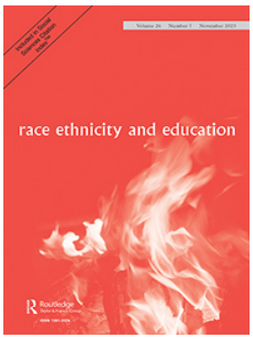

Routledge
Taylor & Francis Group

Race Ethnicity and Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/cree20

Ratings in black and white: a quantcrit examination of race and gender in teacher evaluation reform

Shanyce L. Campbell

To cite this article: Shanyce L. Campbell (2023) Ratings in black and white: a quantcrit examination of race and gender in teacher evaluation reform, Race Ethnicity and Education, 26:7, 815-833, DOI: 10.1080/13613324.2020.1842345

To link to this article: https://doi.org/10.1080/13613324.2020.1842345

RACE ETHNICITY AND EDUCATION
2023, VOL. 26, NO. 7, 815–833
https://doi.org/10.1080/13613324.2020.1842345

Routledge
Taylor & Francis Group

Ratings in black and white: a quantcrit examination of race and gender in teacher evaluation reform

Shanyce L. Campbell

School of Education, University of Pittsburgh, Pittsburgh, PA, USA

ABSTRACT

Education policy reform is often created and implemented in ways that advantage White individuals. This is even more salient when examining the intersection of race and gender identities. Using North Carolina's teacher evaluation system as a policy reform case, I examine whether there are differential observation ratings between Black and White women. The findings show that when Black and White women were similarly effective in the classroom, administrators rated Black women lower. However, there was no statistically significant influence on ratings by having a principal of the same race and gender as teachers. Black women were roughly two times more likely to be placed on a punitive professional development plan than White women. This study contributes to the narrow discourse on the effects of policies on Black women and provides policymakers with recommendations to ensure that teacher evaluation systems are justly implemented.

ARTICLE HISTORY
Received 22 October 2019
Accepted 14 October 2020

KEYWORDS
Teacher evaluation system; QuantCrit; intersectionality; Black women teachers; education reform

i am a black wave in a white sea. always seen and unseen.

- Nayyirah Waheed

Introduction

Education policy reform is often viewed as an objective, ahistorical process that holds a 'problem-solving' perspective on education (Thrupp and Willmott 2003; Gillborn 2005). This view undermines a critical examination of the role of systemic oppression (Thrupp and Willmott 2003). In particular, racism is largely muted and allows for the *tacit intentionality* of policies that advantage White individuals (Gillborn 2005). According to education policy scholar, Amy Stuart Wells, 'even when education policies are "color-blind" on the surface, they interact with school systems and residential patterns in which race is a central factor … (p 38).' Put simply, race matters, as does the intersecting identities of race and gender. This mattering is not only for students' schooling experiences, but also for teachers' professional experiences, a topic which often goes unnoticed and unexamined within the local and national policy landscape.

When examining education policies, David Gillborn (2005) suggests considering the priorities, beneficiaries, and outcomes of the policy to understand how racial inequities

CONTACT Shanyce L. Campbell shanycecampbell@pitt.edu

are maintained. In this study, I consider the latter two using North Carolina's comprehensive teacher evaluation systems (TES) as a case to examine the beneficiaries and outcomes of the policy reform among Black and White women. Driven by federal incentives such as Race to the Top, comprehensive teacher evaluation systems sought to develop new ways to assess teachers' performance through the use of multiple measures – classroom observation ratings, value-added scores, and other measures such as student surveys. Tied to financial support, states rapidly began developing and implementing TES. For instance, according to report by the National Council on Teacher Quality, only two states had not implemented a teacher evaluation system as of 2013 (Doherty and Jacobs 2013). Within the TES, classroom observations were a key aspect of the system such that 44 states and the District of Columbia required classroom observations, whereas, only 27 required the use of value-added scores in their evaluation systems (Doherty and Jacobs 2013). With a strong focus on classroom observations, I center my research on the following questions:

(1) Are there systematic differences in classroom observation ratings from the TES among Black and White women teachers?
(2) If so, does principal-teacher race and gender congruence account for possible differences in ratings?
(3) What are the practical ramifications for systematic differences in ratings?

Black women are centered for several reasons. First, research shows that Black women are exemplars of teacher quality as shown by their ability to improve student's learning, socio-emotional well-being, attendance, and suspension rates (Beauboeuf-Lafontant 2002; Lindsay and Cassandra 2017; Egalite and Kisida 2018; Cherng and Halpin 2016; Dunbar 2009; Dee 2004). In addition to being exemplars, Black women represent the second-largest racial-gendered group in the teacher workforce, yet their experiences often do not often shape educational reforms (Morris 2001). Lastly, studying Black women with an intersectional lens provides a more nuanced analysis of how education policies impacts individuals often overlooked in our current scholarship (Love 2019).

While the use of TES has significantly increased, there is a paucity of research on how this policy reform impacts teachers based on their race-gender[1] identities (Jacob and Walsh 2011; Campbell 2014). This study contributes to our growing knowledge that connects policy reform and teachers' racial-gendered identities. The study also contributes to the growing literature using critical race quantitative methods in the education policy space. I provide strong evidence that teachers Black women receive lower observation ratings than White women within and across schools, even after accounting for teacher effectiveness. I also find Black women were about two times more likely to be placed on a punitive professional development plan than White women. The results have important consequences for the recruitment and retention of Black women, a group of teachers who are necessary to the field yet continue to experience professional genocide. These results suggest that policymakers must intensify TES training, hold schools and districts accountable for racial inequities, and require that classroom observations be video-recorded.

Literature on evaluation systems and teacher's race and gender

Similar to many industries, teacher evaluations serve two primary functions: (1) improving teacher quality and (2) making personnel decisions by school leaders (Arvey and Murphy 1998; Castilla 2008; Scriven 1967). These functions are especially relevant for beginning teachers, who often require additional guidance on instructional practices, classroom management, and other aspects of teaching (Peterson 2004).

While it is important that evaluation systems aid teachers in improving their performance by providing instructionally useful feedback, evaluations must also reflect performance devoid of race-gender bias. While research across other industries has investigated issues of rater bias and fairness on performance ratings and have found racial (DuBois et al. 1993; Wilson and Jones 2008; Dewberry 2001) and gender differences (Oppler et al. 1992; Pulakos et al. 1989); within education, this topic has received little attention. Even less attention is placed on examining multiple oppressive identities of teachers. This section synthesizes the few studies that examine how teachers' race and gender are related to observational ratings. The studies are divided into two categories: 1) those that used main effects or additive approach; and 2) interactional approach to examine the simultaneous effect of race and gender on observation ratings.

Scholarship using race and gender as main effects

Consistent with fields like psychology, social identities are often understood to be unidimensional in education. For example, in most quantitative studies, race, and gender are separately examined in models and implications are made regarding the additive nature of race or gender. This method assumes that individuals' race and gender can somehow be separated and ranked.

In their investigation of Chicago Public Schools' teacher evaluation system, Jiang and Sporte (2016) disaggregated race and gender and found that men and teachers of color received statistically significantly lower observation ratings than women and White teachers, where negative effects were most pronounced among Black teachers. In separate models, the authors included school-fixed effects to test whether differences between school contexts accounted for these relationships. However, effects of teacher gender and race persisted, though the magnitude of the point estimate for Black teachers decreased by about two-thirds. The latter finding suggests that differences between schools accounted for most of the effects for Black teachers. The authors also found that the differential ratings are associated with where Black teachers are employed. Specifically, Black teachers worked in schools that enroll a high proportion of students classified as economically disadvantaged and teachers in these schools tend to receive, on average, low observation rating. Even so, within the same school, Black teachers were rated lower than White teachers.

In a more recent study, Campbell and Ronfeldt (2018) used various methodological approaches and a subsample of teachers randomly assigned to classrooms to examine the relationship between observation ratings and characteristics of teachers and classroom composition. Similar to Jiang and Sporte (2016), they found that observational ratings vary by teachers' gender. Men received lower ratings than women, even after adjusting for the characteristics of the students they teach and differences in teachers' instructional

quality. The authors also found that Black teachers received lower ratings, on average, than White teachers by the equivalent of the average difference in observational ratings between a first- and second-year teacher. However, the differences in ratings were driven by differences in classroom composition. Although the authors did not examine which classroom composition characteristics drove this result, they indicated that Black teachers taught in classrooms that enrolled a significantly larger proportion of Black, boys, special education, and low-academically performing students and smaller proportions of Latinx and low-income students than White teachers.

In a more recent study of low teacher evaluations on employment outcomes, Drake, Auletto, and Cowen (2019) found that teachers with low observation ratings were 5 to 10 times more likely to immediately exit the workforce across public schools. Similar to prior studies, the authors also found that teachers of color and/or men teachers were more likely to receive lower ratings than their peers even after accounting for teacher quality. Black teachers staffed in schools with predominately White colleagues were more likely to receive lower ratings than Black teachers staffed in schools with predominately Black colleagues.

Scholarship using race and gender as interaction effects

The previous studies discussed examined the impact of racial and gendered identities on observation ratings from a unidimensional perspective. An underlying assumption in these studies is that, on average, the effect of observation ratings is similar for Black women and Black men.[2] While it is important to understand how the results of teacher evaluation systems vary based on a single sociodemographic identity, it is equally important to understand how teacher evaluation systems are related to teachers' multiple social identities. However, there are only a few studies that examine observation ratings as they simultaneously relate to specific racial and gender group identities.

Jacob and Walsh (2011) examined the relationship between observation ratings and productivity-related teacher characteristics associated with performance, including educational credentials, experience, and absences as well as race-gender characteristics. They found that teachers who graduated from selective colleges, majored in education, and had more years of teaching experience received higher observation ratings than their peers. They also found that principals rated White women teachers higher than all other racial/ethnic-gendered groups, even after conducting robustness checks for sample selection. The authors note that a major limitation of their analysis was the inability to directly measure teachers' instructional quality using such as students' test performance. Therefore, it is unclear whether their findings reflect principal race-gendered bias in evaluating teachers.

Working to disentangle instructional quality from race-gender bias in observation ratings, Campbell (2014) built upon Jacob and Walsh (2011) by controlling for teachers' effectiveness, as measured by value-added measures. Consistent with Jacob and Walsh (2011), Campbell found that Black women math teachers and White men English language arts received lower ratings than White women counterparts in their respective subjects after adjusting for teacher effectiveness. Moreover, the findings held when teachers were compared only to their colleagues in the same schools. While Campbell (2014) attempts to introduce intersectionality theory, the methodology was more

reminiscent of an interaction approach in that the author privileged all racial-gendered groups to understand systematic bias in the TES.

Taken together, recent literature provides growing evidence of race and/or gender bias associated with the current use of TES. By focusing on a single sociodemographic characteristic or the intersecting characteristics without a critical lens, researchers overlook the deeper issues of bias related to the TES. This study provides additional nuance to the conversation around equitable practices under the widespread teacher evaluation system by exploring how the reform could impact Black women teachers by systematically identifying them as poor-performing teachers.

Conceptual and theoretical frameworks

The study draws upon Quantitative Critical Race Theory and Intersectionality to guide the methods and interpret the data related to the interplay between teachers' race and gender and evaluation policy reforms. Guided by Critical Race Theory (CRT), Quantitative Critical Race Theory (QuantCrit) is a framework coined by Gillborn, Warmington, and Demack (2018) to provide a way to 'apply CRT understandings and insights wherever quantitative data are used in research and/or encountered in policy and practice' (169). CRT purports that racism is common place, built into the social, economic, and political fabric of the United States (Delgado and Stefancic (2017). As a critical race-conscious framework, QuantCrit disrupts the ways current research using quantitative methods reproduces knowledge in ways that reinforce White interests (Crawford 2019). Five principles guide QuantCrit: 1) the centrality of racism, 2) numbers are not neutral, 3) categories are neither 'natural' nor given: for 'race' read 'racism,' 4) voice and insight: data cannot 'speak for itself'; and 5) using numbers for social justice (Gillborn, Warmington, and Demack 2018, 169). It also serves as a framework that legitimizes race equity and justice scholarship.

I also draw upon intersectionality, which was coined by Kimberle´ Crenshaw, a critical race Black feminist (1989), but has roots in Black feminist studies (hooks 1981; Collins 1990). Intersectionality theory is an analytic tool that complicates our understanding of how power and oppression intersect with individuals' multiple social identities. Further theorizing Intersectionality, Dill and Zambrana (2009) offered four core tenets of intersectionality that connect with social identities of privilege and oppression. The four tenets include: 1) centering the lived experiences of people of color in theory and research; 2) emphasizing the complexity of how multiple identities interact in connection to power and oppression; 3) highlighting how structures of power are interconnected that influence our lives, and 4) linking research and practice in a way that promotes social justice and social change.

QuantCrit and intersectionality allow for a nuanced understanding of how policies impact marginalized women. The absence of scholarship using intersectional, quantitative methods to examine how policies impact Black women and marginalized groups more broadly, limits our knowledge claims to White women and men's experiences. From a policy perspective, this translates into policies that serve and protect White teachers both women and men, knowingly or unknowingly.

Methods

Data

The data for this study comes from the North Carolina Department of Public Instruction (NCDPI) administrative data on students, teachers, classrooms, school characteristics, and teacher observation rating data. Unique teacher identifiers are used to link administrative data with teachers' summative observation rating data. I restrict the sample to beginning Black and White women who teach in public middle schools during the 2010–11 and 2012–13 academic years. While the focus on Black women was discussed in the introduction, I also restrict the sample to include White women as a racial comparison. This comparison is made not to normalize Whiteness, but as a way to indirectly understand how power drives policy decisions. Specifically, teachers and school principals are predominately White women, 80% and 53%, respectively (Taie and Goldring 2017; US DOE 2015).

Beginning teachers are defined as those in their first 5 years of teaching. I focus on this population because there is consistent evidence that teacher effectiveness continues to improve during the first 5 years of teaching (Clotfelter, Ladd, and Vigdor 2010; Harris and Sass 2011; Henry, Kevin Fortner, and Bastian 2012). Therefore, more variation in observation ratings may exist among beginning teachers, which improves the precision for estimating differential effects of ratings based on teachers' race-gender. The intentional focus on middle school teachers is driven by the limited research at this grade level, despite the importance of middle grades on student success (Benner, Boyle, and Bakhtiari 2017). Examining middle schools also allows for a more nuanced examination of evaluation ratings across multiple subject areas that are not possible in self-contained elementary classrooms.[3]

Teachers in the study taught English language arts (ELA), mathematics, science, and multiple subjects. The sample was also restricted to teachers with observation ratings across all standards for a given year. Because more experienced teachers are only evaluated on standards 1 (Demonstrate Leadership) and 4 (Facilitate learning for their students), I also restricted the data to teachers with ratings on all standards. The final sample consists of 2,852 unique middle school teachers across the three-year study period.

Sample

Table 1 summarizes the characteristics of teachers in the sample. About 16% of teachers identified as Black women and 84% identified as White women. The average observation ratings across all teachers in the sample was 3.37. Black women had an average rating of 3.24, whereas, White women had an average rating of 3.40. The difference in observation ratings between Black and White women was statistically significant. Across all observation rubric standards, Black women received, on average, statistically significantly lower mean scores than White women. The distribution of years of teaching experience was spread out across the five experience variables with the lowest proportion of teachers being those with 4 years of prior teaching experience. On average, the majority of White women in the sample were prepared via out-of-state certification, master's certification, or a Visiting International Fellowship (43%); Black women were more likely to be prepared via an alternative entry program (49%).

The teachers in the sample were staffed in schools where the principal was White. White women principals made up a slightly larger percentage than men (36% and 33%, respectively). Black principals made up the second-largest racial group and Black women made up a larger percentage of Black principals than men (17% and 12%, respectively). Thirty-eight percent of teachers were staffed in schools with principals of the same race and gender, with White women having a higher proportion of race-gender congruence with their principals than Black women (39% and 31%, respectively). Twenty-one percent of Black women were staffed in schools with a principal of a different race and gender, whereas only 10% of White women were staffed in schools with similar race-gender incongruence. On average, the principals in the sample have a little over 5 prior years of experience as a principal in North Carolina public schools.

Measures

Observation ratings

The observation instrument used to evaluate teachers was developed based on the North Carolina Professional Teaching Standards and included five standards and twenty-five elements that describe the knowledge, skills, and dispositions of an effective teacher. The five standards are: Demonstrate leadership; Establish a respectful environment for a diverse population of students; Know the content they teach; Facilitate learning for their students; and Reflect on their practice (SBE, n.d.). All standards were on a 5-point performance rating scale, where a rating of 1 indicated 'Not Demonstrated' and a rating of 5 indicated 'Distinguished.' The ratings from each standard were averaged within teacher and year to create a teacher-by-year composite measure of teachers' observation rating.[4] In a separate analysis, I create a dichotomous outcome variable to indicate whether the teacher was placed on a Monitored Growth Plan (=1) or required to develop an Individual Growth Plan (=0) as their Professional Development Plan. Individual Growth Plans (IGP) are required for teachers who were rated at least *Proficient* (3 or higher) across all standards, whereas a Monitored Growth Plan (MGP) was required for those who were rated *Developing* (2) on one or more standards and are high-stakes plans requiring teachers 1 year to achieve a *Proficiency* rating on the standard(s). The language across the two plans also suggests the high-stakes nature of the MGP. For example, the North Carolina Teacher Evaluation Process handbook uses the language '*shall develop*' in describing the requirements for IGP, but uses '*shall be placed on*' when describing the requirements for the MGP.

Value-added measures

Teacher effectiveness is measured by aggregated teacher scores using value-added measures (VAM). Acknowledging the recent VAM debate regarding the methodological and practical concerns regarding this measure (AERA 2015; ASA 2014), I do not assume that it is a complete measure of a teacher' instructional quality. However, there is evidence that teachers do account for some of the growth in student learning (American Statistical Association 2014). VAM was constructed from a statistical technique that estimates the contribution of a teacher in a given subject, grade, and year by isolating the effects of the teacher from other factors that influence student performance, such as family, student, and school characteristics. Teacher value-added measures were estimated separately for ELA, math, and science using a three-level hierarchical linear model that accounts for the

Table 1. Teacher and principal characteristics.

	Full Sample	Black Teachers	White Teachers
Teacher Effectiveness			
Teacher Value-Added Score	0.001	0.013	−0.001
	(1.008)	(1.055)	(0.998)
Zero Prior Years of Teaching	24.1%	20.1%	24.8%
One Prior Year of Teaching	20.2%	18.5%	20.5%
Two Prior Years of Teaching	20.3%	19.2%	20.5%
Three Prior Years of Teaching	21.0%	22.8%	20.6%
Four Prior Years of Teaching	14.5%	19.4%	13.6%
In-state Public Undergraduate Prepared	37.3%	32.4%	38.2%
Other Teacher Preparation	38.9%	18.4%	42.8%
Alternative Entry Prepared	23.8%	49.1%	19.0%
Teacher Observation Ratings			
Composite Observation Rating	3.371	3.242	3.395
	(0.546)	(0.508)	(0.550)
Demonstrate Leadership Standard	3.376	3.237	3.402
	(0.639)	(0.572)	(0.648)
Respectful Classroom Standard	3.402	3.287	3.424
	(0.631)	(0.610)	(0.632)
Content Knowledge Standard	3.344	3.225	3.367
	(0.611)	(0.582)	(0.614)
Facilitate Instruction Standard	3.374	3.218	3.403
	(0.622)	(0.603)	(0.622)
Reflect on Practice Standard	3.357	3.241	3.380
	(0.640)	(0.582)	(0.649)
Teacher Content Area			
Mathematics	37.6%	41.0%	37.0%
English Language Arts	48.1%	44.5%	48.7%
Science	9.4%	11.0%	9.1%
Multiple Subjects	4.9%	3.4%	5.2%
Principal Characteristics			
Years of Experience	5.096	4.753	5.161
	(4.687)	(4.410)	(4.735)
Black Women	17.4%	30.5%	14.9%
Black Men	11.8%	23.5%	9.6%
Latinx Men	0.0%	0.1%	0.0%
Native Men	0.1%	0.3%	0.1%
White Women	36.2%	25.1%	38.3%
White Men	33.2%	19.6%	35.7%
Same Race, Same Gender	37.5%	30.8%	38.8%
Same Race, Different Gender	34.2%	23.7%	36.2%
Different Race, Same Gender	16.7%	25.3%	15.1%
Different Race, Different Gender	11.5%	20.2%	9.9%
Observations	4,361	697	3,664

There were no Asian women or men principals, Latinx women principals, or Native women principals in the sample. Other teacher preparation includes out-of-state certified, Master's certification, visiting international fellow, and unclassified preparation.

nesting of students within classrooms, which are within schools. The value-added estimates include a robust set of covariates such as students' prior test scores and other student, classroom, and school characteristics to adjust for factors, which are arguably outside of teachers' control, but affect student performance. Teacher characteristics are omitted from the value-added estimates because of the possible correlation with student performance and observation ratings. VAM estimates were then standardized and averaged across teacher within year.

Focal teacher race and gender characteristics

The intersecting identity of race and gender, i.e., Black women, is the key teacher characteristic of interest in this study. Given the role of Whiteness in policy reform (Gillborn 2005, 2014) and the fact that 80% of the teacher labor market identify as White women (Taie and Goldring 2017), I compare the TES results of this group with that of Black women. Traditional approaches in addressing racial and gender inequities use main effects of each social identity separately in a way that 'conceptualizes people's experiences as separate, independent, and summative' (Bowleg 2008, 314). These approaches often ignore how multiple social identities are intertwined in ways which create differential experiences, opportunities, and outcomes among teachers. Using administrative records of teachers' self-identified race and gender, I created a dichotomous[5] variable to represent the race-gendered identity of the teachers as either Black or White women.[6] Moreover, quantitative scholarship that use race as a statistical variable, often fails to explicitly discuss racism in the data collection, analysis, and interpretation. This striping of racism from race, perpetuates a color-blind scholarship, policies, and practices.

Analysis

To empirically examine differences in evaluation results based on race-gender, I employed a three-level hierarchical linear modeling (HLM) approach that accounts for the fact that observation ratings teachers receive are dependent on the contexts in which they teach. Specifically, teachers (level 1) are nested (or work) in schools (level 2), which are nested within districts (level 3). HLM is also the most appropriate multiple regression estimation strategy from an intersectionality perspective which includes context, heuristic orientation, and complexity (see Scott and Siltanen 2017 for a detailed analysis and discussion how intersectionality is conceptualized in multiple regressions). Unconditional models indicated that the proportion of the total variance of ORs accounted for at the school and district levels was 34% and 15%, respectively.

Several HLM models are estimated; however, the reduced form of the full model is as follows:

$$OR_{ijk} = \delta_{000} + \delta_{001} T_i + \gamma_{002} C_i + \gamma_{010} S_{jk} + \lambda_y + \rho_g + \sigma_z + \vartheta_{ijk} + \tau_{0jk} + \varphi_{00k} \qquad (1)$$

where OR_{ijk} is the observation rating of teacher i, in school j in district k and is a function of fixed and varying teacher characteristics (T), a vector of average classroom characteristics (C), a vector of average school characteristics (S), year, grade, and subject indicators, $(\lambda_y, \rho_g, \sigma_z)$, and random effects $(\vartheta_{ijk}, \tau_{0jk}, \varphi_{00k})$ for teachers, schools, and districts, respectively. Teacher characteristics include race-gender identity (White women as the omitted comparison group), teacher VAM, indicators for years of prior teaching experience, and type of preparation program completed. The classroom characteristics include course track level (i.e., remedial vs. honors), class size, and classroom-aggregated student characteristics including racial identity, family economic status, special education status, gifted status, prior test performance, mobility, and absenteeism. The school characteristics include aggregated racial and family economic status of the schools' student population, average daily student membership, total per pupil expenditures, and principal characteristics. The principal characteristics include years of administrator experience and race and gender congruence with each

teacher. Congruence indicators were included as a mechanism that may account for differences in observation ratings (Grissom and Keiser 2011; Grissom, Nicholson-Crotty, and Keiser 2012). Drawing on Gershenson, Holt, and Papageorge's (2016) study examining teacher-student mismatch, I include four congruence categories – same race and same gender (omitted group); same race and different gender; different race and same gender and different race and different gender to understand the impact of principal race-gender mismatch on teachers' observation ratings.

Because teachers are not randomly assigned to schools it is probable that unobserved school-related factors such as school climate may account for the relationship between observation ratings and Black and White women teachers. To address this potential sorting issue, I also use school-fixed effects to estimate whether Black and White women teachers within the same school receive differential ratings. This approach is only feasible in schools that employ both Black and White women teachers. The school-fixed effects model specification is as follows:

$$OR_{ijk} = \beta_0 + \beta_1 T_{ijk} + \beta_2 C_{ijk} + \beta_3 P_{jk} + \lambda_y + \rho_g + \sigma_z + \varepsilon_{ijk} \quad (2)$$

where OR_{ijk} is the observation rating of teacher i, in school j in district k are a function of fixed and varying teacher characteristics (T), a vector of average classroom characteristics (C), principal-teacher race-gender congruence (P), year, grade, and subject indicators, $(\lambda_y, \rho_g, \sigma_z)$, and a random error (ε_{ijk}) that varies by teacher, school, and district. Standard errors are clustered at the school level.

I replicated the two specifications using the dichotomous variable to predict the odds of being required to complete an MGP between Black and White women teachers. These models used multilevel logistic regression and logistic with school-fixed effects.

Results

Table 2 summarizes the results estimating teachers' observation ratings as a function of teachers' race-gender. Several mechanisms were examined in efforts to rule out potential reasons for differences in observation ratings. Column 1 estimates teachers' observations ratings as a function of race-gender identity only. Column 2 then adds measures of teacher quality – teacher-level VAM scores, years of experience, and education preparation. This specification shows whether variation in the observed effects of ratings between Black and White women teachers are statistically accounted for by differences across the measures of teacher quality. Column 3 includes classroom and school characteristics to understand whether the observed relationship between teachers' race-gender identities and observations ratings are accounted for by the characteristics of the classrooms and schools as found in prior literature. Column 4 adds principal characteristics including years of experience and race-gender congruence indicators, to examine whether differences are accounted for by principals' inexperience or race-gender preferences. Column 5 includes school-fixed effects.

Across all modeling specifications, I find that Black women teachers receive lower observation ratings compared to White women. On average, there is a statistically significant positive relationship between observation ratings and two measures of teacher quality – VAM and years of teaching experience. Moreover, the relationship between the

observation ratings and experience is monotonic – estimates of ratings are higher as years of teaching experience increase.

To this point, the analysis has attempted to isolate the effects of observation ratings based on teacher, classroom, and school characteristics that are observable. These covariate-adjusted models account for available observable factors influencing a teachers' OR; however, they do not account for the fact that teachers tend to teach in schools with particular conditions and characteristics. For example, White women teachers, on average, teach in schools that teach a large, if not exclusive, population of White students. Similarly, the evidence that shows that Black women teachers, on average, also teach in schools that largely enroll Black students (Boyd et al. 2011; Ronfeldt, Kwok, and Reininger 2016; D'amico 2017). However, Black women teachers are more likely to teach in racially diverse schools than White women. To address this concern, I include a school-fixed effects identification strategy that allows for the comparison of Black and White women teachers within the same school. Column 5 represents results for the school-fixed effects model specifications. Consistent with the HLM specifications, the results indicate that within the same school, principals tend to rate White women teachers more favorably than Black women teachers even after accounting for differences in instructional quality and classroom composition.

In terms of the relationship between principal-teacher race-gender congruency and observation ratings, I found that same race, men[7] principals gave teachers more favorable

Table 2. Differences in observation ratings between black and white women teachers.

	Model 1	Model 2	Model 3	Model 4	Model 5
Black Women	−0.061***	−0.103***	−0.093***	−0.097***	−0.098***
	(0.021)	(0.019)	(0.020)	(0.020)	(0.025)
Teacher Value-Added Score		0.091***	0.090***	0.090***	0.087***
		(0.006)	(0.006)	(0.006)	(0.008)
One Prior Year of Teaching		0.222***	0.218***	0.217***	0.214***
		(0.019)	(0.019)	(0.019)	(0.021)
Two Prior Years of Teaching		0.371***	0.369***	0.367***	0.363***
		(0.019)	(0.019)	(0.019)	(0.022)
Three Prior Years of Teaching		0.457***	0.452***	0.451***	0.448***
		(0.019)	(0.019)	(0.019)	(0.025)
Four Prior Years of Teaching		0.516***	0.507***	0.501***	0.495***
		(0.022)	(0.022)	(0.022)	(0.027)
Other Teacher Preparation		−0.008	−0.012	−0.010	−0.011
		(0.015)	(0.015)	(0.015)	(0.019)
Alternative Entry Prepared		0.004	0.010	0.011	0.016
		(0.019)	(0.019)	(0.019)	(0.023)
Principal's Years of Experience				0.004**	0.002
				(0.002)	(0.003)
Same Race, Different Gender				0.069***	0.071*
				(0.023)	(0.040)
Different Race, Different Gender				0.036	0.053
				(0.029)	(0.045)
Different Race, Same Gender				−0.039*	−0.050
				(0.023)	(0.031)
Constant	3.288***	3.030***	3.111***	3.109***	3.009***
	(0.047)	(0.045)	(0.220)	(0.222)	(0.138)
Observations	4,353	4,339	4,332	4,243	4,243
Classroom Characteristics	NO	NO	YES	YES	YES
School Characteristics	NO	NO	YES	YES	NO
School Fixed Effects	NO	NO	NO	NO	YES

Standard errors in parentheses; School level clustered standard errors in parentheses.
*** p < 0.01, ** p < 0.05, * p < 0.1. *Other teacher preparation includes out-of-state certified, Master's certification, visiting international fellow, and unclassified preparation.*

ratings than same race, women principals. For example, Black men principals gave higher ratings than Black women principals to Black women teachers. Women principals of a different racial identity, gave teachers less favorable ratings than women principals of the same race; however, this effect did not persist in the school-fixed effects specification. Moreover, the magnitude of the effect on Black women, was ultimately unchanged after including racial-gendered congruency indicators in the models. This suggests that principal-teacher racial-gendered congruency has little impact on the ratings of Black and White women. As a sensitivity analysis not shown here, I also examined whether there was a differential effect of principal race-gender congruence for Black women teachers and found no statistically significant relationship.[8]

Examining differences in ratings from a practical and policy-relevant perspective, the final analysis examines the type of professional development plan the state required based on the summative observation ratings. Table 3 summarizes the results of the multilevel and school-fixed effects logistic regression specifications (similar to Columns 4 and 5 of Table 2). The results are reported in odds ratios. Estimates greater than 1 suggest teachers are more likely to complete a Monitored Growth Plan; whereas odds ratios less than 1 suggest teachers are less likely to be required to complete a Monitored Growth Plan. Across the two specifications, Black women are about 1.6 – 1.7 times more likely to be required to develop an MGP. Teachers staffed in schools with principals of the same race but different gender or principals of a different race and different gender are less likely to be required develop an MGP.

In a sensitivity analysis not shown here,[9] I also use each standard as a separate outcome variable to examine if the negative ratings among Black women are driven by one or more standards. For instance, it is plausible that Black women may receive lower ratings on the *Demonstrating Leadership* standard, which includes the dimension *Teachers lead in the classroom*, given Black women's cultural norms of communalism, which differs from White women's culture of individualism and authority (Boykin 1997; Moemeka 1998). However, consistent with the previous result, principals tend to evaluate Black women harsher across all five standards.

Discussion

In this study, I extend current literature on teacher evaluation systems by examining whether there are raced-gendered inequities embedded the implementation of TES reform. The results from this study contribute to the growing body of evidence that demonstrates a relationship between a teachers' race-gender and the ratings they receive. Consistent with Jacob and Walsh (2011) and Campbell (2014), I find that even after accounting for differences in teacher quality, Black women are rated lower than White women. Put more succinctly, even when Black and White women are similarly effective in the classroom, after observing their classrooms, administrators rate Black women lower. This was the case both within and between schools. Including measures of teacher quality pushes against scholar's tendency to think that racial inequality is the cause of some inherent deficiency among teachers of color (Gillborn, Warmington, and Demack 2018). Disaggregating the evaluation rubric by each of the five teaching standards did not prove to offer a more nuanced understanding of why Black women's observation ratings were lower.

Table 3. Odds of being placed on a monitored growth plan.

	Model 1	Model 2
Black Women	1.697***	1.553**
	(0.335)	(0.292)
Teacher Value-Added Score	0.622***	0.596***
	(0.046)	(0.042)
One Prior Year of Teaching	0.334***	0.328***
	(0.058)	(0.056)
Two Prior Years of Teaching	0.120***	0.117***
	(0.028)	(0.027)
Three Prior Years of Teaching	0.126***	0.118***
	(0.030)	(0.027)
Four Prior Years of Teaching	0.108***	0.093***
	(0.030)	(0.025)
Other Teacher Preparation	1.259	1.310*
	(0.217)	(0.211)
Alternative Entry Prepared	1.031	1.064
	(0.230)	(0.221)
Principal's Years of Experience	0.992	0.985
	(0.025)	(0.019)
Same Race, Different Gender	0.491**	0.608**
	(0.152)	(0.130)
Different Race, Different Gender	0.375***	0.594*
	(0.131)	(0.161)
Different Race, Same Gender	1.104	1.168
	(0.259)	(0.240)
Observations	2,110	4,477
Classroom Characteristics	YES	YES
School Characteristics	NO	YES
School Fixed Effects	YES	NO

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Other teacher preparation includes out-of-state certified, Master's certification, visiting international fellow, and unclassified preparation*.

The findings also show that principals' racial-gender congruence with the teachers they evaluated did not account for differences in ratings. It is possible that principals provide intangible benefits, such as encouragement, advocacy, and autonomy to teachers who share the same race and gender as found by Grissom and Keiser (2011); however, these benefits do not appear to be transferrable as principals conduct classroom observations. Interestingly, women principals of the same race were harsher in their ratings than principals of the same race-men principals. Despite, same-race men principals, being more lenient in their ratings this effect still did not reduce the magnitude of the effect of being a Black woman on observation ratings.

The more policy-nuanced models examining the type of professional development plan teachers are required to develop told a similar story as the more general results presented above. Black women are more likely asked to develop and implement a Monitored Growth Plan than White women, despite similar teacher quality. The MGP requires more administrator-level oversight with the expectation that teachers will improve within a year. The consequences of misidentifying who is required to complete an MGP are several. First, Black women may be pushed out of the profession, similar to the historical pushout following *Brown v. Board of Education* (Tillman 2004). Black women leaving the profession would remove a significant and influential group of teachers from the labor market and further exacerbate our current teacher-student sociodemographic divide. As previously mentioned, Black women are influential to

student's learning, socioemotional well-being, attendance, and suspension rates. The potential pushout of Black women teachers has a deleterious impact on students' constitutional right to a 'sound, basic education[10].' Second, Black women may decide or be asked to relocate to another school, which increases turnover, an issue shown to negatively impact student learning (Ronfeldt, Loeb, and Wyckoff 2013). Finally, misidentified ratings may create oppressive working conditions at a school, which may result in teacher turnover. In a study of Black women teachers, Farinde, Allen, and Lewis (2016) found that unsupportive school administrators negatively influence Black women's intentions to remain the profession.

While this study adds important new insights to the research on TES and the policy reform, more broadly, there are important limitations to consider. Despite my attempt to understand how the TES impacts Black women, measuring race and gender through quantitative methods only provides a crude approximation to the actual ways in which Black women experience and are impacted by the observation ratings received. Future studies should provide a more in-depth understanding of Black women's experiences and impacts under TES policy reform, using a mixed-methods design which allows scholars to capture the complexities associated with the design, implementation, and impact of teacher evaluation systems for this group of teachers. For broader generalizability, the future studies should also expand this analysis to the elementary and secondary level. This analysis is also useful to understand whether the findings are consistent across grade levels and not unique to middle grades only. While this interrogating teacher effectiveness as a mechanism that could account for differential ratings, is important, one major drawback with the use of VAM is the exclusion of teachers who do not teach in tested subjects; therefore, the results are not generalizable to teachers across subject areas.

Policy recommendations

As principals of QuantCrit recommend, I provide several recommendations that policymakers and educational leaders can consider to establish the implementation of a more socially just teacher evaluation system.

1. *Provide intensive TES training to administrators*. Similar to other states, North Carolina currently provides TES training online, through the completion of several modules. However, we know from the literature on professional development that sustainability and intensity are more likely to produce lasting impacts than short, one-off professional development (Garet et al. 2001). TES training must not be thought of as a one-time, annual check-list, but an ongoing process to ensure administrators have a deep knowledge and understanding of the evaluation rubric and how to effectively use the tool. Effective use also includes ensuring that administrators have the content knowledge and understanding of the classroom environment to evaluate teachers' performance. For example, throughout the academic year, district or state-level trainers could hold professional development sessions that re-iterate state standards, discuss and assess intercultural competence, and discuss ways to incorporate anti-oppressive evaluation strategies. TES trainings should center equity and provide opportunities for administrators to interrogate their own biases and the ways in which these biases may affect how they evaluate teachers. In addition to the internal *spirit work* that must be

done, these trainings must also offer practical ways that administrators can interrupt themselves when biases arise (Dillard 2008).

As a measured form of accountability, the TES training should also include calibration of administrators to ensure they are capable of accurately and consistently evaluating teachers equitably. Rubric calibrating would provide administrators with a deepened understanding of the rubric; shift their attention back to the rubric if classroom observations do not occur throughout the year; and potentially reduce administrators' likelihood of evaluating teachers based on their own biases (Buie 2018). Moreover, the increase in the reliability of observations also allows teachers to receive relevant and useful feedback to support their instructional improvement. Along with rubric calibration, the TES training should also include a certification process, where administrators must pass an exam that certifies them as capable of conducting teacher evaluations. Administrators unable to pass the certification should not be allowed to evaluate teachers, in which case another certified school administrator or a certified district-level evaluator would conduct observation ratings for the school. While these recommendations attempt to ensure that administrators are trained and certified, there is still the likelihood that principal error will occur (White 2018).

2. *Hold districts and schools accountable for the lack of parity in observation ratings*. Our current understanding of school accountability is based primarily on student test score performance (Figlio and Loeb 2011). However, in considering a more nuanced understanding of school accountability, another recommendation is to require that schools report teachers' observation ratings in aggregate and by subgroups (including an intersecting approach[11]) to district and state officials. This will allow district officials to examine rating variation and stability across time to detect patterns of inequality in ratings across teachers and intervene with solutions that hold administrators accountable. The goal is that this reporting requirement may elicit increased transparency by schools and districts in ways that promote a just teacher evaluation system.

3. *Require video-recorded observations*. Currently, evaluators are required to observe and evaluate 25 different teaching practices during the classroom observation as well as collect evidence of additional performance responsibilities demonstrated by the teacher. While this may be time-consuming, a more critical point of consideration is the ability or inability of a principal to retain each of the standards in real-time, which creates opportunities for race-gender bias in the evaluation process. As an attempt to circumvent potential bias, requiring all principals to record classroom observations not only allows the principal the option to review the teaching practice before assigning the rating; it also provides an artifact that can be leveraged during the post-observation conference. During the post-observation conference, principals would be required to highlight and justify their ratings of the teachers' performance using the classroom observation video. Additionally, as a protection for teachers, I recommend that a district instructional coach be present during the post-observation conference to act as an expert advocate on behalf of the teacher. A least costly option would require a district-level instructional coach only for teachers required to complete a monitored or directed growth plan.

These policy recommendations represent possible steps that policymakers and educators could consider to ensure that TES are justly implemented. However, a larger conversation is needed that acknowledges the past and present influence of racism in the ways education policies are created, implemented, and maintained. This study also is

a call for education policy scholars to conduct more intersectional and critical scholarship that explicitly addresses racism in the data collection, analysis, and interpretation.

Drawing upon QuantCrit's principal that data cannot speak for itself, I would like to explicitly acknowledge how my social position as a Black woman that was partially raised by a Black woman grandmother who was a classroom teacher for over 25 years, influences the importance and urgency of this study. Additionally, my heavy reliance on Black women, other mothers throughout my academic career are in large part, why I was able to thrive in educational spaces that inflict harm. Therefore, the research questions, results, and implications of this study matter beyond knowledge creation, but deeply impacts educators that I share a kindred connection with.

Notes

1. The hyphen here indicates that I am thinking about race and gender as intersecting identities. 'Race and Gender' indicates that I am viewing these identities separately.
2. In these two studies presented, women and men were the only two gender identities discussed.
3. This assumes that ELA, math, Algebra I and science teachers are mutually exclusive, which is not the case in all schools across North Carolina. For example, if a teacher teaches math and science, of the three required observations, principal evaluate the teacher solely on science performance, vice versa or a combination of the two courses.
4. At the end of the school year, the principal conducts a summative evaluation to determine each teacher's formal summative rating on each standard. The final rating is not a simple average of the four observation ratings; principals use multiple pieces of evidence including the previous observations as well as other artifacts (i.e., lesson plans, student work, and service on committees) to determine teachers' final ratings on each standard.
5. Another strategy used by critical quantitative scholars is effect coding, which does not require the normalizing White women, but allows for each group to be compared to the overall average; however, for the purpose of this study, I decided to use an indicator/dummy coding strategy (Mayhew and Simonoff 2015).
6. Racial descriptors provided in the data include: Gender descriptors provided in the data include: woman and man. Because these were the only descriptors provided, this study cannot provide a more nuanced understanding of different racial groups such as multi-raced women or gendered groups such as transgendered women.
7. I am aware that this may not be traditional speak; however, I do not assume that the labels 'male' and 'men' identify the same person.
8. Regression tables are available upon request.
9. Regression tables are available upon request.
10. This is a constitutional right based on the 1997 North Carolina Supreme Court ruling in the *Leandro v. State* case that declared all children in North Carolina have a fundamental state constitutional right to receive a sound basic education. *Leandro v. State*, 346 N.C. 336, 347 (1997).
11. To be clear, this is not to be confused with intersectionality. While the underlying goal of the recommendation is to detect inequities experienced by women teachers of color, the report would include all race-gendered teachers.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

American Educational Research Association. 2015. "AERA Statement on Use of Value-added Models (VAM) for the Evaluation of Educators and Educator Preparation Programs." *Educational Researcher* 44 (8): 448–452. doi:10.3102/0013189X15618385.

American Statistical Association. 2014. "ASA Statement on Using Value-Added Models for Educational Assessment." http://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf

Arvey, R. D., and K. R. Murphy. 1998. "Performance Evaluation in Work Settings." *Annual Review of Psychology* 49 (1): 141–168. doi:10.1146/annurev.psych.49.1.141.

Beauboeuf-Lafontant, T. 2002. "A Womanist Experience of Caring: Understanding the Pedagogy of Exemplary Black Women Teachers." *The Urban Review* 34 (1): 71–86. doi:10.1023/A:1014497228517.

Benner, A. D., A. E. Boyle, and F. Bakhtiari. 2017. "Understanding Students' Transition to High School: Demographic Variation and the Role of Supportive Relationships." *Journal of Youth and Adolescence* 46 (10): 2129–2142. doi:10.1007/s10964-017-0716-2.

Bowleg, L. 2008. "When Black+ Lesbian+ Woman≠ Black Lesbian Woman: The Methodological Challenges of Qualitative and Quantitative Intersectionality Research." *Sex Roles* 59 (5–6): 312–325. doi:10.1007/s11199-008-9400-z.

Boyd, D., H. Lankford, S. Loeb, M. Ronfeldt, and J. Wyckoff. 2011. "The Role of Teacher Quality in Retention and Hiring: Using Applications to Transfer to Uncover Preferences of Teachers and Schools." *Journal of Policy Analysis and Management* 30 (1): 88–110. doi:10.1002/pam.20545.

Boykin, A. W., R. J. Jagers, C. M. Ellison, and A. Albury. 1997. "Communalism: Conceptualization and Measurement of an Afrocultural Social Orientation." *Journal of Black Studies* 27 (3): 409–418. doi:10.1177/002193479702700308.

Buie, C. 2018. "The 3 Benefits of Calibrating Regularly." Teachstone (blog). March 6. https://info.teachstone.com/blog/the-3-benefits-of-calibrating-reguarly

Campbell, S. L. 2014. "Quality Teachers Wanted: An Examination of Standards-based Evaluation Systems and School Staffing Practices in North Carolina Middle Schools (Order No. 3633946)." *ProQuest Dissertations & Theses A&I. (1612601875)*.

Campbell, S. L., and M. Ronfeldt. 2018. "Observational Evaluation of Teachers: Measuring More than We Bargained For?" *American Educational Research Journal* 55 (6): 1233–1267. doi:10.3102/0002831218776216.

Castilla, E. J. 2008. "Gender, Race, and Meritocracy in Organizational Careers." *American Journal of Sociology* 113 (6): 1479–1526. doi:10.1086/588738.

Cherng, H.-Y. S., and P. F. Halpin. 2016. "The Importance of Minority Teachers: Student Perceptions of Minority versus White Teachers." *Educational Researcher* 45 (7): 407–420. doi:10.3102/0013189X16671718.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor. 2010. "Teacher Credentials and Student Achievement in High School a Cross-subject Analysis with Student Fixed Effects." *Journal of Human Resources* 45 (3): 655–681. doi:10.1353/jhr.2010.0023.

Collins, P. H. 1990. "Black Feminist Thought in the Matrix of Domination." In *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*, 221–238.

Crawford, C. E. 2019. "The One-in-ten: Quantitative Critical Race Theory and the Education of the 'New (White) Oppressed'." *Journal of Education Policy* 34 (3): 423–444. doi:10.1080/02680939.2018.1531314.

Crenshaw, K. 2018. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory, and Antiracist Politics [1989]." In *Feminist Legal Theory*, 57–80. Routledge.

D'amico, D., R. J. Pawlewicz, E. Penelope M., and A. P. McGeehan. 2017. "Where are All the Black Teachers? Discrimination in the Teacher Labor Market." *Harvard Educational Review* 87 (1): 26–49. doi:10.17763/1943-5045-87.1.26.

Dee, T. S. 2004. "Teachers, Race, and Student Achievement in a Randomized Experiment." *Review of Economics and Statistics* 86 (1): 195–210. doi:10.1162/003465304323023750.

Delgado, R., and J. Stefancic. 2017. *Critical Race Theory: An Introduction*. NYU Press.

Dewberry, C. 2001. "Performance Disparities between Whites and Ethnic Minorities: Real Differences or Assessment Bias?" *Journal of Occupational and Organizational Psychology* 74 (5): 659–673. doi:10.1348/096317901167578.

Dill, B. T., and R. E. Zambrana. 2009. "Critical Thinking about Inequality: An Emerging Lens." In *Emerging Intersections: Race, Class, and Gender in Theory, Policy, and Practice*, 1–21.

Dillard, C. B. 2008. "Re-membering Culture: Bearing Witness to the Spirit of Identity in Research." *Race Ethnicity and Education* 11 (1): 87–93. doi:10.1080/13613320701845848.

Doherty, K. M., and S. Jacobs. 2013. "State of the States 2013: Connect the Dots-Using Evaluations of Teacher Effectiveness to Inform Policy and Practice." *National Council on Teacher Quality*.

Drake, S., A. Auletto, and J. M. Cowen. 2019. "Grading Teachers: Race and Gender Differences in Low Evaluation Ratings and Teacher Employment Outcomes." *American Educational Research Journal* 56 (5): 1800–1833. doi:10.3102/0002831219835776.

DuBois, C. L., P. R. Sackett, S. Zedeck, and L. Fogli. 1993. "Further Exploration of Typical and Maximum Performance Criteria: Definitional Issues, Prediction, and White-black Differences." *Journal of Applied Psychology* 78 (2): 205. doi:10.1037/0021-9010.78.2.205.

Dunbar, R. B. 2009. "Finding Their Way: A Critical Ethnography of Five African American Women Educators' Early Experiences to Develop into Culturally Relevant Pedagogues."

Egalite, A. J., and B. Kisida. 2018. "The Effects of Teacher Match on Students' Academic Perceptions and Attitudes." *Educational Evaluation and Policy Analysis* 40 (1, March): 59–81. doi:10.3102/0162373717714056.

Farinde, A. A., A. Allen, and C. W. Lewis. 2016. "Retaining Black Teachers: An Examination of Black Female Teachers' Intentions to Remain in K-12 Classrooms." *Equity & Excellence in Education* 49 (1): 115–127. doi:10.1080/10665684.2015.1120367.

Figlio, D., and S. Loeb. 2011. "School Accountability." In *Handbook of the Economics of Education*. Vol. 3, 383–421. Elsevier.

Garet, M. S., A. C. Porter, L. Desimone, B. F. Birman, and K. Suk Yoon. 2001. "What Makes Professional Development Effective? Results from a National Sample of Teachers." *American Educational Research Journal* 38 (4): 915–945. doi:10.3102/00028312038004915.

Gershenson, S., S. B. Holt, and N. W. Papageorge. 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.

Gillborn, D. 2005. "Education Policy as an Act of White Supremacy: Whiteness, Critical Race Theory and Education Reform." *Journal of Education Policy* 20 (4): 485–505. doi:10.1080/02680930500132346.

Gillborn, D. 2014. "Racism as Policy: A Critical Race Analysis of Education Reforms in the United States and England." *The Educational Forum* 78 (1): 26–41. doi:10.1080/00131725.2014.850982.

Gillborn, D., P. Warmington, and S. Demack. 2018. "QuantCrit: Education, Policy, 'Big Data' and Principles for a Critical Race Theory of Statistics." *Race Ethnicity and Education* 21 (2): 158–179. doi:10.1080/13613324.2017.1377417.

Grissom, J. A., J. Nicholson-Crotty, and L. Keiser. 2012. "Does My Boss's Gender Matter? Explaining Job Satisfaction and Employee Turnover in the Public Sector." *Journal of Public Administration Research and Theory* 22 (4): 649–673. doi:10.1093/jopart/mus004.

Grissom, J. A., and L. R. Keiser. 2011. "A Supervisor like Me: Race, Representation, and the Satisfaction and Turnover Decisions of Public Sector Employees." *Journal of Policy Analysis and Management* 30 (3): 557–580. doi:10.1002/pam.20579.

Harris, D. N., and T. R. Sass. 2011. "Teacher Training, Teacher Quality and Student Achievement." *Journal of Public Economics* 95 (7–8): 798–812. doi:10.1016/j.jpubeco.2010.11.009.

Henry, G. T., C. Kevin Fortner, and K. C. Bastian. 2012. "The Effects of Experience and Attrition for Novice High-school Science and Mathematics Teachers." *Science* 335 (6072): 1118–1121. doi:10.1126/science.1215343.

hooks, B. 1981. *Ain't I a Woman: Black Women and Feminism*. Boston, MA: South End Press.

Jacob, B. A., and E. Walsh. 2011. "What's in a Rating?" *Economics of Education Review* 30 (3): 434–448. doi:10.1016/j.econedurev.2010.12.009.

Jiang, J. Y., and S. E. Sporte. 2016. "Teacher Evaluation in Chicago: Differences in Observation and Value-Added Scores by Teacher, Student, and School Characteristics." *Research Report*. University of Chicago Consortium on School Research.

Lindsay, C. A., and M. D. H. Cassandra. 2017. "Exposure to Same-race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina." *Educational Evaluation and Policy Analysis* 39 (3): 485–510. doi:10.3102/0162373717693109.

Love, B. L. 2019. *We Want to Do More than Survive: Abolitionist Teaching and the Pursuit of Educational Freedom*. Boston, MA: Beacon Press.

Mayhew, M. J., and J. S. Simonoff. 2015. "Non-White, No More: Effect Coding as an Alternative to Dummy Coding with Implications for Higher Education Researchers." *Journal of College Student Development* 56 (2): 170–175. doi:10.1353/csd.2015.0019.

Moemeka, A. A. 1998. "Communalism as a Fundamental Dimension of Culture." *Journal of Communication* 48 (4): 118–141. doi:10.1111/j.1460-2466.1998.tb02773.x.

Morris, J. E. 2001. "Forgotten Voices of Black Educators: Critical Race Perspectives on the Implementation of a Desegregation Plan." *Educational Policy* 15 (4): 575–600. doi:10.1177/089590480101500400 4.

Oppler, S. H., J. P. Campbell, E. D. Pulakos, and W. C. Borman. 1992. "Three Approaches to the Investigation of Subgroup Bias in Performance Measurement: Review, Results, and Conclusions." *Journal of Applied Psychology* 77 (2): 201. doi:10.1037/0021-9010.77.2.201.

Peterson, K. 2004. "Research on School Teacher Evaluation." *NASSP Bulletin* 88 (639): 60–79. doi:10.1177/019263650408863906.

Pulakos, E. D., L. A. White, S. H. Oppler, and W. C. Borman. 1989. "Examination of Race and Sex Effects on Performance Ratings." *Journal of Applied Psychology* 74 (5): 770. doi:10.1037/0021-9010.74.5.770.

Ronfeldt, M., A. Kwok, and M. Reininger. 2016. "Teachers' Preferences to Teach Underserved Students." *Urban Education* 51 (9): 995–1030. doi:10.1177/0042085914553676.

Ronfeldt, M., S. Loeb, and J. Wyckoff. 2013. "How Teacher Turnover Harms Student Achievement." *American Educational Research Journal* 50 (1): 4–36. doi:10.3102/0002831212463813.

Scott, N., and J. Siltanen. 2017. "Intersectionality and Quantitative Methods: Assessing Regression from a Feminist Perspective." *International Journal of Social Research Methodology* 20 (4): 373–385. doi:10.1080/13645579.2016.1201328.

Scriven, M. 1967. "The Methodology of Evaluation." *Perspectives of curriculum evaluation and AERA monograph Series on Curriculum Evaluation, No. 1*. Chicago: Rand McNally.

Taie, S., and R. Goldring. 2017. "Characteristics of Public Elementary and Secondary School Teachers in the United States: Results from the 2015–16 National Teacher and Principal Survey." *First Look. NCES 2017–072*. National Center for Education Statistics.

Thrupp, M., and R. Willmott. 2003. *Educational Management in Managerialist Times*. McGraw-Hill Education (UK).

Tillman, L. C. 2004. "(Un)intended Consequences? The Impact of the Brown V. Board of Education Decision on the Employment Status of Black Educators." *Education and Urban Society* 36 (3): 280–303. doi:10.1177/0013124504264360.

U.S. Department of Education. 2015. "National Center for Education Statistics, National Teacher and Principal Survey (NTPS)." *Public School Principal Data File*. –16.

Waheed, N. 2013. *Salt*. CreateSpace Independent Publishing Platform.

Wells, A. S. 2014. "Seeing past the 'Color-blind' Myth of Education Policy." *The Education Digest* 80 (3): 38.

White, M. C. 2018. "Rater Performance Standards for Classroom Observation Instruments." *Educational Researcher* 47 (8): 492–501. doi:10.3102/0013189X18785623.

Wilson, K. Y., and R. G. Jones. 2008. "Reducing Job-irrelevant Bias in Performance Appraisals: Compliance and Beyond." *Journal of General Management* 34 (2): 57–70. doi:10.1177/030630700803400204.