Urban Education
2023, Vol. 58(3) 427–456

DOI: 10.1177/0042085919873691
journals.sagepub.com/home/uex
SAGE

Bias or Empathy in Universal Screening? The Effect of Teacher–Student Racial Matching on Teacher Perceptions of Student Behavior

Ericka S. Weathers[1]

Abstract

This study uses linear probability models with student and teacher fixed effects to assess whether the racial match between teachers and students affects "at-risk" ratings on a teacher-completed universal screener of student internalizing and externalizing behavior. The data are from a large, urban California school district. I find that Asian and Black teachers are more likely to rate their same-race students "at-risk" for internalizing behavior compared with how the same Asian and Black students would be rated by White teachers. These findings have implications for policy and practice aimed at enhancing universal screening for externalizing and internalizing behavior.

Keywords

race, identity, urban school district, teacher perceptions, student behavior

Introduction

Universal screening is a systematic process of identifying students who are at risk of academic, behavioral, and emotional difficulties. Screening allows for

[1]The Pennsylvania State University, University Park, USA

Corresponding Author:
Ericka S. Weathers, The Pennsylvania State University, 300 Rackley Building, University Park, PA 16802, USA.
Email: eweathers@psu.edu

the early identification of and intervention for students facing such challenges in educational settings (Cook, Volpe, & Livanis, 2010; Hughes & Dexter, 2011). Screening is typically among the first steps in determining whether a student needs additional support. Universal screening is common in a Multitiered Systems of Support (MTSS) framework that exists in many schools across the country. Although the conceptualization of MTSS can vary across states, in its most basic form, it is a schoolwide, evidence-based, integrated, and early intervention framework designed to improve academic, behavioral, and emotional outcomes for students (Harlacher, Sakelaris, & Kattelman, 2014).

The literature on universal screening largely focuses on screening for academics, with less attention being given to behavior (Cook et al., 2010). Academic screening involves administering performance assessments to students to determine accuracy and fluency in reading, writing, and mathematics (Hughes & Dexter, 2011; Jenkins, Hudson, & Johnson, 2007). Screening for behavioral risk typically requires the use of teacher assessments of student externalizing (e.g., aggressive, overactive, and impulsive) and/or internalizing (e.g., anxiousness and social withdrawal) behavior, although in some schools office discipline referrals are used as the screening mechanism (McIntosh, Horner, Chard, Boland, & Good, 2006). Because academic screening involves the direct administration of assessments to students, whereas behavioral screening typically requires teachers and other school staff to reflect on and evaluate students' behavior, behavioral screening is arguably more subjective than academic screening. Students with multiple teachers can have several different assessments of their behavior, and different teachers may rate the same student differently.

There is evidence that differences between two teachers' ratings of the same student are not entirely random. Teachers' assessments of students vary systematically by teacher and student race and other demographic characteristics. In a meta-analysis of teacher expectations by student race, teachers were found to have the highest expectations of Asian students; more positive expectations of White students than Black and Latinx students; make more negative (e.g., discipline) and fewer positive (e.g., gifted program) referrals for Black and Latinx students; and direct more positive speech (e.g., praise and affirmation) toward White students than Black and Latinx students (Tenenbaum & Ruck, 2007). Other studies found that Black students are more likely to be referred to the office for discipline by their teachers even after controlling for teacher expectations, student behavior, academic performance, gender, special education status, socioeconomic status, and age (Bryan, Day-Vines, Griffin, & Moore-Thomas, 2012; Rocque, 2010). Furthermore, variation in teachers' perceptions of students is related to the

racial match between teachers and students (i.e., students and teachers are of the same race). For example, non-Black teachers are less likely to have favorable expectations and perceptions of the behaviors and academic capabilities of Black students (Dee, 2005; Downey & Pribesh, 2004; Ehrenberg, Goldhaber, & Brewer, 1995; Fox, 2015; Gershenson, Holt, & Papageorge, 2016).

Current Study

The potential for subjectivity in the universal screening of behavior and the documented knowledge on the sources of variation in teachers' perceptions of students is a criticism of behavior screening. Racial matching helps to explain variation in teachers' perceptions of students, but is racial matching also predictive when teachers' assessments are potentially consequential for subsequent student support? The current study assesses whether the racial match between teachers and students affects ratings on a teacher-completed universal screener of student externalizing and internalizing behavior used in an urban school district. I use the term urban to refer to the size of the city in which the study district is located as well as the broad array of racial, ethnic, linguistic, and socioeconomic diversity of the students within the study district (Milner & Lomotey, 2014).

Approximately 50% of elementary and secondary school students in the United States are non-White, yet non-White teachers comprise less than 20% of the teaching workforce (National Center for Education Statistics, 2012, 2016). This racial mismatch between the student and teacher population is particularly prevalent in urban school districts (Milner & Lomotey, 2014) and can have implications for educational practice. For example, Easton-Brooks (2014) argues, "within a system heavily populated by ethnic minority students and nonethnic minority teachers, developing connections or building continuity between school and home cultures can be challenging. . . " (p. 97). Furthermore, students with externalizing and internalizing behavior challenges are at greater risk of adverse outcomes such as poor academic performance, office discipline referrals, physical health problems, alcohol and substance abuse, future unemployment, and difficult relationships with peers (Baker, Grant, & Morlock, 2008; Cook et al., 2011; Flook, Repetti, & Ullman, 2005; Menzies & Lane, 2012; Oakes et al., 2010), yet internalizing behaviors are argued to be substantially underreported and underserved (Cook et al., 2011). If the receipt of behavioral support services in school is partially determined by multiple and subjective ratings on a teacher-completed screener of student behavior; if there are challenges in bridging school and home cultures in urban schooling environments that are at least partially related to the racial

mismatch between a largely White teaching workforce and a more racially and ethnically diverse student population; if students with untreated externalizing and internalizing behavior symptoms are at greater risk of adverse outcomes; and if students' internalizing behaviors are less likely to be reported and served, this could be consequential for students. Differential ratings could lead to referrals for some students and no referrals for other students, regardless of the actual need. This occurrence would complicate service delivery, thereby presenting another challenge to student success in urban school districts.

Although the existing literature on the impact of the racial match between students and teachers is informative, it is incomplete. Studies in this area have largely focused on school suspension, teachers' perceptions of students' academic abilities, and externalizing behaviors. Few studies use teachers' perceptions of students' internalizing behavior as an outcome. This might be because externalizing behaviors are arguably more identifiable by teachers than internalizing behaviors (Hinshaw, Han, Erhardt, & Huber, 1992) or because students with internalizing behaviors have traditionally been overlooked and underserved in schools (Cook et al., 2011). Of the studies that do examine teachers' perceptions of students' internalizing behaviors, they do not specifically examine the role of racial matching between students and teachers (e.g., Sbarra & Pianta, 2001) nor use teachers' assessments of students that have the same direct consequences as a universal screener (e.g., Wright, Gottfried, & Le, 2017). In addition, these studies tend to only have two teacher ratings per student (e.g., Dee, 2005; Ehrenberg et al., 1995; Fox, 2015; Gershenson et al., 2016). This small sample of teachers means there is relatively little variation in teachers' ratings within students. Furthermore, many of the studies assessing the effects of racial matching on teachers' perceptions of students' behavior and academic capabilities have identification strategies (or methodological designs) that do not remove potential bias. For example, these studies tend to only use models with student fixed effects (Dee, 2005; Fox, 2015; Gershenson et al., 2016). Although student fixed effects allow for within-student comparisons to eliminate omitted variable bias due to unobserved time-invariant characteristics of students (e.g., academic engagement), such an approach does not eliminate important sources of teacher bias (e.g., rating patterns). For example, without controlling for unobserved time-invariant characteristics of teachers, there could be concern that White teachers who rate White students may not be the same White teachers who are rating Black students, resulting in selection bias.

This study adds to the to the growing body of literature on universal screening for behavior as well as the existing knowledge on the impact of the racial match between teachers and students by using behavior measures that

can have consequences for students, internalizing behaviors as one of the outcome measures, a larger sample of teachers assessing the behavior of the same student, and teacher fixed effects to control for unobserved time-invariant characteristics of teachers.

Literature Review and Theoretical Context

Studies Linking Racial Matching and Teachers' Perceptions of Students

Prior studies examining teachers' perceptions of student competency, future educational attainment, and behavior have found that racial matching between teachers and students is associated with favorable perceptions of students. For example, Ehrenberg and colleagues (1995) examined how the race, gender, and ethnicity of teachers and students influenced teachers' subjective evaluations of their students. The survey asked teachers whether they expected the student to attend college, would recommend the student for academic honors, believed the student related well to others, communicated with the student outside of class, and believed the student worked hard. They found that students were perceived more favorably when they were evaluated by teachers who shared their same race, gender, and ethnicity. Using student and teacher fixed effects to control for unobserved characteristics of teachers and students, Ouazad (2014) investigated whether teachers assess the academic skills of their same-race students more favorably. Overall, the author found that White teachers provided less favorable assessments of Latinx and Black students. An investigation of racial mismatch (i.e., teachers and students are not of the same race) on teacher perceptions of 10th-grade students' competency and future educational attainment found that non-Black teachers have lower expectations of Black students compared with Black teachers' expectations of the same Black students (Gershenson et al., 2016). Fox (2015) found similar results in that Black teachers were more likely to expect Black students to complete more than high school.

Rong (1996) studied the effects of race and gender on teachers' perceptions of students' social behavior. The results showed that White female teachers provided lower ratings for Black students than White students. Dee (2005) examined whether student assignment to a demographically similar teacher influenced teachers' assessments of students' behavior and academic performance. The author found that the odds of a teacher perceiving a student to be frequently disruptive, consistently inattentive, and rarely completing homework were larger when the teacher was of a different race than the student. McGrady and Reynolds (2013) explored the consequences of racial

mismatch on teachers' perceptions of the classroom behavior of Asian, Black, Latinx, and White students. They found that when rated by White teachers, Asian students were viewed more favorably than White students were, Black students were viewed more negatively than White students were, and on average Latinx students were not rated differently than White students. Okonofua and Eberhardt (2015) conducted an experiment to assess how student race may influence teachers' responses to classroom misbehavior of Black and White students. They found that Black students were more likely labeled troublemakers than White students were. Teachers believed Black students warranted harsher consequences than White students did after two infractions. Teachers were more likely to view the misbehavior of Black students as part of a pattern and were more likely to envision themselves suspending Black students in the future. Although their study did not explicitly examine the racial match between teachers and students, it is important to note that roughly 80% of the teachers in their sample were White. Lindsay and Hart (2017) assess whether exposure to same-race teachers affects exclusionary discipline rates for Black students. They found reductions in exclusionary discipline rates for Black students with exposure to larger shares of Black teachers.

In sum, prior literature provides evidence that the racial match between teachers and students influences teachers' perceptions of students. Students are perceived more favorably when assessed by teachers of their same race. This is true for teachers' perceptions of students' academic capabilities, behavior, and future educational attainment.

Theoretical Explanations for Racial Matching and Teacher Perceptions

The literature puts forth several mechanisms through which racial matching could influence teachers' perceptions and expectations of students. Racial matching may function through systematic or implicit bias (Downey & Pribesh, 2004; Gershenson et al., 2016). Teachers may have biased perceptions and expectations of students of a different race. Another mechanism relating racial matching and teachers' perceptions and expectations of students is racial stereotypes. White teachers may have negative stereotypes of minority students (McGrady & Reynolds, 2013; Okonofua & Eberhardt, 2015). It is also possible that racial matching influences teachers' perceptions and expectations of students through stereotype threat. Students who believe that their teachers hold negative stereotypes about their racial group may be at a greater risk of conforming to these stereotypes (Steele & Aronson, 1995).

Racial matching may also perform through role model effects (Lusher, Campbell, & Carrell, 2015; Villegas & Irvine, 2010). Teachers are role models for students. This is particularly true for teachers of color of students of color. Teachers of color have been found to increase the self-worth of students of color and decrease feelings of isolation experienced by many students of color in educational settings (Villegas & Irvine, 2010). The strength of the teacher–student relationship is another mechanism linking racial matching to teacher perceptions. For example, Saft and Pianta (2001) examined the extent to which teacher perceptions of their relationships with students varied as a function of the ethnic match between the teacher and student. They found that when a student's and teacher's ethnicity were the same, teachers reported higher levels of closeness with the student, lower levels of conflict, and lower levels of student dependency of teacher. Furthermore, teachers of color tend to have high expectations of students, use culturally relevant pedagogy, forge caring and trusting relationships with students, confront issues of racism through their teaching, and serve as advocates and cultural brokers. This advocacy takes the form of communicating with teachers and administrators on behalf of students, providing information about relevant opportunities and services, questioning rules that are not in the best interest of students, and relating to students of color in ways that White teachers may not be able to identify with (Villegas & Irvine, 2010). These characteristics are associated with better rapport between same-race teachers and students (Lindsay & Hart, 2017).

The existing literature provides evidence that racial matching is related to teachers' perceptions of students' academic capabilities, future educational attainment, and behavior and that this relationship likely exists because of phenomenon such as implicit bias, negative stereotypes, stereotype threat, role model effects, and the strength of the teacher–student relationship. However, exactly how these mechanisms play out when teacher perceptions of student behavior are used to help identify students in need of behavioral support and intervention is more nuanced and largely dependent on how teachers perceive the process of universal screening and the associated externalizing and internalizing behaviors that they are evaluating.

Teachers likely interpret the universal screening process and externalizing and internalizing behaviors in one of three ways. First, teachers make no distinction between externalizing and internalizing behaviors nor between general assessments of students and assessments of students that are directly linked to behavioral support. If teachers have this interpretation, the results would follow the same general pattern as much of the previously cited existing literature on racial matching and teacher perceptions of students. More specifically, students would be less likely to be identified as needing

additional externalizing and internalizing behavioral support when assessed by teachers of their same race. This could be because of implicit bias, negative stereotypes, and stereotype threat that manifest in the interactions between teachers and students of a different race or because of the strength of the relationship between same-race teachers and students, all of which are cited in the literature on racial matching and teachers' perceptions of students.

Second, teachers view the assessment of both externalizing and internalizing behaviors as a means of identifying students in need of behavioral support. For example, an administrator in the study district noted that some students may "externalize," whereas other students may "internalize" and both could be symptomatic of an underlying challenge in need of additional support. If teachers complete the universal screener under the premise that they are helping students get necessary services, the strength of the relationship between teachers and students may factor into this process. Because the teacher–student relationship is arguably stronger when teachers and students are of the same race, teachers may have more empathy for their same-race students who externalize and internalize. This empathy from teachers might present in the form of wanting to make sure their same-race students receive the necessary support services to excel in school. Thus, when teachers view the screening process as a form of help, externalizing and internalizing behavior screeners completed by teachers for students of the same race will be more likely to identify students as in need of additional support compared with universal screeners completed by teachers for students of a different race.

Third, teachers have differential views of externalizing and internalizing behaviors. Administrators and staff in the study district also commented that the internalizing behavior screener focused less on behaviors that typically lead to disciplinary action and more on behaviors that might elicit empathy from teachers. Externalizing behaviors have been found to be more disruptive to the classroom and challenge teachers' authority than internalizing behaviors, providing a potential explanation as to why teachers may have differential views/responses to externalizing and internalizing behaviors (Gresham & Kern, 2004). If internalizing behaviors elicit empathy from teachers, then internalizing behavior screeners completed by teachers for students of the same race will be more likely to identify a student as in need of further behavioral support. This could be because of the strength of the relationship between same-race teachers and students. Same-race teachers may have more empathy for their same-race students or students with strong relationships with their same-race teachers may feel more comfortable exhibiting internalizing behaviors, thus making the behaviors easier for their same-race

teachers to detect. If externalizing behaviors are instead more problematic for classroom management, then externalizing behavior screeners completed by teachers for students of the same race will be less likely to identify a student as in need of additional behavioral support. This could be because the effects of racial matching on teacher perceptions of student behavior are moderated by strong relationships between same-race teachers and students. Same-race teachers may be better able to manage the externalizing behaviors of their same-race students. Or students may have more self-regulation (thus exhibit fewer externalizing behaviors) around their same-race teachers whom they view as role models and have strong relationships with. This hypothesized relationship could also be because of implicit bias, negative stereotypes, or stereotype threat that manifests in the interactions between teachers and students of a different race.

These possible interpretations of the universal screening of externalizing and internalizing behaviors, as well as the mechanisms through which racial matching is theorized to be related to teachers' perceptions of students, suggest that bias and empathy are the most salient factors. Although the current study is unable to determine how teachers interpret the universal screening of externalizing and internalizing behaviors and therefore the specific mechanisms at play, it provides several different lenses through which the relationship between racial matching and teacher perceptions of students' behavior can be understood.

Method

Data

The data for this study come from administrative records from a large, urban California school district. Data are for the 2014-2015 school year for sixth-, seventh-, and eighth-grade students and their teachers. Data include student and teacher demographics as well as categorical ratings and total scores on a universal screener of externalizing and internalizing behavior completed in Fall 2014. Each screener contains seven items measured on a 4-point Likert-type scale ranging from 0 (*never exhibiting the behavior in question*) to 3 (*frequently exhibiting the behavior*). Externalizing behavior items ask whether a student is defiant or oppositional to adults, fights or argues with peers, bullies others, gets angry easily, lies to get out of trouble, disrupts class activities, and has difficulty sitting still. Internalizing behavior items evaluate whether a student is nervous or fearful, bullied by peers, spends time alone, clings to adults, is withdrawn, seems sad or unhappy, and complains about being sick or hurt. The measure requires summing individual item scores to

get total scores. Total scores range from 0 to 21. Each total score is used to determine whether a student is "not at-risk," "on the radar," or "at-risk" for externalizing or internalizing behavior challenges. Externalizing behavior scores between 0 and 4 are designated "not at-risk"; 5 through 7, "on the radar"; and 8 through 21, "at-risk" for externalizing behavior challenges. Internalizing behavior scores between 0 and 4 are categorized as "not at-risk"; 5 through 8, "on the radar"; and 9 through 21, "at-risk" for internalizing behavior challenges.

The externalizing behavior measure used in this study was modeled after the Student Risk Screening Scale (SRSS; Drummond, Eddy, Reid, & Bank, 1994), but is slightly different. The Student Internalizing Behavior Screener (SIBS) was the measure used to assess student risk for internalizing behavior (Cook et al., 2011). The SRSS (though slightly different from the externalizing measure used in the current study) and the SIBS have been shown to be valid and reliable tools for assessing student risk for externalizing and internalizing behavior challenges. Using a sample of 674 high school students, Lane, Robertson Kalberg, Parks, and Carter (2008) found that the Student Externalizing Behavior Screener (SEBS) has high internal consistency (i.e., each item measure produces a similar score), test–retest stability (i.e., consistently produces the same results over time), interrater reliability (i.e., the degree of agreement among different raters), and convergent validity (i.e., correlated with another well-established measure designed to assess the same constructs). Oakes and colleagues (2010) also found strong internal consistency and test–retest stability for the SEBS using a sample of 1,142 elementary school students. Cook and colleagues (2011) examined the psychometric properties of the SIBS using a sample of 1,357 elementary school students. They found high internal consistency and temporal stability (i.e., stability of results over time) and moderate convergent validity with the SEBS.

Teachers use an online tool to complete the behavior screener. Teachers assess their students' behavior twice during the academic year, once in the fall and then again in the spring. Teachers only complete screeners for students they have known for at least 30 days. The screening process identifies students potentially in need of behavioral support. Students receiving "at-risk" ratings on the externalizing or internalizing screener from at least one teacher may be eligible for participation in a behavior intervention, although the behavior risk ratings are not solely the deciding factor.

For the current study, the sample was limited to teachers and students who identified as Asian, Black, Latinx, or White. This resulted in the exclusion of 36 teachers (12% of the original sample) who identified as mixed race, Native Hawaiian or Pacific Islander, American Indian or Alaska Native, or declined to state their race. I also excluded 751 students (10% of the sample) who

Table 1. Descriptive Statistics for the Analytic Sample.

Students	
% Asian	51.78
% Black	9.44
% Latinx	28.57
% White	10.22
% Female	49.26
% Male	50.74
% in sixth grade	32.59
% in seventh grade	35.84
% in eighth grade	31.58
Average screeners per student	3.87
Number of students	6,549
Teachers	
% Asian	21.88
% Black	5.86
% Latinx	9.38
% White	62.89
% Female	56.64
% Male	43.36
Average age	42.07
Average number of years in district	10
Average number of screeners per teacher	99
Number of teachers	256

identified as mixed race, Native Hawaiian or Pacific Islander, and American Indian or Alaska Native. I made these restrictions because of the small number of students and teachers who identified with these racial categories and because of the inability to create racial matching variables for teachers and students who identified as mixed race. I also excluded observations for one middle school in the sample because there were only five universal screeners completed across the entire school. The final sample included 256 teachers who completed 25,369 screeners for 6,549 students across 12 schools in the study district. On average, there were roughly four screeners completed for each student and 99 screeners completed by each teacher. Most screeners were completed by White teachers for Asian and Latinx students and by Asian teachers for Asian students. Table 1 displays descriptive characteristics of students, teachers, and screeners. The study sample has a slightly greater share of White teachers, a substantially greater share of male teachers, and a slightly smaller share of Latinx teachers compared with the entire district.

Table 2. Screeners Completed by Teacher–Student Race Dyad.

	n
Same-race Asian	3,189
Teacher Asian student Black	465
Teacher Asian student Latinx	1,448
Teacher Asian student White	536
Same-race Black	163
Teacher Black student Asian	588
Teacher Black student Latinx	353
Teacher Black student White	107
Same-race Latinx	847
Teacher Latinx student Asian	997
Teacher Latinx student Black	264
Teacher Latinx student White	256
Same-race White	1,731
Teacher White student Asian	8,664
Teacher White student Black	1,451
Teacher White student Latinx	4,310
N	25,369

The study sample has slightly greater representation of Asian students and a substantially smaller share of White students as compared with the entire district. Table 2 describes the number of screeners completed by teacher and student race. Nine of the 12 schools in the sample completed screeners for at least 98% of the students in their respective schools.

Analytic Strategy

The study district uses "at-risk" ratings (i.e., universal screener scores above 7 for externalizing behaviors and above 8 for internalizing behaviors) on the universal screener to identify students who may need additional behavioral support. Because I am interested in variation in teachers' perceptions of students' behavior when there is an attached consequence (i.e., identification and possible referral for behavior services), I am primarily concerned with screener scores greater than or equal to the "at-risk" threshold of 8 or more for externalizing behavior and 9 or more for internalizing behavior. Therefore, I collapsed the three-category outcome variables into binary outcomes. The "not at-risk" and "on-the radar" categories were merged into a "not at-risk" assignment, and the "at-risk" category was kept intact. More specifically, screeners scoring between 0 and 7 for the externalizing behavior measure and

between 0 and 8 for the internalizing behavior measure are categorized as "not-at-risk." I used the race of teachers and students to create 16 interaction variables for the analysis. These interactions are termed teacher–student race dyads (e.g., same-race Asian).

Multiple screenings completed for each student and multiple ratings given by each teacher provide the ability to make within-student and within-teacher comparisons. The identifying variation comes from these within-student and within-teacher differences. This approach eliminates bias due to unobserved time-invariant characteristics of students and teachers. For example, if students and teachers are not randomly assigned to classrooms, I would expect unobservable characteristics of students such as behavior, to correlate with observable characteristics of teachers such as gender. This would bias my estimates. Including student and teacher fixed effects eliminates bias created by nonrandom sorting of students and teachers into schools and classrooms. Perhaps some teachers give lower ratings regardless of their students' behavior or race; if this were true, this could lead to biased estimates. This bias could be downward (i.e., muting the effect of racial matching) or upward (i.e., overstating the effect of racial matching) depending on whether these teacher rating patterns are correlated with teacher race. Including teacher fixed effects also eliminates this form of bias.

Linear probability models with student and teacher fixed effects were used to estimate the effect of racial matching between teachers and students on at-risk externalizing and internalizing behavior ratings. I fit the following model:

$$E_{st} = \beta_0 + \beta_1 Same\text{-}Race Teacher - Student Dyad_{st} + \beta_2 Different\text{-}Race Teacher - Student\ Dyad_{st} + \zeta_s + \zeta_t + \varepsilon_{st},$$

where E_{st} is the binary externalizing or internalizing behavior risk rating for student *s* rated by teacher *t*. $Same\text{-}Race Teacher - Student\ Dyad_{st}$ is a vector of same-race teacher–student dyad covariates (e.g., teacher Asian and student Asian). $Different\text{-}Race Teacher - Student\ Dyad$ is a vector of controls for different-race teacher–student dyads (e.g., teacher Asian and student Black). ζ_s is a student fixed effect that controls for unobserved characteristics of students such as academic engagement that may influence teachers' ratings. ζ_t is a teacher fixed effect that controls for unobserved teacher characteristics that may influence their ratings of students (e.g., some teachers may give higher ratings regardless of their student's behavior or race). ε_{st} is a student-by-teacher-specific error term representing unobserved characteristics of the teacher–student race dyad that influenced teacher ratings of students.

Table 3. Screeners Rated At-Risk by Student and Teacher Race.

	Externalizing	Internalizing
Students		
% Asian	16.77	27.74
% Black	36.85	26.29
% Latinx	39.79	35.79
% White	6.59	10.18
Teachers		
% Asian	24.54	13.87
% Black	4.36	6.26
% Latinx	11.12	11.19
% White	59.98	68.68

Results

Descriptive

Seventy-seven percent of the 25,369 completed screeners were for different-race teacher–student dyads. Of the 6,549 students in the sample, 14% of students received at least one at-risk rating for externalizing behavior, 10% of students received at least one at-risk rating for internalizing behavior, and 5% of students received at least one at-risk rating for both. Of the 25,369 completed screeners, only 7% rated students at-risk for externalizing behavior and 4% rated students at-risk for internalizing behavior. The discrepancy between the percentage of students receiving an at-risk rating from at least one of their teachers and the percentage of screeners assigned an at-risk rating could be evidence that teachers are rating the same student differently or that students with more screeners are being rated at-risk less.

Table 3 presents descriptive statistics of at-risk ratings for externalizing and internalizing behavior by teacher and student race. Of the externalizing and internalizing behavior screeners rated at-risk, the majority were for Black and Latinx students for externalizing behaviors and Latinx students for internalizing behaviors. Of the screeners rated at-risk for externalizing or internalizing behavior, White teachers completed many of these assessments. Table 4 presents descriptive statistics of at-risk ratings in each teacher–student race dyad. Three percent of all assessments completed by Asian teachers for Asian students were rated at-risk.

At-Risk Versus Not At-Risk

Results from models estimating the effects of teacher and student racial matching on at-risk ratings for externalizing and internalizing behaviors are

Table 4. Screeners Rated At-Risk by Teacher–Student Race Dyad.

	Externalizing	Internalizing
% Same-race Asian	3.14	1.6
% Teacher Asian student Black	24.95	4.52
% Teacher Asian student Latinx	12.29	2.97
% Teacher Asian student White	4.29	1.68
% Same-race Black	25.77	14.72
% Teacher Black student Asian	1.19	2.04
% Teacher Black student Latinx	5.95	5.67
% Teacher Black student White	3.74	0
% Same-race Latinx	8.97	4.72
% Teacher Latinx student Asian	2.11	2.11
% Teacher Latinx student Black	29.92	10.61
% Teacher Latinx student White	5.08	4.3
% Same-race White	4.16	4.1
% Teacher White student Asian	1.81	1.89
% Teacher White student Black	26.81	11.16
% Teacher White student Latinx	9.30	5.03

Note. Columns do not total 100%.

shown in Tables 5 and 6. All models hold students constant, and White teachers are the reference category. Model 1 includes controls for teacher race, student race, and different-race teacher–student dyads. Model 2 includes controls for teacher race, different-race teacher–student dyads, and student fixed effects. Model 3 includes controls for student race, different-race teacher–student dyads, and teacher fixed effects. Model 4 is the fully specified and preferred model, it includes controls for different-race teacher–student dyads and student and teacher fixed effects. Joint *F* tests assess the null hypothesis that the predictors in the models are zero.

In Model 2 with student fixed effects, Black students who were rated by Black teachers were less likely to be rated at-risk for externalizing behavior compared with how the same Black student would be rated by a White teacher. However, when accounting for unobserved characteristics of teachers (i.e., using a teacher fixed effect) in Model 4, the same-race Black coefficient is no longer statistically significant. This is an important callout because conventional models (student fixed effects, subject fixed effects, and/or controls for other teacher and student observables) used in similar studies assessing the effect of racial matching on teachers' perceptions of students find positive effects (i.e., more favorable ratings/perceptions) for Black students with same-race teachers.

Table 5. Effects of Racial Matching on At-Risk Ratings for Externalizing Behavior.

	(1)	(2)	(3)	(4)
Same-race Asian	0.012 (0.013)	−0.005 (0.012)	0.018 (0.013)	0.004 (0.012)
Same-race Black	−0.006 (0.031)	−0.064* (0.029)	−0.023 (0.031)	−0.031 (0.029)
Same-race Latinx	−0.012 (0.018)	0.000 (0.017)	0.003 (0.018)	0.005 (0.017)
N	25,369	25,369	25,369	25,369
R^2	.084	.006	.062	.109
Joint *F* test	154.11***	10.02***	137.69***	4.25***
Student race controls	x		x	
Teacher race controls	x	x		
Different-race teacher–student dyad controls	x	x	x	x
Teacher fixed effect			x	x
Student fixed effect		x		x

Note. All models hold students constant. White teachers are the reference category. Standard errors are in parentheses. Joint *F* test to test the null hypothesis that the predictors in the models are zero.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 6. Effects of Racial Matching on At-Risk Ratings for Internalizing Behavior.

	(1)	(2)	(3)	(4)
Same-race Asian	0.021* (0.010)	0.019 (0.010)	0.027** (0.010)	0.027** (0.010)
Same-race Black	0.077 (0.024)	0.073** (0.025)	0.050* (0.023)	0.065** (0.025)
Same-race Latinx	−0.005 (0.014)	−0.014 (0.014)	0.001 (0.014)	−0.011 (0.014)
N	25,369	25,369	25,369	25,369
R^2	.02	.006	.01	.124
Joint *F* test	35.85***	8.84***	21.78***	6.02***
Student race controls	x		x	
Teacher race controls	x	x		
Different-race teacher–student dyad controls	x	x	x	x
Teacher fixed effect			x	x
Student fixed effect		x		x

Note. All models hold students constant. White teachers are the reference category. Standard errors are in parentheses. Joint *F* test to test the null hypothesis that the predictors in the models are zero.
*$p < .05$. **$p < .01$. ***$p < .001$.

Same-race teacher–student dyads were stronger predictors of at-risk ratings for internalizing behavior. Asian students who were rated by Asian teachers were 2.7 percentage points more likely to be rated at-risk for internalizing behavior compared with how the same Asian student would be rated

by White teachers. Black students who were rated by Black teachers were 6.5 percentage points more likely to be rated at-risk for internalizing behavior compared with how the same Black student would be rated by White teachers. The effects for same-race Asian and same-race Black are robust and hold up across most of the four models. I did not find any significant effects for same-race Latinx.

Alternative Configurations of the Outcome Variables

To further test the strength of the effects of teacher and student racial matching on ratings of externalizing and internalizing behavior, I fit Models 1 through 4 using different configurations of the externalizing and internalizing behavior outcome variables. More specifically, I fit the models by dichotomizing the outcome variables between "not at-risk" and "any level of risk." In this configuration, not-at-risk encompasses scores between 0 and 4 for externalizing behavior and 0 and 4 for internalizing behavior. Any level of risk collapses at-risk and on-the-radar ratings (5 and above for externalizing behavior and 5 and above for internalizing behavior). I also fit the models using the continuous 0 to 21 rating scale.

Of the 6,549 students in the sample, 25% of students received at least one any level of risk rating (includes on-the-radar ratings and at-risk ratings) for externalizing behavior, 24% received at least one any level of risk rating for internalizing behavior, and 12% received at least one any level of risk rating for both externalizing and internalizing behaviors. Twelve percent of the 25,369 completed screeners had any level of risk for externalizing behavior and 9% had any level of risk for internalizing behavior.

Tables 7 and 8 display results from models estimating the effects of teacher and student racial matching on ratings for externalizing and internalizing behaviors with outcome variables dichotomized between not at-risk and any level of risk. There were no significant effects for externalizing behavior in the preferred model. There was a significant effect for internalizing behavior for same-race Asian in Model 4. Asian students who were rated by Asian teachers were 2.9 percentage points more likely to be rated as having any level of risk compared with how the same Asian student would be rated by White teachers.

To interpret the coefficients in models using the continuous 0 through 21 rating scale, it is necessary to understand the distribution of rating scores. The distribution of externalizing and internalizing behavior scores measured on scales ranging from 0 through 21 clustered at 0. The distribution of the continuous scale scores for externalizing and internalizing behavior is depicted in Figures 1 and 2. The average total score for externalizing behaviors was

Table 7. Effects of Racial Matching on Any Level of Risk Ratings for Externalizing Behavior.

	(1)	(2)	(3)	(4)
Same-race Asian	0.012 (0.017)	−0.009 (0.016)	0.007 (0.016)	−0.012 (0.015)
Same-race Black	0.053 (0.040)	−0.042 (0.038)	0.013 (0.039)	−0.028 (0.037)
Same-race Latinx	−0.047* (0.024)	−0.032 (0.022)	−0.027 (0.023)	−0.03 (0.021)
N	25,369	25,369	25,369	25,369
R^2	.1	.003	.075	.161
Joint *F* test	188.53***	5.13***	169.71***	2.24*
Student race controls	x		x	
Teacher race controls	x	x		
Different-race teacher–student dyad controls	x	x	x	x
Teacher fixed effect			x	x
Student fixed effect		x		x

Note. All models hold students constant. White teachers are the reference category. Standard errors are in parentheses. Joint *F* test to test the null hypothesis that the predictors in the models are zero.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 8. Effects of Racial Matching on Any Level of Risk Ratings for Internalizing Behavior.

	(1)	(2)	(3)	(4)
Same-race Asian	−0.02 (0.015)	−0.009 (0.016)	0.021 (0.013)	0.029* (0.013)
Same-race Black	0.032 (0.036)	0.02 (0.039)	−0.008 (0.031)	0.025 (0.032)
Same-race Latinx	−0.029 (0.021)	−0.01 (0.022)	−0.013 (0.018)	−0.016 (0.019)
N	25,369	25,369	25,369	25,369
R^2	.023	.004	.016	.371
Joint *F* test	39.09***	6.62***	33.60***	3.30**
Student race controls	x		x	
Teacher race controls	x	x		
Different-race teacher–student dyad controls	x	x	x	x
Teacher fixed effect			x	x
Student fixed effect		x		x

Note. All models hold students constant. White teachers are the reference category. Standard errors are in parentheses. Joint *F* test to test the null hypothesis that the predictors in the models are zero.
*$p < .05$. **$p < .01$. ***$p < .001$.

1.51 ($SD = 3.86$) and the average total score for internalizing behaviors was 0.952 ($SD = 2.69$).

Tables 9 and 10 display results from models estimating the effects of teacher and student racial matching on rating scores for externalizing and

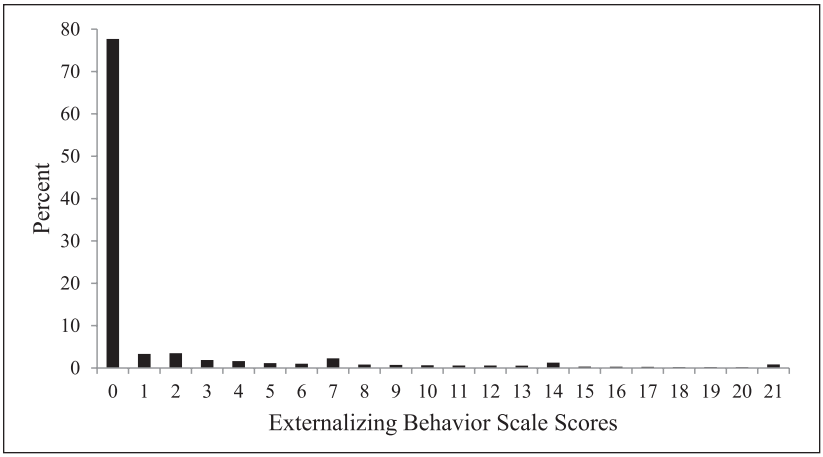

Figure 1. Distribution of externalizing behavior scale scores.

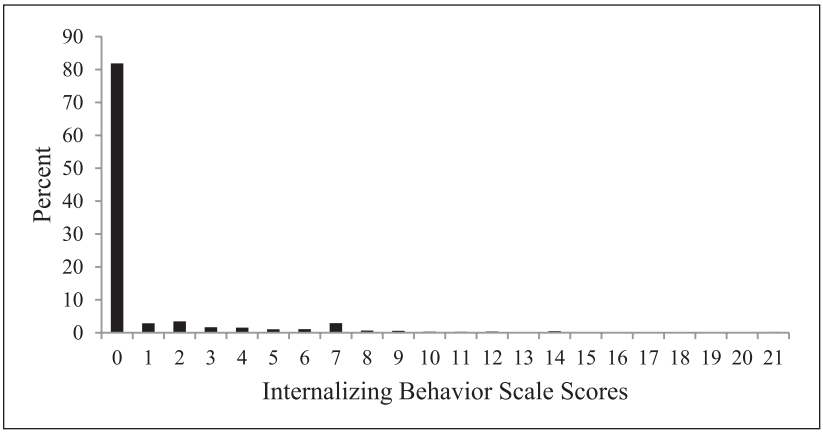

Figure 2. Distribution of internalizing behavior scale scores.

internalizing behaviors using the continuous outcome variables. I did not find significant effects for externalizing behavior in the preferred model. Again, there was a significant effect for internalizing behavior for same-race Asian in Model 4. The coefficients in the models were moderately large relative to the standard deviations of the externalizing and internalizing scale scores. Asian students rated by Asian teachers were more likely to have higher total scores than if these same Asian students were rated by White teachers.

Table 9. Effects of Racial Matching on Continuous Scale Scores for Externalizing Behavior.

	(1)	(2)	(3)	(4)
Same-race Asian	0.254 (0.194)	−0.01 (0.167)	0.226 (0.188)	−0.023 (0.157)
Same-race Black	0.389 (0.467)	−0.731 (0.405)	−0.098 (0.460)	−0.515 (0.388)
Same-race Latinx	−0.313 (0.277)	−0.122 (0.234)	−0.005 (0.271)	−0.036 (0.222)
N	25,369	25,369	25,369	25,369
R^2	.125	.007	.097	.188
Joint *F* test	241.56***	10.28***	225.22***	4.62***
Student race controls	x		x	
Teacher race controls	x	x		
Different-race teacher–student dyad controls	x	x	x	x
Teacher fixed effect			x	x
Student fixed effect		x		x

Note. All models hold students constant. White teachers are the reference category. Standard errors are in parentheses. Joint *F* Test to test the null hypothesis that the predictors in the models are zero.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 10. Effects of Racial Matching on Continuous Scale Scores for Internalizing Behavior.

	(1)	(2)	(3)	(4)
Same-race Asian	0.041 (0.142)	0.049 (0.143)	0.312* (0.127)	0.327** (0.125)
Same-race Black	0.437 (0.342)	0.441 (0.349)	−0.046 (0.312)	0.41 (0.308)
Same-race Latinx	−0.251 (0.203)	−0.216 (0.201)	−0.066 (0.183)	−0.177 (0.176)
N	25,369	25,369	25,369	25,369
R^2	.034	.006	.022	.307
Joint *F* test	59.02***	9.64***	48.10***	6.34***
Student race controls	x		x	
Teacher race controls	x	x		
Different-race teacher–student dyad controls	x	x	x	x
Teacher fixed effect			x	x
Student fixed effect		x		x

Note. All models hold students constant. White teachers are the reference category. Standard errors are in parentheses. Joint *F* test to test the null hypothesis that the predictors in the models are zero.
*$p < .05$. **$p < .01$. ***$p < .001$.

Overall, the results in Tables 7 to 10 show the robustness of the effect for same-race Asian. No matter how the outcome variable is operationalized, the results for same-race Asian hold. This suggests that the mechanisms for

assigning ratings for internalizing behavior are the same for the continuous rating scale, any level of risk ratings, and at-risk ratings. In other words, even if total rating scores are not high enough to warrant further support, Asian teachers are recognizing internalizing behaviors in Asian students. Contrarily, there were no significant effects for same-race Black on these different configurations of the internalizing behavior outcome variable. This could mean that the results are sensitive to different configurations of the outcome variable. It is also possible that the mechanisms that result in different ratings are specific to certain levels of risk and are more pronounced when assigning at-risk ratings than they are for the assignment of any level of risk ratings. This suggests that Black teachers recognize internalizing behaviors in Black students when they are severe enough to warrant further support (i.e., possible referral for behavioral intervention and services).

Discussion

Summary

This study sought to understand one aspect of the subjectivity in the universal screening of students' internalizing and externalizing behaviors. I used linear probability models with student and teacher fixed effects to assess the causal relationship between teacher and student racial matching and at-risk ratings on a universal screener of externalizing and internalizing behaviors. The findings in the current study suggest that there is subjectivity along racial lines in the universal screening of students' behavior. More specifically, Asian students were more likely to be rated at-risk for internalizing behavior when rated by Asian teachers compared with how the same Asian students would be rated by White teachers. The same was true for Black students rated by Black teachers. These findings were robust and held up across several different models. Furthermore, there was still a significant effect for same-race Asian even when using two alternative configurations of the internalizing behavior outcome variable.

Making Sense of the Results

Although these estimates provide a strong causal warrant that racial matching predicts teachers' perceptions of students' internalizing behavior, the interpretation that teachers have of the universal screening process and the mechanisms explaining these results are not completely clear. The fact that I found significant effects for internalizing behavior, but not for externalizing behavior, does suggest that teachers may in fact be differentiating

between internalizing and externalizing behaviors. If internalizing behaviors evoke more empathy than externalizing behaviors and if teachers believe the universal screening process is a means of garnering additional support for students in need, the results of the current study would suggest that students fare better when rated by teachers of their same race. However, if teachers make no distinction between externalizing and internalizing behaviors nor between general assessments of students and assessments of students that are directly linked to behavioral support, these findings would suggest that racial matching does not buffer against potential bias in teachers' perceptions of students.

The null findings for externalizing behaviors may be because there is less subjectivity (and thus less variation) along racial lines in the screening of externalizing behaviors. This might be because externalizing behaviors are arguably more identifiable and familiar to teachers than are internalizing behaviors (Hinshaw et al., 1992). For example, in the study district, some teachers and administrators reported feeling confident in addressing externalizing behavior, but were less equipped to deal with internalizing behavior challenges. Alternatively, even if teachers were aware that ratings on the externalizing and internalizing behavior screener would help to identify students in need of additional support, if teachers felt that the behaviors assessed on the externalizing behavior measure more closely resembled behaviors "that typically lead to disciplinary action," they may be concerned about assigning an at-risk label. This could be especially true if considering the educational climate in the United States over the last several years: particularly as it relates to racial disproportionality in school discipline and special education. The U.S. Department of Education Office for Civil Rights (OCR) released a brief documenting disproportionate trends in school discipline. The overall findings of the brief indicated that Black students experienced out-of-school suspensions and expulsions at substantially higher rates than White students did; students with disabilities were twice as likely to receive an out-of-school suspension compared with students without disabilities; and roughly 25% of Black boys and 20% of Black girls with disabilities received an out-of-school suspension (OCR, 2014). Although these disproportionate trends existed long before the release of this report, the 2014 findings placed discipline and special education at the center of policy conversations, program implementation, and academic research at the national, state, and school district level.

Most relevant to the current study is action that took place in California. For example, in September 2014, the Governor of California signed legislation banning expulsions for willful defiance. Willful defiance was a catchall category of discipline infractions that could include behavior such as talking

back to school personnel, violating the school dress code, or refusing to complete assignments. This ban was due in large part to the category's disproportionate impact on Black and Latinx students (California Assembly Bill 420, 2014). Another example is the California Department of Education's callout of school districts with disproportionate suspension, expulsion, and special education rates and the requirement that these districts have their policies and procedures reviewed and/or implement a plan to reduce disparities. This callout was in accordance with amendments made to the Individuals With Disabilities Education Act (IDEA). The completion of the behavioral screener in the study district comes on the heels of the OCR brief, during a period of increased focus on disparities in school discipline and special education, and substantial change in educational policy and program implementation regarding school discipline and special education at the Federal level and in the state of California.

In the context of this information, teachers may have been considerably thoughtful yet cautious in assessing each student out of concern that even though the purpose of the universal screener was to identify students in need of additional support, they would contribute to the disproportionate trends in student discipline and special education. This might be especially true for externalizing behaviors because these behaviors are characteristic of the types of infractions that generally warrant office discipline referrals and disciplinary consequences in the state of California (California Assembly Bill 1729, 2012). Ultimately, there could have been a tension between not wanting to perpetuate systemic bias and identifying students in need of additional behavioral support. This could also potentially explain the lack of considerable variation in scores on the externalizing and internalizing behavior outcome variables.

Comparing Results to the Extant Knowledge on Racial Matching

It is important to note that the results of the current study differ from Wright and colleagues' (2017) examination of teacher perceptions of students' externalizing and internalizing behavior. They found that having a same-race teacher was unrelated to teachers' ratings of students' internalizing behaviors, but related to teachers' ratings of students' externalizing behaviors. These differences in results could be due to bias from unobserved time-invariant student characteristics in their models or could be because of the differences between kindergarten and middle school. Arguably, middle school is a period of significant biological, psychological, and social change. Preadolescents can be sensitive about how they appear to others, need supportive relationships with adults, and are concerned about establishing peer networks. In

middle school, there is exposure of students to different teachers and peers as students change classrooms throughout the day. This mobility limits the amount of time students spend with a single teacher and a set group of peers, thus complicating some of the aforementioned needs and concerns. It is also argued that teachers spend more time monitoring behavior in middle school than teachers in elementary school (Fabes, Carlo, Kupanoff, & Laible, 1999; Roeser, Midgley, & Urdan, 1996). Perhaps these unique features of middle school and early adolescence alter the dynamics of the teacher–student relationship.

It is also important to note that similar studies assessing the effect of racial matching on teachers' perceptions of students' behavior and academic capabilities find significant effects for same-race Black–Black teachers rate Black students more favorably than White teachers do. The current study also finds a significant effect for externalizing behavior for Black students rated by Black teachers when using the conventional student fixed effects model. However, adding teacher fixed effects to the model nullifies this effect. This finding suggests that controlling for unobserved teacher characteristics might be an important source of eliminating bias.

Contrary to the findings for internalizing behavior for same-race Asian and same-race Black, I did not find significant effects for same-race Latinx. This finding is counter to what I expected. Although, it should be pointed out that in the extant literature examining the effects of racial matching on teachers' perceptions of students, there is not a consistent effect across studies (sometimes even within studies looking at multiple outcomes) for same-race Latinx. These inconsistent findings across and within studies suggest that racial matching operates differently for Black and Latinx groups.

Limitations

The small proportion of screeners rated at-risk for externalizing and internalizing behavior and the small sample of Black teachers rating Black students limit the external validity of the current study. Although the generalizability of these findings may be restricted, the takeaways from these findings are still relevant. The lack of considerable variation in the outcome variables, particularly the internalizing behavior outcome, may be an artifact of a larger challenge highlighted in the universal screening literature which says that internalizing behaviors are traditionally overlooked and underserved in schools (Cook et al., 2011). This is because internalizing behaviors are challenging to observe, are not generally disruptive to the classroom, do not challenge teachers' authority, and tend to meet teachers' behavioral expectations (Gresham & Kern, 2004). Yet and still, students with internalizing behavior

problems are at greater risk of lower academic achievement, physical health challenges, limited opportunities for employment, and difficult relationships with peers (Baker et al., 2008; Flook et al., 2005). If the goal of universal screening for externalizing and internalizing behavior is to identify students in need of additional support and teachers are the authority tasked with identifying these students, then understanding the mechanisms for the subjective identification of students is important for both research and practice.

Implications and Next Steps

This study found that Black and Asian students were more likely to be rated at-risk for internalizing behavior when assessed by their same-race Black and Asian teachers. This finding has implications for educational policy and practice. If teachers view the assessment of both externalizing and internalizing behaviors as a means of identifying students in need of behavioral support or if teachers have differential views of externalizing and internalizing behaviors, the study findings could speak to the strength of the relationship between same-race teachers and students that may be patterned by student comfort and teacher empathy. Such a finding and explanation would further support the calls for recruiting, training, and sustaining racial and ethnic minority teachers in urban schools as well as the rationale for such recruitment efforts (e.g., Villegas & Irvine, 2010) and suggests that the racial match between teachers and students may help to "bridge home and school cultures."

Regardless of teachers' perceptions of the universal screening process and externalizing and internalizing behaviors or whether the study findings are because of bias, stereotypes, stereotype threat, role model effects, or the strength of the relationship between teachers and students, it is important to ensure that all teachers are sufficiently prepared to teach in urban schooling environments, especially if student behavioral support is partially contingent upon teachers' identification of student risk or need. Urban schools are commonly characterized as lacking adequate teaching practices, sufficient resources, strong administrative decision making, advanced counseling and psychological services, and inclusive curriculum that meets the needs of all students (Milner & Lomotey, 2014). Furthermore, teachers and administrators sometimes adopt "context-neutral mindsets" that ignore how the political and social contexts of schools can shape opportunities and outcomes for students (Milner, 2012; Williamson, Apedoe, & Thomas, 2016). This contextual neutrality likely explains some of the aforementioned challenges in urban school settings. Sufficient teacher preparation requires an understanding of context. One such approach might be urban teacher residency programs. Urban teacher residencies are an innovative approach for recruitment,

preparation, and retention of teachers in urban schools. Context is at the core of urban teacher residency programs. Williamson and colleagues (2016) argue that

> the challenge for urban teacher residency programs is to provide teachers with the knowledge and skills to teach in a particular urban environment, while recognizing the similarities and differences that may exist across different urban contexts, particularly as it relates to their unique cultural landscapes, organization of the schools, conflicts over space, and the flow of capital. (p. 1173)

This style of teacher preparation might allow teachers in urban school environments to better understand their students and relevant social contexts and thus may make it easier for teachers to identify students in need of support.

Future research should unpack the mechanisms through which racial matching influences teachers' perceptions of students' externalizing and internalizing behavior, particularly in the context of universal screening. Classroom observations and surveys of teachers and students are some ways to explore potential mechanisms. Research exploring these mechanisms can guide policy and practice aimed at enhancing the universal screening process, increasing teachers' knowledge and capacity to identify student risk (with a significant emphasis on internalizing behavior), and ultimately improving service delivery for students with externalizing and internalizing behavior challenges.

Author's Note

Student-level district data were used for this study. The study district has requested to remain anonymous. This anonymity in combination with the use of restricted access student-level data prevents me from making these data publicly accessible.

Acknowledgments

The author thank Sean Reardon and Tom Dee for their valuable feedback on earlier versions of this article.

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: The research reported here was

supported in part by the Institute of Education Sciences (IES), U.S. Department of Education, through Grant R305B090016 to Stanford University. The opinions expressed are those of the author and do not necessarily represent views of the Institute or the U.S. Department of Education.

ORCID iD


Ericka S. Weathers https://orcid.org/0000-0002-9008-0838

References

Baker, J. A., Grant, S., & Morlock, L. (2008). The teacher-student relationship as a developmental context for children with internalizing or externalizing behavior problems. *School Psychology Quarterly*, *23*, 3-15. doi:10.1037/1045-3830.23.1.3

Bryan, J., Day-Vines, N. L., Griffin, D., & Moore-Thomas, C. (2012). The disproportionality dilemma: Patterns of teacher referrals to school counselors for disruptive behavior. *Journal of Counseling & Development*, *90*, 177-190. doi:10.1111/j.1556-6676.2012.00023.x

California Assembly Bill 1729. Pupil Rights: Suspension or Expulsion: Alternatives and Other Means of Correction (2012).

California Assembly Bill 420. Pupil Discipline: Suspensions Expulsions: Willful Defiance (2014).

Cook, C. R., Rasetshwane, K. B., Truelson, E., Grant, S., Dart, E. H., Collins, T. A., & Sprague, J. (2011). Development and validation of the student internalizing behavior screener: Examination of reliability, validity, and classification accuracy. *Assessment for Effective Intervention*, *36*, 71-79. doi:10.1177/1534508410390486

Cook, C. R., Volpe, R. J., & Livanis, A. (2010). Universal screening beyond academic performance: Introduction to the special series. *Assessment for Effective Intervention*, *35*, 195-196. doi:10.1177/1534508410379843

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *The American Economic Review*, *95*, 158-165.

Downey, D. B., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education*, *77*, 267-282. doi:10.1177/003804070407700401

Drummond, T., Eddy, J. M., Reid, J. B., & Bank, L. (1994, November). *The student risk screening scale: A brief teacher screening instrument for conduct disorder*. Paper presented at Prevention Conference, Washington, DC.

Easton-Brooks, D. (2014). Ethnic-matching in urban schools. In H. R. Milner & K. Lomotey (Eds.), *Handbook of urban education* (pp. 97-113). New York, NY: Routledge.

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review*, *48*, 547-561.

Fabes, R. A., Carlo, G., Kupanoff, K., & Laible, D. (1999). Early adolescence and prosocial/moral behavior I: The role of individual processes. *Journal of Early Adolescence*, *19*, 5-16.

Flook, L., Repetti, R. L., & Ullman, J. B. (2005). Classroom social experiences as predictors of academic performance. *Developmental Psychology*, *41*, 319-327. doi:10.1037/0012-1649.41.2.319

Fox, L. (2015). Seeing potential: The effects of student-teacher demographic congruence on teacher expectations and recommendations. *AERA Open*, *2*(1), 1-17. doi:10.1177/2332858415623758

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student-teacher demographic match on teacher expectations. *Economics of Education Review*, *52*, 209-224. doi:10.1016/j.econedurev.2016.03.002

Gresham, F. M., & Kern, L. (2004). Internalizing behavior problems in children and adolescents. In R. B. Rutherford, M. M. Quinn, & S. R. Mathur (Eds.), *Handbook of research in emotional and behavioral disorders* (pp. 262-281). New York, NY: The Guilford Press.

Harlacher, J. E., Sakelaris, T. L., & Kattelman, N. M. (2014). *Practitioner's guide to curriculum-based evaluation in reading*. New York, NY: Springer.

Hinshaw, S. P., Han, S. S., Erhardt, D., & Huber, A. (1992). Internalizing and externalizing behavior problems in preschool children: Correspondence among parent and teacher ratings and behavior observations. *Journal of Clinical Child Psychology*, *21*, 143-150.

Hughes, C. A., & Dexter, D. D. (2011). Response to intervention: A research-based summary. *Theory into Practice*, *50*, 4-11. doi:10.1080/00405841.2011.534909

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review*, *36*, 582-600. doi:10.1111/1540-5826.00049

Lane, K. L., Robertson Kalberg, J., Parks, R. J., & Carter, E. W. (2008). Student Risk Screening Scale: Initial evidence for score reliability and validity at the high school level. *Journal of Emotional and Behavioral Disorders*, *16*, 178-190. doi:10.1177/1063426608314218

Lindsay, C. A., & Hart, C. M. D. (2017). Exposure to same-race teachers and student disciplinary outcomes for Black students in North Carolina. *Educational Evaluation and Policy Analysis*, *39*, 485-510.

Lusher, L., Campbell, D., & Carrell, S. (2015, September). *TAs like me: Racial interactions between graduate teaching assistants and undergraduates* (Working Paper No. 21568). Cambridge, MA: National Bureau of Economic Research. doi:10.3386/w21568

McGrady, P. B., & Reynolds, J. R. (2013). Racial mismatch in the classroom: Beyond Black-White differences. *Sociology of Education*, *86*, 3-17. doi:10.1177/0038040712444857

McIntosh, K., Horner, R. H., Chard, D. J., Boland, J. B., & Good, R. H. (2006). The use of reading and behavior screening measures to predict nonresponse to school-wide positive behavior support: A longitudinal analysis. *School Psychology Review*, *35*, 275-291.

Menzies, H. M., & Lane, K. L. (2012). Validity of the student risk screening scale: Evidence of predictive validity in a diverse, suburban elementary setting. *Journal of Emotional & Behavioral Disorders*, *20*, 82-91.

Milner, H. R. (2012). Beyond a test score: Explaining opportunity gaps in educational practice. *Journal of Black Studies*, *43*, 693-718. doi:10.1177/0021934712442539

Milner, H. R., & Lomotey, K. (Eds.). (2014). *Handbook of urban education*. New York, NY: Routledge.

National Center for Education Statistics. (2012). *Digest of education statistics*. Retrieved from https://nces.ed.gov/programs/digest/d13/tables/dt13_209.10.asp

National Center for Education Statistics. (2016). *Back to school statistics*. Retrieved from https://nces.ed.gov/fastfacts/display.asp?id=372

Oakes, W. P., Wilder, K. S., Lane, K. L., Powers, L., Yokoyama, L. T. K., O'Hare, M. E., & Jenkins, A. B. (2010). Psychometric properties of the student risk screening scale: An effective tool for use in diverse urban elementary schools. *Assessment for Effective Intervention*, *35*, 231-239. doi:10.1177/1534508410379796

Office for Civil Rights. (2014). *Data snapshot: School discipline* (Vol. 1). Retrieved from https://www2.ed.gov/about/offices/list/ocr/docs/crdc-discipline-snapshot.pdf

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

Ouazad, A. (2014). Assessed by a teacher like me: Race and teacher assessments. *Education Finance and Policy*, *9*, 334-372.

Rocque, M. (2010). Office discipline and student behavior: Does race matter? *American Journal of Education*, *116*, 557-581. doi:10.1086/653629

Roeser, R. W., Midgley, C., & Urdan, T. C. (1996). Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning: The mediating role of goals and belonging. *Journal of Educational Psychology*, *88*, 408-422.

Rong, X. L. (1996). Effects of race and gender on teachers' perception of the social behavior of elementary students. *Urban Education*, *31*, 261-290.

Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly*, *16*, 125-141. doi:10.1521/scpq.16.2.125.18698

Sbarra, D. A., & Pianta, R. C. (2001). Teacher ratings of behavior among African American and Caucasian children during the first two years of school. *Psychology in the Schools*, *38*, 229-238.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797-811. doi:10.1037/0022-3514.69.5.797

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology*, *99*, 253-273. doi:10.1037/0022-0663.99.2.253

Villegas, A. M., & Irvine, J. J. (2010). Diversifying the teaching force: An examination of major arguments. *Urban Review*, *42*, 175-192. doi:10.1007/s11256-010-0150-1

Williamson, P., Apedoe, X., & Thomas, C. (2016). Context as content in urban teacher education: Learning to teach in and for San Francisco. *Urban Education*, *51*, 1170-1197. doi:10.1177/0042085915623342

Wright, A., Gottfried, M. A., & Le, V.-N. (2017). A kindergarten teacher like me: The role of student-teacher race in social-emotional development. *American Educational Research Journal*, *54*, 78-101.

Author Biography

Ericka S. Weathers is an assistant professor of Education (Educational Theory and Policy) in the Department of Education Policy Studies at The Pennsylvania State University. Her research uses quasi-experimental methodologies to understand the causes and consequences of racial and socioeconomic inequality in education.