racialized peers might shift the odds of facing more severe disciplinary responses.

Although the present research observed consistent context-based disparities in disciplinary decisions across studies, there were noteworthy differences between preservice and in-service teachers. First, the effects of classroom composition on perceptions of the student were more pronounced among preservice teachers. Second, preservice teachers' perceptions of the entire classroom context depended on the classroom composition, whereas in-service teachers' perceptions did not. Thus, preservice teachers confronted with disruptive behavior in racialized classrooms perceived the entire classroom as more challenging and negatively, an effect not observed among in-service teachers. These observed differences between preservice and in-service teachers pose questions for future research. For example, it is debated whether expertise and work experience increase or decrease stereotyping and discrimination (Kemme et al., 2020; e.g., Sim et al., 2013). While the present research suggests that context-based disparities might be smaller among experienced teachers, some limitations warrant caution in this interpretation. First, although we recruited a substantial sample of in-service teachers, Study 2 might still have been at the lower boundary in terms of sample size and power in a between-subjects design.[8] Second, while in-service teachers reported substantial work experience, we do not know whether or to what extent the presented scenarios in racialized classrooms corresponded with their lived experiences. These limitations might have created boundary conditions for the observed effects of classroom composition. Consequently, more experimental research is necessary that more thoroughly assesses teachers' lived experiences (e.g., teaching experiences; whether they live in rural or urban areas), ideally based on large sample sizes to achieve high statistical power.

The present research used a manipulation of the immediate context (class registers) and demonstrated the impact of racialized classroom composition on teachers' decision-making. The scenarios lacked information about the environments surrounding classrooms (e.g., neighborhoods). Because adding such missing context information could potentially amplify disparities (Jasperse et al., 2022), we believe that we conducted a relatively conservative test of context-based disparities. School environments are typically not isolated from their surrounding geographic contexts. These contexts might themselves be targets of stereotypes depending on their demographic composition (Essien & Rohmann, 2024), which, in turn, could reinforce stereotyping and harsh decision-making in the classroom. Consequently, future research should model structural features likely present in the real world that have been shown to be related to context-based disparities.

[8] For example, post-hoc power analyses for the effect of classroom composition on suggested discipline (d = 0.19) indicated achieved power of $1 - \beta = 0.60$.

An important avenue for future research is to explore other teacher outcomes beyond disciplinary decisions. School discipline is just one of many areas where context-based disparities might occur (see Osei-Twumasi & Pinetta, 2023). Given the infancy of research on context-based disparities, there are numerous other important questions to address, such as the cognitive or affective processes underlying context-based differences in teachers' decision-making. These questions highlight the need for further research.

Constraints on Generality

In the present research, we observed that racialized classroom composition influenced participants' suggested use of disciplinary measures. These effects were robust against alternative modeling decisions (see Supplement) and consistent across studies with effect sizes of 0.20 and 0.19, respectively. Importantly, the effects were observed in two different samples, including a sample of in-service teachers (Study 2). We thus expect our findings to replicate in well-powered studies with university students and/or teachers.

Our teacher sample—although a convenience sample—was more balanced in terms of gender composition and slightly older compared to the population of teachers in Germany (see Statistisches Bundesamt, 2024a, 2024b). The age distribution of the sample was roughly comparable to that of the target population (see Fig S1 in the Supplement). Over 80 percent of participants in the teacher sample self-categorized as White. While we would expect a similar racialized composition in the target population, we acknowledge that such data are not systematically collected at federal or state levels in Germany. Previous research suggests that the racialized (mis)match between teachers and the student body is related to context-based disparities (Osei-Twumasi & Pinetta, 2023). Osei-Twumasi and Pinetta (2023) observed that White (but not Black) teachers in classrooms with higher percentages of Black students provided less emotional support, lower-quality instructional support, and were less able to effectively manage their classrooms. This suggests that the racialized composition of teacher samples may be an important factor in the emergence of effects of classroom composition on the use of disciplinary measures.

Our research was conducted in a specific local context, which may raise questions about the generalizability of our findings to other contexts. However, observational studies in the United States have shown similar links between racialized school composition and disciplinary practices using diverse methodologies (e.g., Osei-Twumasi & Pinetta, 2023; Skiba et al., 2014; Welch & Payne, 2010), suggesting that our findings may extend to other contexts.

The stimuli used in both studies consisted of multiple scenarios describing a student misbehaving in school. These scenarios were adapted from research conducted in the US context (Jarvis & Okonofua, 2020; Okonofua & Eberhardt, 2015), suggesting that situations described in these scenarios were applicable to another societal and educational context (i.e., Germany). Consequently, we expect that our findings generalize to future studies using similar materials. Nonetheless, empirical validation is necessary, and we encourage future research to examine the effects of classroom composition on disciplinary measures in other societal contexts.

Lastly, we manipulated the perceived classroom composition by presenting class registers that varied in the percentage of student names that signaled stigmatized racialized group membership. The success of future replications will depend on tailoring this manipulation to the relevant intergroup relations in a different societal context (e.g., Black-White intergroup relations in the United States). Our investigation focused on specific classroom compositions (60% vs. 0% Arabic names), leaving open questions about the effects at different ratios. Future research should explore these variations to better understand the impact of racialized classroom composition. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context (see Simons, Shoda, & Lindsay, 2017).

Conclusion

This research provided, to our knowledge, first experimental evidence of context-based disparities in school. We hope to provide starting points for a deeper understanding of how contextual factors may shape teachers' decision-making. By demonstrating that classroom composition influences teachers' disciplinary decisions, we have identified another potential vicious cycle between teachers and students—one that may exacerbate educational inequalities by systematically shaping

the learning environments students experience.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by a German Research Foundation grant (466464566) awarded to the first author.

ORCID iD

Iniobong Essien https://orcid.org/0000-0003-0277-7908
Birte Siem https://orcid.org/0000-0003-4697-7856
Lysann Zander https://orcid.org/0000-0001-7075-3436
Anette Rohmann https://orcid.org/0000-0003-4757-1460

Supplemental Material

Supplemental material for this article is available online.

References

Anyon, Y., Jenson, J. M., Altschul, I., Farrar, J., McQueen, J., Greer, E., … Simmons, J. (2014). The persistent effect of race and the promise of alternatives to suspension in school discipline outcomes. *Children and Youth Services Review*, *44*, 379–386. https://doi.org/10.1016/j.childyouth.2014.06.025

Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown*. Retrieved from https://github.com/crsh/papaja

Bailey, D. H., Jung, A. J., Beltz, A. M., Eronen, M. I., Gische, C., Hamaker, E. L., … Murayama, K. (2024). Causal inference on human behaviour. *Nature Human Behaviour*, *8*(8), 1448–1459. https://doi.org/10.1038/s41562-024-01939-z

Childs, T. M., & Wooten, N. R. (2023). Teacher bias matters: An integrative review of correlates, mechanisms, and consequences. *Race Ethnicity and Education*, *26*(3), 368–397. https://doi.org/10.1080/13613324.2022.2122425

Conigrave, J. (2023). *Corx: Create and format correlation matrices*. Retrieved from https://CRAN.R-project.org/package=corx

Dowle, M., & Srinivasan, A. (2023). *Data.table: Extension of 'data.frame'*. Retrieved from https://CRAN.R-project.org/package=data.table

Edwards, L. (2016). Homogeneity and inequality: School discipline inequality and the role of racial composition. *Social Forces*, *95*(1), 55–76. https://doi.org/10.1093/sf/sow038

Essien, I., & Rohmann, A. (2024). Space-focused stereotypes of immigrant neighbourhoods. *British Journal of Social Psychology*, *63*(4). https://doi.org/10.1111/bjso.12756

Fincke, G., & Lange, S. (2012). *Segregation an grundschulen: Der einfluss der elterlichen schulwahl*. Sachverständigenrat deutscher Stiftungen für Integration und Migration. Retrieved from https://d-nb.info/1046700308/34

García, E. (2020). *Schools are still segregated, and black children are paying a price*. Retrieved from https://www.epi.org/publication/schools-are-still-segregated-and-black-children-are-paying-a-price/

Henschel, S., Heppt, B., Weirich, S., Edele, A., Schipolowski, S., & Stanat, P. (2019). Zuwanderungsbezogene disparitäten. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich, & S. Henschel (Eds.), *IQB-bildungstrend 2018* (pp. 295–336). Münster; New York: Waxmann.

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

Jasperse, L., Stillerman, B. S., & Amodio, D. M. (2022). Missing context from experimental studies amplifies, rather than negates, racial bias in the real world. *Behavioral and Brain Sciences*, *45*, e78. https://doi.org/10.1017/s0140525x21000868

Kelley, K. (2023). *MBESS: The MBESS r package*. Retrieved from https://CRAN.R-project.org/package=MBESS

Kemme, S., Essien, I., & Stelter, M. (2020). Antimuslimische einstellungen in der polizei? Der zusammenhang von kontakthäufigkeit und -qualität mit vorurteilen und stereotypen gegenüber muslimen. *Monatsschrift für Kriminologie Und Strafrechtsreform*, *103*(2), 129–149. https://doi.org/10.1515/mks-2020-2048

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Müller, K. (2020). *Here: A simpler way to find your files*. Retrieved from https://CRAN.R-project.org/package=here

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science*, *26*(5), 617–624. https://doi.org/10.1177/0956797615570365

Okonofua, J. A., Walton, G. M., & Eberhardt, J. L. (2016). A vicious cycle: A social-psychological account of extreme racial disparities in school discipline. *Perspectives on Psychological Science*, *11*(3), 381–398. https://doi.org/10.1177/1745691616635592

Osei-Twumasi, O., & Pinetta, B. J. (2023). Quality of classroom interactions and the demographic divide: Evidence from the measures of effective teaching study. *Urban Education*, *58*(5), 899–930. https://doi.org/10.1177/0042085919893744

Payne, A. A., & Welch, K. (2010). Modeling the effects of racial threat on punitive and restorative school discipline practices. *Criminology*, *48*(4), 1019–1062. https://doi.org/10.1111/j.1745-9125.2010.00211.x

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

R Core Team. (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Selker, R., Love, J., & Dropmann, D. (2023). *Jmv: The 'jamovi' analyses*. Retrieved from https://CRAN.R-project.org/package=jmv

Sim, J. J., Correll, J., & Sadler, M. S. (2013). Understanding police and expert performance: When training attenuates (vs. Exacerbates) stereotypic bias in the decision to shoot. *Personality and Social Psychology Bulletin*, *39*(3), 291–304. https://doi.org/10.1177/0146167212473157

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2160588

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

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. https://doi.org/10.3102/0002831214541670

Skiba, R. J., Michael, RobertS., Nardo, A., & Peterson, ReeceL. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review*, *34*(4), 317–342. https://doi.org/10.1023/A:1021320817372

Statistisches Bundesamt. (2024a). Anteil der weiblichen lehrkräfte an allgemeinbildenden schulen in deutschland im schuljahr 2023/2024 nach schulart [graph]. In *Statista*. Retrieved from https://de.statista.com/statistik/daten/studie/1129852/umfrage/frauenanteil-unter-den-lehrkraeften-in-deutschland-nach-schulart/

Statistisches Bundesamt. (2024b). Verteilung der lehrkräfte an allgemeinbildenden schulen in deutschland im schuljahr 2023/2024 nach altersgruppen [graph]. In *Statista*. Retrieved from https://de.statista.com/statistik/daten/studie/1129882/umfrage/verteilung-der-lehrerinnen-in-deutschland-nach-altersgruppen/

Stürmer, S., Rohmann, A., Froehlich, L., & van der Noll, J. (2019). Muslim immigration, critical events, and the seeds of majority members' support for radical responses: An interactionist perspective. *Personality and Social Psychology Bulletin*, *45*(1), 133–145. https://doi.org/10.1177/0146167218780989

Welch, K., & Payne, A. A. (2010). Racial threat and punitive school discipline. *Social Problems*, *57*(1), 25–48. https://doi.org/10.1525/sp.2010.57.1.25

Welsh, R. O., & Little, S. (2018). The school discipline dilemma: A comprehensive review of disparities and alternative approaches. *Review of Educational Research*, *88*(5), 752–794. https://doi.org/10.3102/0034654318791582

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wiernik, Brenton M. (2022). *Preprint templates.* https://doi.org/10.17605/OSF.IO/HSV6A

Wilke, C. O. (2024). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. Retrieved from https://CRAN.R-project.org/package=cowplot