Retaining Teachers of Color in an Era of High-Stakes Teacher Evaluation: Investigating Racial Differences in Teacher Evaluation Ratings and Teacher Turnover

Jason A. Grissom[1]
Brendan Bartanen
Ashley A. Jones
Vanderbilt University

***

Teacher diversity has been linked to improved student outcomes including increased academic achievement, higher rates of gifted assignment, and lower rates of exclusionary discipline. To recognize these benefits, there has been a growing research and policy focus on recruiting and hiring teachers of color. However, studies have shown that teachers of color have high turnover rates, so retention of teachers of color has also become an important part of the conversation. There are numerous school-related factors that may differentially impact teachers of color and result in higher rates of turnover. We look to update these findings on rates of turnover and factors related to turnover using data that was collected after the widespread implementation of rigorous evaluation systems. We find that on average Black teachers in Tennessee receive lower evaluation scores than their White counterparts and those differences are a significant predictor of turnover.

***

Numerous recent studies have demonstrated the benefits of a racially and ethnically diverse teacher workforce for students, particularly students of color. These benefits are evident in increased academic achievement (Clotfelter, Ladd, & Vigdor, 2010; Egalite, Kisida, & Winters, 2015), higher rates of gifted assignment (Grissom, Nicholson-Crotty, & Nicholson-Crotty, 2009; Grissom, Rodriguez, & Kern, 2017), lower rates of exclusionary discipline (Holt & Gershenson, 2015; Lindsay & Hart, 2017; Kinsler, 2011), and reduced absenteeism (Farkas et al., 1990; Holt & Gershenson, 2015), among other outcomes.

To increase presence of diverse teachers, there has been a growing policy focus on the need to recruit, hire and retain teachers of color. Most research finds that teachers of color are

[1] This research was facilitated by the Tennessee Education Research Alliance. We thank Nate Schwartz, Laura Booker, Erin O'Hara, and Sylvia Flowers for helpful comments.

more likely than White teachers to turnover (Ingersoll & May, 2011; Putnam, Hansen, Walsh, & Quintero, 2016). However, they are more likely than White teachers to persist at "difficult-to-staff" schools, characterized by large proportions of low-income and minority students (Villegas & Irvine, 2010; Scafidi, Sjoquist, & Stinebrickner, 2007). Increases in teacher retention may result in fewer understaffed minority and low-income schools. Identifying strategies for stemming turnover among teachers of color requires a good understanding of what factors lead them to turn over at higher rates.

Although several studies have examined the factors that drive turnover among teachers of color (see Achinstein et al., 2010, for a review), relatively few studies have made use of large-scale administrative databases to examine this phenomenon. Of the studies that do exist on this topic, most predate Race to The Top and the era of high-stakes evaluation systems. Revisiting this analysis in this new era is important if we suspect that differences in how White teachers and teachers of color are evaluated may contribute to racial gaps in teachers' turnover propensities.

This is particularly important for teachers of color, who we find tend to have lower observation scores as this leads to questions about potential rater biases. Studies have shown that Black ratees tend to be given lower ratings than their White peers even when performance measures are equal (Stauffer & Buckley, 2005; Kraiger & Ford, 1985). If these lower ratings lead to greater job dissatisfaction, as (Poon, 2004) finds possible, this could be a contributing factor to higher rates of teacher turnover.

We make use of data from Tennessee from 2011-12 to 2015-16, which correspond to the first five years of implementation of the state's multiple-measures teacher evaluation system, to examine teacher turnover among Black and White teachers. We address three research questions. First, what are patterns of turnover among Black and White teachers in Tennessee? Second, what

individual, classroom, and school factors predict teacher turnover and, more specifically, differential turnover among Black teachers? Third, to what extent do Black teachers receive different evaluation scores than their White colleagues, and do these differences explain differences in turnover among Black and White teachers?

Background

Greater teacher racial and ethnic diversity is associated with a variety of positive student outcomes for both students of color and their White peers. Exposure to teachers of color is associated with greater achievement gains among Black and Hispanic students (Dee, 2004; Egalite, Kisida, & Winters, 2015). Research has found similar effects for non-achievement outcomes, including lower absenteeism rates (Farkas et al., 1990; Holt & Gershenson, 2015), lower rates of exclusionary discipline (Lindsay & Hart, 2017), and higher likelihood of being identified as gifted (Grissom, Rodriguez, & Kern, 2017; Grissom & Redding, 2016). Moreover, the benefits of teacher diversity on students of color appear also to carry forward to future years, yielding, for example, increased likelihood of taking honors and Advanced Placement courses (Grissom, Kabourek, & Kramer, 2017) and lower probability of dropping out of high school (Gershenson, Hart, Lindsay, & Papageorge, 2017). White students appear to respond positively to the presence of teachers of color as well; Cherng and Halpin (2017) find that students, regardless of race, rate Black and Latino teachers higher on numerous measures, including how well the teacher motivates students to high academic standards and builds supportive relationships with students.

Research on the benefits of a racially and ethnically diverse teacher workforce has focused attention on factors that undermine that diversity. Much of the national conversation on increasing teacher racial/ethnic diversity focuses on stemming turnover among teachers of color.

Ingersoll and May's (2011) look at national data from the Schools and Staffing Survey shows that, in 2003-04, 47,600 teachers of color entered the teacher workforce, but by the next year, 56,000 had left teaching, meaning that public schools were losing substantially more teachers of color than they were adding in that timeframe. Evidence suggests that the rate of turnover among teachers of color continues to increase. A more recent report by Ingersoll and May (2016) shows that the overall turnover rate (both moves to other schools and exits) of Black teachers grew from 13% in 1988-89 to 19% in 2011-12. This 19% rate stands in contrast to the rate of turnover among White teachers, which stood at only 15% in 2011-12.

One important set of factors that explain higher rates of turnover among teachers of color relate to the characteristics of the schools in which they work. Teachers of color are disproportionately found in urban schools with high proportions of low-income and low-achieving students and students of color, and numerous studies have shown that teachers are more likely to leave such schools (Clotfelter, Ladd, & Vigdor, 2007; Ingersoll, 2001). A likely reason is poorer working conditions (Grissom, Viano, & Selin, 2016; Simon & Johnson, 2015). Teachers in schools with large numbers of low-income students are more likely to report inadequate instructional materials, low-quality facilities, ineffective leadership, less teacher input into pedagogical decisions, higher rates of student discipline, and less collegial relationships among staff (e.g., Grissom, 2011; Ladd, 2011; Loeb, Darling-Hammond, & Luczak, 2005; Simon & Johnson, 2015; Torres, Santos, Peck, & Cortes, 2004), which push teachers to seek employment elsewhere. Although all teachers are more likely to leave such schools, research also shows that teachers of color are more likely than White teachers to persist in such "hard-to-staff" schools, perhaps due to attachment or commitment to serving students of color (Clotfelter, Ladd, & Vigdor, 2007; Achinstein et al., 2010; Villegas & Irvine, 2010; Scafidi et al., 2007).

Among the factors contributing to higher turnover rates in schools with large proportions of traditionally disadvantaged students, school leadership may be particularly important (Boyd et al., 2011; Grissom, 2011; Grissom & Bartanen, 2017; Ladd, 2011). Principals influence many aspects of teacher work, including the overall school climate, how school resources are allocated, how well teachers are supported, the quality of feedback and coaching teachers receive, and other aspects of how the school as an organization is managed (e.g., Grissom & Loeb, 2011; Grissom, Loeb, & Master, 2013). In more recent years, many states have expanded another key way that principals impact teachers by increasing the rigor of teacher evaluation systems. These systems emphasize increased frequency of teacher observation using standardized rubrics and formalized feedback conversations based on observations, typically in combination with achievement-based measures, including teacher value-added, though classroom observations carry more weight in evaluations than these other measures (Whitehurst, Chingos, Lindquist, 2014). With perceived rigor has come higher stakes attached to teacher evaluation, with districts and states creating expectations for principals to place low-scoring teachers on intervention or improvement plans or to remove them from the school (Grissom & Youngs, 2016). These observations have been shown to be useful in increasing teachers value-added scores in years post-observation (Taylor & Tyler, 2012). However, if teachers feel that these subjective measures, which often carry high stakes (Goldring et al., 2015), are perceived to be unfair, they may persuade teachers to exit the profession.

The usefulness and fairness of teacher evaluation systems has been widely debated (Duffett, Farkas, Rothertham, & Silva, 2008; Weisberg, Sexton, Mulhern, & Keeling, 2009), but research shows that principals find the new systems useful for how they approach managing their teachers. Principals are especially focused on the information collected during teacher

observations because they believe it shows a more holistic view of teacher's performance (Goldring et al., 2015). However, principals' ratings are not strictly based on what is observed in classrooms. For example, some principals give high value-added teachers low ratings when they perceive these teachers as exerting too little effort or contributing little to the school community (Harris, Ingle, & Rutledge, 2014). Also, principals frequently depart from their true assessments of a teacher when assigning an observation rating depending on whether the evaluation is high or low stakes, and they may be more likely to give more positive ratings to novice teachers than to more experienced ones (Grissom & Loeb, 2017).

Both black and white ratees are rated lower when evaluated by someone of a different race (Elvira & Town, 2001)—this likely impacts Black teachers differentially as principals, or raters, or more likely to be White. In their comparison of principals' low- and high-stakes ratings of the same teachers, Grissom and Loeb (2017) found evidence of more positive bias in personnel ratings for White and Hispanic teachers than for Black teachers. One way to mitigate the impact of observer biases on observation scores and increase reliability would be to have the observation scored by more than one observer (Johnson, Penny, & Gordon, 2000). This however is not feasible in many schools and may lead to higher rates of teacher turnover as employees are more likely to leave their job if they perceive lower ratings to be a product of manipulation due to rater biases (Poon, 2004).

In a study on how raters determine between two choices on a scale, Lumley (2002) finds that extraneous factors that are not included in the rubric impact scores given when evaluators are unable to decide between two score points or they rely on comparisons to previous subjects. This may translate to principal observations when asked to rate teachers on a scale as scores are not entirely objective. Differences in ratings may be a result of other extraneous factors such as

what time of day observations are conducted (Graham, Milanowski, & Miller, 2012); however the systematically lower observation scores that Black teachers suggest there may be other factors involved.

Teachers of color may be disadvantaged in teacher evaluation ratings due to classroom composition. Black teachers are more likely to teach low-income and low-achieving students within their schools (Kalogrides, Loeb, & Beteille, 2013). Teachers in classrooms with higher concentrations of low-income and racial minority students receive lower evaluation ratings (Borman & Kimball, 2005). One explanation is that students with higher achievement levels are better able to demonstrate evaluation rubric items such as formulating questions and responding to teacher questions (Whitehurst et al., 2014). Boykin (1994) studied cultural styles of African-American students, and finds characteristics such as verve, mobility, oral tradition, communalism, but because these do not fit the traditional schooling model, they are seen as hyperactivity, immaturity, and a sign of low cognitive ability. Cultural incongruence may lead White raters to dock Black teachers for not having strong classroom management based on a rubric that does not account for possible cultural differences.

The arbitrariness of whether some teachers are rated more positively or negatively is especially concerning if teachers of color are more likely to receive lower scores from their observations. Numerous studies across professions find that Black employees face bias in performance ratings (e.g., Kraiger & Ford, 1985). Studies in education settings find that Black teachers and male teachers are more likely to be rated below acceptable (Bailey, Bocala, Shakman, & Zweig, 2016) and that when rated by a Black and White supervisor, the White supervisor tends to rate them more negatively (Stauffer & Buckley, 2005; Kraiger & Ford, 1985). These differences in ratings based on rater race are also found when worker productivity

is controlled for (Baldi & McBrier, 1997; Elvira & Town, 2001). Research linking lower evaluation scores to lower satisfaction and higher turnover rates means that evaluation bias may contribute to differences in the mobility rates of Black and White teachers (Grissom & Bartanen, 2017; Koedel, Li, Springer, & Tan, 2017).

Data

This study analyzes administrative data from Tennessee, a state made up of 146 districts operating roughly 1,800 schools that serve 996,000 students. Thirty-one percent of the state's students are Black or Hispanic and 58% are eligible for the federal subsidized lunch program.[2] Tennessee was a first-round winner of the Obama administration's Race to the Top competition, and instituted a number of educational reforms under its auspices under the state's "First to the Top" legislation and reform initiatives. Particularly useful for this study, these reforms included implementation of a statewide educator evaluation system, the Tennessee Educator Acceleration Model (TEAM), beginning in the 2011-12 school year. Similar to the evaluation systems adopted throughout the country, TEAM assigns teachers a summative evaluation rating based on a weighted combination of components that include classroom observation scores, individual/school value-added scores, and achievement scores. We describe these components further below.

For this analysis, the Tennessee Department of Education (TDOE) provided us with longitudinal administrative data files covering all public education personnel in the state from 2001–02 to 2015–16 via the Tennessee Education Research Alliance (TERA) at Vanderbilt University. These files contain information about employees' personal and professional characteristics, including job positions, gender, race and ethnicity, years of experience, and

[2] https://www.tn.gov/education/topic/report-card

highest degree earned. We use these files to construct additional experience measures, such as years employed in their current school. Beginning in 2006–07, we merge these data with information on the characteristics of the schools and districts in which teachers and principals currently work from annual student demographic, enrollment, and achievement data from TDOE. While our descriptive analyses of teacher workforce composition and teacher turnover span 2002–03 to 2015–2016, our teacher turnover regression models exclude school years prior to 2006–07 because we do not have school-level student demographic and achievement data. Additionally, our analyses of evaluation ratings start in 2011–12, the first school year in which TEAM was implemented. Table 1 contains descriptive statistics for Tennessee teachers.

Figure 1 shows the composition of Tennessee's teacher workforce across all years of available data. In addition to the patterns for White, Black, and other race/ethnicity teachers, we also show the intersection of race and gender groups. As noted in Table 1, the Tennessee teacher workforce is overwhelmingly White and female. Depending on the year, 65 to 70 percent of teachers are White women. White men are the second largest group at slightly under 20 percent. Similar to White teachers, Black teachers are much more likely to be female, comprising approximately 15 percent of the workforce. Black men are the least represented group (among Black and White teachers) at roughly three percent. The number of non-White, non-Black teachers in Tennessee is extremely low. In any given year, they constitute less than one percent of the teacher workforce.[3] The most striking aspect of Figure 1 is the lack of change in the composition of Tennessee's teacher workforce across the last fifteen years. The percentage of Black and White teachers has remained flat over time, and there are little to no changes when

[3] Because of this, we henceforth focus our analysis on Black and White teachers.

looking at the intersection of race and gender. Similar to the national numbers, non-White teachers are severely underrepresented in Tennessee relative to the share of non-White students.[4]

Operationalizing Teacher Turnover

From the longitudinal administrative data files, we create both binary and categorical teacher turnover variables. The binary variable takes a value of 1 if a teacher leaves his or her school during or immediately following year *t* (i.e., they are not a teacher in the same school in year *t+1*). The categorical indicator includes four turnover types (not including teachers who stay in their schools): moving to another school in the same district, moving to a school in a different district, changing positions (e.g., instructional coach, assistant principal), and leaving the education system. Pooling all available years of data, the annual teacher turnover rate in Tennessee is roughly 17 percent (see Table 1). Transfers comprise the largest portion of turnover; 6.3 percent of teachers move to another school in the same district and 2.2 percent move across districts. Exits are also a large share of turnover, with 7 percent of teachers leaving the education system annually.[5] Finally, approximately one percent of teachers change positions each year.

Measures of Teacher Performance

As outlined above, Tennessee teachers receive summative evaluation ratings as part of the statewide educator evaluation system implemented in 2011–12. These scores—termed as

[4] Also similar to national patterns is that the percentage of non-White students in Tennessee is growing over time. In 2006–07, 32 percent of students were non-White, compared to 36 percent in 2015–16. This increase is driven by Hispanic/Latino students (4.7 percent in 2006–07 to 9.2 percent in 2015–16).

[5] Among this group, approximately 15 percent eventually return to the education system after an absence of one or more years.

"level of effectiveness" (LOE)—take a discrete value from 1 to 5 and are generated according to an underlying composite score constructed from weighted combinations of observation, growth, and achievement measures. Classroom observation scores receive the greatest weight in determining teachers' summative evaluation ratings. For the typical teacher, their average observation score accounts for 50% of their LOE.[6] For teachers in tested grades and subjects (i.e., grades 3-8 in core subjects and some select high school courses), 35% of the summative evaluation comes from the teacher's individual TVAAS score, which is a value-added measure of student performance on end-of-year statewide exams. For teachers not in tested grades and subjects, their growth component is the schoolwide TVAAS score. Finally, the remaining 15% comes from an achievement measure, which is typically the same for all teachers in a school. Achievement measures vary greatly between schools, but most common are levels measures of student achievement in math and reading.

Methods

We conduct three main analyses in this article. First, we quantify the teacher turnover gap between Black and White teachers and examine the extent to which differences in teacher characteristics (e.g., education level, experience) and school characteristics (e.g., student demographics, locale) explain the gap. We then conduct a parallel analysis that examines gaps in evaluation scores. Finally, we evaluate whether race-based disparities in evaluation scores contribute to turnover disparities. We describe each of these analyses below.

For our teacher turnover analyses, we estimate the following linear probability model:

$$\Pr(Teacher\ turnover)_{ijt} = T_{ijt}\phi + S_{jt}\eta + \tau_t + \epsilon_{ijt} \qquad (1)$$

[6] Depending on the teaching assignment, the availability of prior student achievement data, and the school year, the weight given to the observation score ranges from 50-85%. For more information, see http://team-tn.org/.

where the probability that teacher i leaves their position in year t is a function of teacher characteristics T (race, gender, age, experience, and highest educational attainment), school characteristics S (achievement index, enrollment, proportion of Black and Hispanic/Latino students, proportion of students qualifying for free/reduced price lunch, proportion of gifted students, proportion of students with disabilities, school level, and school locale), and year fixed effects τ_t. In addition to a binary teacher turnover model, we also estimate multinomial models that differentiate among three types of turnover: transferring to another school (combining within- and across-district moves), exiting from the education system, and changing to a non-teaching position.

Our focus is estimating the difference in turnover rates between Black and White teachers while accounting for average differences in their other characteristics and school context. In equation 1, this difference is represented by the coefficient on an indicator for whether a teacher is Black. However, there are several reasons why estimating equation 1 may produce misleading results. First, there may be unobserved school factors that contribute to higher or average lower turnover rates among the teaching staff and that are uncorrelated with the school characteristics included in the model. If such factors are present in schools primarily staffed by Black teachers (or vice-versa), they will bias the estimated turnover gap between Black and White teachers. To the extent that these factors are fixed over time, we can control for them by adding school fixed effects to equation 1. This effectively isolates the comparison to Black and White teachers who work in the same school.

However, Black and White teachers may be *differentially responsive* to school context. For instance, Black and White teachers may have different preferences for working with Black students and/or working in rural schools. Failing to account for race-based differences in

responsiveness to school context could also bias our estimate of the Black-White teacher turnover gap, even in models that include school fixed effects.[7] To account for differential responsiveness, we add to equation 1 interactions between teacher race and each of the school characteristics. To find the residual Black-White turnover gap, we estimate the contrast between Black and White teachers.

To identify Black-White gaps in evaluation ratings, we estimate versions of equation 1 that replace teacher turnover with each of the five evaluation measures (LOE, average observation score, growth score, individual growth score, and achievement score). Similar to the teacher turnover models, we account for unobserved school heterogeneity and/or differential responsiveness by estimating models that include school fixed effects or interactions between teacher race and each of the school characteristics. The justification for interactions remains the same. While prior work demonstrates that student demographics affect teachers' evaluation ratings (particularly subjective measures like classroom observations), Black and White teachers may be differentially affected. An alternative possibility is that teachers' actual effectiveness is affected by school context, but that the relationships are different for Black and White teachers. Finally, when examining observation scores, we also estimate models that include controls for teachers' individual value-added score (TVAAS composite). This accounts for the possibility that disparities in observation scores between Black and White teachers are a function of differences in effectiveness (specifically, the ability to improve student test scores).

[7] We can also estimate models that include both school fixed effects and interactions between race and school characteristics. We prefer a specification without school fixed effects when including interactions, as school fixed effects restricts the identification of differential responsiveness to within-school variation. Even over a relatively long panel, student demographics are fairly stable within schools, which limits our ability to estimate the interactions.

The third part of our analysis examines the extent to which Black-White disparities in evaluation scores mediate disparities in teacher turnover rates. For the given evaluation measure, we first re-estimate equation 1 for the sample of teachers without missing scores and calculate the contrast between Black and White teachers. Then, we simply control for the evaluation measure in the turnover model, such that any change in the Black-White contrast provides evidence of mediation. We repeat this procedure for using of the evaluation measures for both binary and multinomial turnover outcomes.

Results

We begin by documenting yearly average turnover rates for Tennessee teachers. Figure 2 shows binary turnover rates and Figure 3 shows rates for specific turnover types. In Figure 2, the solid line shows the overall turnover rate, but we also show breakdowns for race and the combination of race and gender. In Tennessee, the teacher turnover rate ranges between 15 and 20 percent. The overall upward trend in recent years coincides with the passage of teacher tenure reforms and the implementation of the statewide educator evaluation system. The turnover rate for White teachers tracks the overall rate closely, as White teachers are a strong majority in the workforce. White women are the least likely to leave their positions, while White men have turnover rates slightly above the overall rate.

The most obvious pattern in Figure 2 is that Black teachers have substantially higher turnover rates than White teachers. Pooling across years, the gap is approximately 6 percentage points. Turnover rates for Black teachers also fluctuate more than for White teachers, as Black teachers are a smaller group work in a smaller number of districts, both of which make year-to-year shocks in turnover more apparent. The largest shock occurs in the 2013–14 school year, which is in the midst of a merger and subsequent de-merger between the Memphis City and

Shelby County school districts, which collectively account for the majority of Black teachers in Tennessee. Similar to the pattern for White teachers, Black men are more likely to leave their positions than Black women.

Figure 3 shows that Black teachers' increased propensity to move to another school in their district almost completely accounts for the turnover disparity between Black and White teachers. Black teachers are more than twice as likely to transfer schools (11 percent vs. 5 percent), with no consistent difference between men and women. One important factor is that, on average, Black teachers work in larger, more urban school districts than White teachers, which yields more opportunities for intra-district movement. Across-district moves are much less common than within-district moves, and the difference between Black and White teachers is less clear. Prior to the implementation of evaluation reforms in 2011–12, White male teachers consistently had the highest rates of across-district movement (roughly 3 percent each year), while Black women had the lowest rates (1.5 percent each year). In the post-evaluation reform years, the overall rate of across-district movement has increased, with larger increases for Black teachers.

An oft-repeated concern is that teachers of color may leave the profession at higher rates than White teachers, which ultimately hampers efforts to diversify the teacher workforce. In Tennessee, the evidence for this claim is mixed. Outside of a spike in exits in 2009, Black women tend to exit the education system at lower rates than any other group, including during the post-evaluation years. On the other hand, Black men have the highest exit rates among the four race-by-gender groups, except for the two most recent years. White men and women consistently exit the education system at rates of roughly 7 percent each year.

Pulling together the descriptive findings from Figures 2 and 3, we identify two important takeaways about teacher turnover in Tennessee. First, while *turnover* is much greater among Black teachers than White teachers, *attrition* (i.e., teachers who leave the profession) is roughly equal. Thus, the widening racial/ethnic diversity gap between students and teachers is more a function of the lack of racial/ethnic diversity among new teachers entering Tennessee's workforce than failure to retain teachers of color after they enter. That is not to say that higher rates of transfer are not problematic, however. From the standpoint of a school leader, frequent staff turnover may harm school performance. Prior research demonstrates that teacher turnover is particularly detrimental in schools serving disadvantaged populations—the types of schools more likely to be staffed by Black teachers in Tennessee. Additionally, higher rates of transfer among Black teachers may signal dissatisfaction with working conditions, school climate, or other factors. Besides raising turnover rates among Black teachers, these factors could also discourage entry into the teaching profession.

The second takeaway is that, in addition to examining differences by race, it is also important to consider the intersection of race and gender. While men are more likely than women to leave their positions overall, we find evidence that Black men are particularly more likely to turn over. Further, while Black men constitute the lowest percentage of teachers in Tennessee (not counting the minute number of non-Black, non-White teachers), they also leave the education system at the highest rates in most of the years we observe. Thus, in addition to exploring the factors associated with race-based disparities in turnover, we also examine how these patterns may change for gaps across the combination of race and gender.

Differences in Individual and Work Contexts for White and Black Teachers

Up to now, we have focused purely on describing differences in turnover rates between Black and White teachers. On their own, these patterns provide useful information about Tennessee's teacher labor market and the mismatch between the demographic composition of the workforce and the students they serve. However, it is also important to investigate the drivers of this turnover gap. In the current analysis, we investigate three sets of factors that may contribute to the turnover gap: (1) differences in individual characteristics, (2) differences in work contexts, and (3) differences in evaluation ratings.

Table 2 shows averages of individual and school characteristics by teacher race, teacher gender, and the intersection of race and gender. We do not compute significance tests between groups, as we have the population of Tennessee teachers. One potential explanation for differences in turnover rates is that Black and White teachers have different demographic traits (besides race). If, for example, Black teachers are older or have more experience than White teachers, their elevated turnover rates could be explained by greater rates of retirement.

Comparing the first two columns (Black vs. White teachers), we find little to no substantive difference in age or experience. The average Black teacher is roughly one year older than the average White teacher, but both groups have an average of 12 years of experience in the Tennessee education system. As would be expected given higher transfer rates among Black teachers, the average White teacher has 5 years of experience in her current school, compared to 4 years for the average Black teacher. Finally, Black teachers in Tennessee have greater average educational attainment than White teachers. Overall, there is little evidence that Black and White teachers differ dramatically in terms of observable characteristics.

Similarly, there are few substantive differences in teacher characteristics between men and women. Men are slightly experienced than women, but the difference is less than a year.

There is more variation when looking at the intersection of race and gender. While White women and White men have similar characteristics, Black women and Black men are different. Specifically, Black men have two fewer years of experience than Black women, who have the most experience of any group. This corresponds to the patterns in exit rates; Black men were the most likely to exit the education system while Black women were the least likely. Higher educational attainment among Black teachers is completely driven by Black women. More than 70% of Black female teachers have a master's degree or higher in Tennessee, compared to 53% of Black male teachers.

While Black and White teachers have similar individual characteristics, there are stark differences in the average contextual characteristics of the schools in which they work. Comparing the average achievement index (i.e., the school average of student test score levels), the average White teacher works in a school with students scoring 0.23 SD above the state average, while the average Black teachers works in a school where students score 0.68 SD below the state average. There are also large differences in the average composition of the student body. The average White teacher works in a school where 18 percent of the students are Black, compared to 67 percent for the average Black teacher. In addition to working in schools with lower achievement levels, Black teachers also work in schools that serve more low-income students (77% versus 57% of students qualifying for free/reduced price lunch, respectively). Finally, three quarters of Black teachers work in urban schools, compared to only one quarter of White teachers. Turning to gender, the only difference is that men are far more likely than women to work in high schools (52% versus 19%), and less likely to work in elementary schools (20% versus 59%). Once accounting for the average differences by race and gender, there are no additional patterns when looking at the intersection of race and gender.

Given prior findings that teacher turnover tends to be correlated with school characteristics, we expect that these average differences between Black and White teachers explains some of the turnover disparity. Specifically, schools with lower average achievement and larger proportions of disadvantaged students tend to have higher teacher turnover rates, which is consistent with our descriptive findings. However, we also consider the possibility that Black and White teachers are differentially responsive to these contextual characteristics.

Differences in Evaluation Ratings for White and Black Teachers

Table 3 shows average ratings on each component of the teacher evaluation system by race, gender, and the intersection of race and gender. Each measure is on a 1 to 5 scale. On average, Black teachers score lower than White teachers across all measures. For example, the average White teacher in Tennessee receives a 3.84 (out of 5) on their summative evaluation (LOE), compared to 3.63 for the average Black teacher, a difference of roughly 0.2 standard deviation units. The difference in average observation scores is .09 points, or 0.15 standard deviation units. Similarly sized gaps exist for growth and achievement scores.

Except for achievement scores, male teachers score lower than female teachers across all components. In the case of average observation scores, the male-female gap is larger than the Black-White gap (roughly one-third of a standard deviation). However, looking at the patterns by the intersection of race and gender reveals even larger disparities for Black men than would be suggested by simply adding the Black-White and male-female gaps. Compared to White women, Black men receive summative evaluation ratings almost half a point lower (~0.5 SD), on average). This disparity is driven by average differences in observation scores, where Black men score well below any other group, on average.

While the large disparities in evaluation ratings deserve attention, we must exercise caution in directly interpreting these findings. As demonstrated in Table 2, there is a tremendous amount of teacher sorting in Tennessee, and differences in school context may contribute to disparities in evaluation ratings. Nonetheless, the combination of findings in Tables 2 and 3 suggests that higher turnover rates among Black teachers may be at least partially a function of receiving lower evaluation ratings. In the next section, we more formally test this hypothesis in a regression framework that accounts for the confounding role of school context.

Multivariate Analysis of Disparities in Turnover and Evaluation Ratings

We begin our multivariate analysis by estimating the Black-White gap in teacher turnover while conditioning on other factors, such as individual and school contextual characteristics. Table 4 shows the results of estimating Equation 1. The top row of the table (Panel A) shows the estimated contrast in teacher turnover between Black and White teachers, while the estimated coefficients are shown below (Panel B). In models with no interactions, the contrast is simply the coefficient on an indicator for whether the teacher is Black.

Column 1 shows the descriptive difference in teacher turnover rates between Black and White teachers (6.2 percentage points). Adding teachers' other individual characteristics (gender, highest degree, age, and experience) attenuates the gap only slightly (now 5.8 percentage points), which is unsurprising given the small average differences in these characteristics between Black and White teachers (Table 2). It is worth noting that the estimated coefficients are precisely estimated and align with existing studies of teacher turnover. Both age and experience have a "U-shaped" relationship with turnover; teachers are less likely to leave their positions as they become older and gain experience, until they reach retirement age. Male teachers and teachers with advanced degrees are also more likely to turn over.

Column 3 adds both school contextual characteristics and teacher characteristics to the model. Once accounting for differences in school context, the turnover disparity is completely erased and actually becomes negative. In other words, Black teachers are less likely to leave their positions once you have controlled for the characteristics of the schools in which they tend to work. Examining the model coefficients, we find that teachers are more likely to leave their positions in schools with lower average achievement levels and larger proportions of Black and Hispanic/Latino students. That Black teachers work in schools with far more Black students and lower average achievement explains the complete reversal in the estimated Black-White turnover gap when controlling for these characteristics. Even with the rich set of school characteristics included in the model, there may be unobserved school factors that confound our estimate of the Black-White turnover gap. To examine this possibility, we add school fixed effects in Column 4. The estimated turnover gap is virtually identical to Column 3, which suggests that our included school characteristics adequately control for any (time-invariant) school factors that are both correlated with teacher race and a teacher's propensity to leave their position.[8]

The results in Columns 3 and 4 demonstrate that teachers are very responsive to the contextual characteristics of schools. However, prior studies suggest that Black and White teachers may respond differentially to school context, which would imply that our previous models are misspecified. We test for differential responsiveness in Columns 5 and 6 by including interactions between teacher race and school contextual characteristics. Column 5 includes only a single interaction with the proportion of Black students in the school, while Column 6 includes interactions with the full set of school controls. In Column 5, we see that the interaction between Black teacher and the proportion of Black students is negative and highly significant. Further, the

[8] There still could be time-varying school factors that are uncorrelated with our school controls. This is a limitation of our analysis (and most analyses of turnover).

magnitude of the interaction is more than half the size of the main effect, which provides strong evidence that Black teachers respond differentially to the proportion of Black students in their school. Once we take into account this differential responsiveness, the estimated Black-White turnover gap again becomes positive (1.3 percentage points). Column 6 shows that Black teachers are differentially responsive to most of the school characteristics, and the Black-White turnover gap grows larger once we account for these interactions.

We further demonstrate the concept of differential responsiveness (with regards to turnover) in the top row of Figure 4, which plots the predicted probability of teacher turnover as a function of the proportion of Black students in the school. Each graph has three lines; the solid line plots the prediction for the "average" teacher (i.e., pooling Black and White teachers), while the dashed lines show Black and White teachers separately. Panel 1a shows predictions from a turnover model that includes teacher characteristics, the proportion of Black students in the school, and an interaction between teacher race and the proportion of Black students. Panel 1b adds the rest of the school controls, and Panel 1c adds the school controls and the full set of interactions.

The graphs demonstrate a consistent point about differential responsiveness and the implications for estimating the Black-White turnover gap. Even in the simplest model (1a), the estimated relationship between turnover and the proportion of Black students is different for Black and White teachers. While lines for White and Black teachers are both upward sloping, the line for Black teachers is flatter, suggesting that they are less averse than White teachers to working in schools with larger numbers of Black students. In fact, once accounting for other school factors and Black teachers' differential responsiveness to those factors, turnover rates for Black teachers actually *decrease* in schools with larger proportions of Black students (Panel 1c).

Figure 4 also demonstrates why the models without interactions produce a misleading conclusion about the Black-White turnover gap. Because White teachers greatly outnumber Black teachers in Tennessee, a model that excludes interactions estimates an "average" slope that is much closer to the slope for White teachers. This model effectively over-controls for the relationship between Black students and turnover, given that Black teachers are concentrated in schools with large numbers of Black students. For example, Panel 1c shows that in a model without interactions the predicted turnover rate for a teacher in a school with 80 percent Black students is roughly 23 percent. This overstates the likelihood of turnover for Black teachers in such schools, however, as their predicted turnover rate is only 18 percent in the correctly-specified model.

Based on these findings, we take Column 6—a model with the full set of interactions and without school fixed effects— as our preferred specification for estimating the Black-White gap in turnover.[9] The estimated contrast in Column 6 shows that once accounting for teacher and school contextual characteristics, there is a residual Black-White turnover gap of 2.5 percentage points. Put another way, much of the descriptive difference in teacher turnover between Black and White teachers is explained by differences in their individual characteristics and the characteristics of the schools in which they work (much more the latter), but roughly 40 percent of the gap remains unaccounted for in our models.

Table 5 shows the results for a multinomial turnover outcome. Specifically, we model Black-White gaps in transfers, exits, and position changes.[10] For each outcome, we show the

[9] While the model excludes school fixed effects, which could introduce bias from unobserved school heterogeneity, the consistency between Columns 3 and 4 suggests that this is not a major problem. Specifically, while school fixed effects contribute to model fit, it does not appear that they alleviate any bias in estimating the Black-White contrast. As a check, we also estimated models that include interactions and school fixed effects. The results are fairly consistent with Column 6, though the Black-White gap is closer to 0.

[10] Due to the low number of across-district moves, we combine within- and across-district moves into a single category. Results are not substantively different when modeling these outcomes separately.

unadjusted gap (no covariates in the model) and the preferred specification from Table 4. Column 1 shows that Black teachers are 6.6 percentage points more likely than White teachers to transfer to another school. Once we account for individual and school contextual characteristics in Column 2, this gap falls to 1.8 percentage points. Turning to exits, there is a small descriptive gap between Black and White teachers (0.4 percentage points) that increases slightly in the preferred specification (0.8 percentage points). Black teachers are also slightly more likely to change positions (0.6 percentage points), but the gap does not change appreciably when controlling for individual and school characteristics.

Next, we model disparities in evaluation ratings between Black and White teachers. Our primary focus is on observation scores, as they receive the greatest weight in determining the summative score and are the only measure that reflects individual performance across the full population of teachers. In Table 6, we estimate the Black-White contrast in observation scores across several specifications. Like the turnover models, we show the estimated contrast for each specification in Panel A, with the model coefficients shown in Panel B. Column 1 shows that, on average, Black teachers receive observation ratings that are 0.15 standard deviations lower than White teachers. Once we control for teacher characteristics in Column 2, the gap grows to 0.18 SD. One important finding in Column 2 is that men receive substantially lower observation scores than women (0.30 SD), and this gap is larger than the race-based gap. Other teacher characteristics are highly predictive of observation scores in the ways one might expect. Teachers with more education and more experienced teachers receive higher observation scores, on average. For example, compared to teachers with 15 to 24 years of experienced, teachers with less than five years of experience receive scores approximately half of a standard deviation lower.

Next we control for both individual and school contextual characteristics (Column 3). Like with the turnover models, controlling for school context greatly attenuates the gap between Black and White teachers (the gender gap attenuates only slightly). The coefficients on the school characteristics demonstrate that teachers in more advantaged schools tend to have higher observation scores. For instance, a one standard deviation increase in a school's average test score level is associated with a 0.14 standard deviation increase in a teacher's average observation score. Similarly, teachers have higher observation scores in schools with more gifted students, while higher proportions of low-income and Hispanic/Latino students are associated with significantly lower observation scores. Middle school and high school teachers have scores approximately 0.17 SD below elementary school teachers, on average. If we isolate within-school variation by including school fixed effects (Column 4), the Black-White observation score gap increases to slightly larger than in the uncontrolled model (0.16 SD). Even in the school fixed effects model, the coefficients on the proportion of Hispanic/Latino students and students qualifying for free/reduced price lunch are negative and significant, which aligns with prior findings that teachers' observation scores reflect in part the demographic characteristics of their students (Steinberg & Garrett, 2016).

Unlike the teacher turnover models, including school fixed effects changes substantively the magnitude of the observation score gap between Black and White teachers, which suggests that the included school contextual characteristics do not fully capture school-level heterogeneity that is correlated with teacher race and observation scores. Specifically, models that exclude school fixed effects may understate the magnitude of the disparity.

In Column 5 we test whether the Black-White disparity in observation scores is explained by differences in teacher value-added (TVAAS). Once controlling for teachers' direct impacts on

student achievement, any residual disparity in observation scores between Black and White teachers provides stronger evidence of systematic bias on the part of observers. The TVAAS coefficient shows a strong relationship between a teacher's observation scores and their ability to raise student achievement—a one standard deviation increase in the TVAAS composite is associated with a 0.32 SD increase in average observation score. Importantly, differences in value-added explain only a small portion of the Black-White gap in observation scores. Comparing teachers whose students achieved the same level of growth, Black teachers still received observation scores 0.13 SD lower than White teachers in the same school.[11]

Similar to the turnover models, the influence of school context on observation scores may be different for Black and White teachers, which necessitates the inclusion of interaction terms between teacher race and the school characteristics. Column 6 shows several statistically significant and substantively important interactions. For instance, while White teachers' observation scores are uncorrelated with the proportion of Black students in their school (conditional on the other school characteristics), Black teachers receive substantially higher scores in schools with large numbers of Black students. We demonstrate this relationship graphically in the bottom row of Figure 4. Holding all else equal, a Black teacher working in a school with no Black students has a predicted observation score roughly half of a standard deviation below a White teacher. As the proportion of Black students increases, the observation scores of Black teachers rise precipitously. In schools with a majority of Black students, Black teachers outscore White teachers. Again, the "average teacher" line demonstrates why omitting interactions leads to an underestimation of the Black-White gap. Given that more than 80 percent

[11] To ensure the differences between columns 4 and 5 are not driven by sample selection (since only a subset of teachers have individual TVAAS scores), we checked whether the estimated contrast in column 4 changes if the sample is restricted to those in column 5. The Black-White gap is essentially identical for these teachers.

of Tennessee teachers are White, the average slope for the relationship between observation score and proportion of Black students is heavily informed by White teachers, despite Black teachers performing substantially worse in these schools.

We find other significant interactions as well. Black teachers receive higher scores in urban schools and schools with larger proportions of Hispanic/Latino students, with no relationship for White teachers. Both Black and White teachers receive higher observation scores in schools with more gifted students, but the magnitude of the relationship is more than twice as large for Black teachers. Black teachers also score disproportionately lower in middle schools and higher in urban schools. Once we account for these interactions with school context, the estimated Black-White contrast grows to 0.34 SD. In Column 7 we control for teachers' individual TVAAS scores and again find that their inclusion has little effect on the magnitude of the observation score gap.

In Table 7 we show the estimated Black-White contrast (in SD units) for the other evaluation measures. The first column shows the descriptive difference, the second column includes controls and school fixed effects (mirroring Column 4 in the previous table), and the third column replaces school fixed effects with interactions (mirroring Column 6 in the previous table). Descriptively, Black teachers score lower on all measures in the evaluation system. When controlling for teacher and school controls and including school fixed effects, the disparities become smaller in magnitude, though all but the growth score remain statistically significant. In the preferred specification (interactions instead of school fixed effects), Black teachers score roughly 0.10 to 0.15 SD below White teachers, on average. The exception is growth scores; Black and White teachers score similarly once controlling for school contextual characteristics.

Do Differences in Evaluation Ratings Mediate Differences in Turnover Rates?

The previous analyses establish that, even when controlling for differences in individual and school contextual characteristics, Black teachers have higher average turnover rates and lower average evaluation scores than White teachers. Next, we test the extent to which higher turnover rates among Black teachers are mediated by each of the evaluation measures by estimating turnover models with and without controls for evaluation scores. Table 8 displays two estimated contrasts in turnover for each combination of turnover type (columns) and evaluation measure (rows). We estimate the first contrast using the preferred specification from Table 4 for the sample of teachers with a non-missing evaluation score. The second contrast controls for the given evaluation score. Attenuation of the Black-White provides evidence that evaluation scores mediate turnover disparities between Black and White teachers.

Looking across evaluation measures, we find evidence of mediation with observation scores. Comparing Columns 1 and 2 (results for a binary turnover outcome) in Panel A, the estimated contrast decreases from 3 to 2 percentage points when controlling for observation scores[12], which suggests that lower observation scores among Black teachers partially explain their higher turnover rates. Looking across turnover types (Columns 3 through 8), this mediation is concentrated among transfers and exits. Without controlling for observation scores, Black teachers have predicted transfer rates 2.6 percentage points greater than White teachers, which drops to 2.0 percentage points when controlling for observation scores. While the contrast in exits is not statistically significant in either model, the absolute change in the coefficient between the baseline and mediation model is actually larger (0.008) than for transfers. Finally, while Black teachers change positions at significantly higher rates, controlling for their observation

[12] The contrast in Column 1 varies among the evaluation measures because the sample is restricted to teachers who have non-missing scores for the given evaluation measure. For example, among teachers who have individual growth scores, the estimated turnover gap between Black and White teachers (without controlling for the individual growth score) is 3.6 percentage points.

scores does not appreciably change this gap. Overall, we find evidence that lower observation scores lead Black teachers to transfer schools and exit the education system at higher rates.

In contrast, we find little evidence of mediation for growth and achievement scores. Despite the finding in Table 7 that Black teachers receive lower scores on all components of the evaluation system, differences in growth and achievement scores do not appear to drive differences in turnover rates by race. Comparing the baseline and mediation models across turnover types, there are little substantive changes in the estimated Black-White contrasts. The differences in Panel E, which examines teachers' LOE scores, are roughly half the size of those in Panel A, which makes sense given that observation scores typically account for half of teachers' summative ratings.

Even with Black teachers receiving lower average scores on all evaluation measures that produce a teacher's summative rating, only observation scores show evidence of mediating the relationship between teacher race and turnover. There are several potential explanations for this finding. First, even after the wave of reforms to evaluation systems, very few teachers receive ratings that would subject them to administrative removal. Tennessee is no different; less than five percent of teachers receive the lowest summative evaluation rating each year. To the extent that a relatively small number of teachers are removed through administrative processes, we would not expect the modest differences in growth and achievement scores to drive turnover differences. Unfortunately, we cannot directly observe whether teacher's leave their schools via administrative removal.

More likely is that observation scores are more salient to teachers and principals in terms of signaling job performance. For teachers, observation scores not only receive the greatest weight in the evaluation system, but provide immediate, direct, and personalized feedback about

performance. Additionally, most classroom observations are conducted by the school principal or assistant principal, such that the source of feedback is someone with whom the teacher regularly interacts. In contrast, for the majority of teachers, growth and achievement scores are based on school-level performance on standardized tests, [13] rather than the performance of their individual students, and the scores are typically not available until after the end of the school year. Receiving low observation scores, then, may constitute a much stronger negative signal to teachers than receiving low growth or achievement scores. For principals, prior work finds that when making human capital decisions, principals give more weight to teachers' observation scores than value-added or achievement scores (Goldring et al., 2015). Additionally, in other work we find that effective principals experience lower rates of turnover among teachers with high observation scores, but also *higher* rates of turnover among teachers with low observation scores (Grissom & Bartanen, 2017). This type of strategic retention holds only for observation scores—no such pattern exists for growth or achievement scores.

While we cannot say definitively whether the connection between observation scores and turnover is driven by the responsiveness of teachers versus principals, we can present suggestive evidence by drawing on survey data that assesses teacher job satisfaction. By re-estimating our mediation models using satisfaction as the outcome, we can examine the extent to which differences in observation scores mediate differences in satisfaction between Black and White teachers. Table 9 shows the estimated contrasts in job satisfaction between Black and White teachers. Comparing Columns 1 and 2 shows how the gap changes once we control for differences in each evaluation measure. Column 1 shows that, conditional on the full set of teacher and school controls, Black teachers report slightly higher job satisfaction than White

[13] Some schools and districts use alternative measures, such as graduation rates.

teachers. Panel A shows that once we control for observation scores, the gap grows from 0.04 to 0.09 SD. In other words, lower average observation scores among Black teachers lowers their satisfaction relative to White teachers. The remaining panels show little to no substantive changes in the satisfaction gap when controlling for growth or achievement scores. These results provide some suggestive evidence that lower observation scores contribute to higher rates of voluntary turnover among Black teachers.

Examining Patterns by the Intersection of Race and Gender

In our descriptive analyses, we noted that patterns in turnover and evaluation scores varied not just by race and gender, separately, but also by the intersection of race and gender. In Figures 5 and 6, we conduct a parallel set of analyses by estimating contrasts between White women, White men, Black women, and Black men.[14] Figure 5 shows the estimated gaps in observation scores across four specifications: (1) no covariates, (2) teacher covariates only, (3) teacher covariates, school covariates, and interactions, and (4) adding individual TVAAS scores. Each bar shows the difference between the stated group and White women, who are the omitted group.

As outlined in the prior descriptive statistics, White women have the highest average observation scores, such that the bars are negative, indicating lower average scores for White men, Black women, and Black men. For White men, the descriptive contrast is roughly 0.30 SD. Controlling for other teacher characteristics reduces the gap only slightly, but controlling for school characteristics reduces the gap by roughly 25 percent. Even controlling for teacher

[14] In models with interactions, we allow the slopes of school characteristics to vary by the combination of race and gender, rather than just by race. However, we find little evidence that this differential responsiveness varies by gender, such that the results are substantively equivalent if we only include interactions between school characteristics and teacher race.

characteristics, school characteristics, and individual growth scores, White men receive observation scores more than 0.20 SD lower than White women, on average.

Turning to Black women, we start with a descriptive gap of 0.14 SD (relative to White women). Unlike with White men, controlling for teacher and school characteristics increases the magnitude of the estimated contrast. In model 3b (controls for teacher and school characteristics), the observation score gap grows to more than 0.35 SD. In other words, once we account for differences in the types of schools in which White and Black women work (and, implicitly, average differences in teacher observation scores in those schools), Black women fare much worse than when simply comparing average observation scores. Controlling for individual TVAAS scores explains little of the residual gap.

Black men have the lowest average observation scores, and unlike White men and Black women, controlling for individual and school characteristics has no substantive effect on the estimated gap. In all four specifications, Black men score more than half a standard deviation lower than White women, on average.

Figure 6 shows the estimated turnover gaps. Again, White women are the omitted group, and all of the estimated contrasts are positive (higher turnover rates). Models 1–3 mirror the specifications for observation scores and Model 4 adds controls for observation scores. White men have turnover rates roughly one percentage point above White women, and controlling for teacher and school characteristics does not appreciably change this gap. However, as demonstrated in Figure 5, White men have substantially lower observation scores than White women. Comparing models 3a and 4a in Figure 6, controlling for differences in observation scores cuts the turnover gap in half for White men. This suggests that observation scores mediate gender differences in turnover rates.

Considering Black women and Black men separately yields similar findings to those in Tables 4 and 8. Controlling for other teacher characteristics explains little of the turnover gap relative to White women, but controlling for school contextual characteristics and differential responsiveness reduces the estimated gap by 30–40 percent. Finally, comparing models 3 and 4, controlling for observation scores explains roughly one-third of the residual turnover gap.

Overall, Figure 6 demonstrates that differences in observation scores mediate race- and gender-based differences in turnover rates. However, we do not find strong evidence that intersectionality plays a large role. That is, the turnover patterns for Black men are roughly equal to the sum of race- and gender-based differences. Importantly, even after controlling for differences in individual characteristics, school contextual characteristics, and teachers' evaluation scores, there remains a fairly substantial turnover gap between Black and White teachers. Thus, while the evaluation system (or at least the evaluation scores that teachers receive) does contribute to race-based turnover gaps, much of the gap remains unexplained.

Discussion

Growing evidence demonstrates the benefits of a diverse teacher workforce, particularly for students of color. Yet even as students of color now constitute a majority in U.S. K-12 public schools, the teacher workforce remains overwhelmingly White. We examine the extent to which teacher turnover undermines teacher diversity in the state of Tennessee. Counter to some national estimates, we find that Black teachers in Tennessee are no more likely to exit the teaching profession than White teachers. We do find, however, that they are substantially more likely to leave their schools to transfer elsewhere, especially Black male teachers. These higher rates of transfer are an important phenomenon to understand given that Black teachers are more likely to

teach vulnerable students for whom teacher turnover may be especially harmful (Ronfeldt et al., 2013).

Leveraging longitudinal administrative data, we find that few individual characteristics of teachers differ between Black and White teachers, nor do these characteristics explain turnover gaps. Like prior work, however, we find that differences in school contexts matter. Black teachers in Tennessee work in schools that have lower achievement and larger numbers of students from traditionally marginalized groups, and accounting for these factors closes—but does not eliminate—the turnover gap between Black and White teachers.

We then examine differences in teacher evaluation ratings, particularly differences in teacher observation scores, and find that Black teachers receive substantially lower ratings than their White colleagues, even when limiting the comparison to other White teachers in the same school with the same level of student achievement growth. When we account for observation ratings, we can further reduce the residual gap in turnover between Black and White teachers, suggesting that evaluation differences indeed contribute to turnover differences. Importantly, however, inclusion of a large number of individual and school covariates alongside measures of teacher performance cannot fully explain turnover differences between Black and White teachers, which demonstrates the need for further investigation of unobserved factors driving the higher turnover propensities of Black teachers.

Our results have a number of policy implications. First, the descriptive fact that exit rates of Black teachers in Tennessee are not higher than those of White teachers—and in recent years may be lower—focuses attention on increasing pipelines into teaching rather than stemming exits from the profession as perhaps a more important strategy, at least in this state. Nevertheless, the high rates of Black teacher movement highlights the need for policy attention at keeping Black

teachers in their schools. Addressing working conditions on those schools may be one approach. Our results suggest another: addressing differences in how Black and White teachers are assessed by the teacher evaluation system. At this stage, we are unable to say what causes Black teachers' lower evaluation scores, but the fact that Black teachers receiving lower observation ratings—and by extension, lower total evaluation scores—than similar White teachers should give policymakers pause. Further research into the drivers of Black-White evaluation gaps is important, particularly given the relationship we demonstrate between these gaps and higher turnover rates among Black teachers.

Our study faces a number of limitations. First, we analyze data from only one state, so our results may not be generalizable to other contexts. Second, we lack measures of teacher working conditions that we can observe over time (e.g., via surveys), which prevents us from going as deep as some other studies to connect teacher race, school contexts, teacher working conditions, and turnover. Third, we are unable to observe reasons for teacher turnover, including even whether a given turnover event is voluntary or involuntary, which limits our ability to draw firm conclusions about why lower observation scores are associated with higher turnover rates, particularly for Black teachers. Further investigation of the evaluation-turnover relationship among White teachers and teachers of color is a ripe area for future work.

References

Achinstein, B., Ogawa, R. T., Sexton, D., & Freitas, C. (2010). Retaining teachers of color: A pressing problem and a potential strategy for "hard-to-staff" schools. *Review of Educational Research*, *80*(1), 71-107.

Bailey, J., Bocala, C., Shakman, K., & Zweig, J. (2016). Teacher demographics and evaluation: A descriptive study in a large urban district.

Baldi, S., & McBrier, D. B. (1997). Do the determinants of promotion differ for blacks and whites? Evidence from the US labor market. Work and Occupations, 24(4), 478-497.

Banks, J. A. (1994). *An introduction to multicultural education*. Allyn and Bacon Inc., 160 Gould St., Needham Heights, MA 02194.

Borman, G. D., & Kimball, S. M. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps?. *The elementary school journal*, *106*(1), 3-20.

Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, *48*(2), 303-333.

Bristol, T. (2014). Black Men of the Classroom: An Exploration of how the Organizational Conditions, Characteristics, and Dynamics in Schools Affect Black Male Teachers' Pathways into the Profession. Columbia University doctoral dissertation.

Cherng, H. Y. S., & Halpin, P. F. (2016). The Importance of Minority Teachers Student Perceptions of Minority Versus White Teachers. *Educational Researcher*, 45(7), 407-420.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673-682.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (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.

Cole, B. P. (1986). The Black educator: An endangered species. *The Journal of Negro Education*, *55*(3), 326-334.

Dee, T. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, 86(1), 195–210.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter?. *The American economic review*, *95*(2), 158-165.

Dee, T. S. (2015). Social identity and achievement gaps: Evidence from an affirmation intervention. *Journal of Research on Educational Effectiveness*, *8*(2), 149-168.

Desimone, L. (1999). Linking parent involvement with student achievement: Do race and income matter?. *The Journal of Educational Research*, *93*(1), 11-30.

Donovan, M. S., & Cross, C. T. (2002). Minority students in special and gifted education, committee on minority representation in special education. *Washington, DC: National Academy of Education*.

Duffett, A., Farkas, S., Rothertham, A. J., & Silva, E. (2008). Waiting to be Won Over: Teachers Speak on the Profession, Unions, and Reform. *Education Sector*.

Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review*, *45*, 44-52.

Farkas, G., Grobe, R. P., Sheehan, D., & Shuan, Y. (1990). Cultural resources and school success: Gender, ethnicity, and poverty groups within an urban school district. *American sociological review*, 127-142.

Foster, M. (1993). Educating for competence in community and culture: Exploring the views of exemplary African-American teachers. *Urban Education*, 27(4), 370–394.

Fuller, M. L. (1992). Teacher education programs and increasing minority school populations: An educational mismatch. *Research and multicultural education: From the margins to the mainstream*, 184-200.

Gershenson, S., Hart, C., Lindsay, C., & Papageorge, N. W. (2017). The Long-Run Impacts of Same-Race Teachers. IZA working paper.

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.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, *44*(2), 96-104.

Graham, M., Milanowski, A., & Miller, J. (2012). Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings. *Online Submission*.

Grissom, J. A. (2011). Can good principals keep teachers in disadvantaged schools? Linking principal effectiveness to teacher satisfaction and turnover in hard-to-staff environments. *Teachers College Record*, *113*(11), 2552-2585.

Grissom, J.A., Kabourek, S., & Kramer, J.W. (2017). Student-teacher racial and ethnic congruence and the progression of student course-taking in high school mathematics. Vanderbilt University working paper.

Grissom, J. A., Nicholson-Crotty, J., & Nicholson-Crotty, S. (2009). Race, region, and representative bureaucracy. *Public Administration Review*, *69*(5), 911-919.

Grissom, J. A., Kern, E. C., & Rodriguez, L. A. (2015). The "representative bureaucracy" in education: Educator workforce diversity, policy outputs, and outcomes for disadvantaged students. *Educational Researcher*, *44*(3), 185-192.

Grissom, J. A., Rodriguez, L. A., & Kern, E. C. (2017). Teacher and principal diversity and the representation of students of color in gifted programs: Evidence from national data. *The Elementary School Journal*, *117*(3), 396-422.

Grissom, J. A., Viano, S. L., & Selin, J. L. (2016). Understanding employee turnover in the public sector: Insights from research on teacher mobility. *Public Administration Review*, *76*(2), 241-251.

Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. *American Educational Research Journal*, *51*(1), 73-112.

Hollins, E. (1982). The Marva Collins story revisited: Implications for regular classroom instruction. Journal of Teacher Education, 33(1), 37–40.

Holt, S. B., & Gershenson, S. (2015). The impact of teacher demographic representation on student attendance and suspensions. IZA working paper.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American educational research journal*, *38*(3), 499-534.

Ingersoll, R. M., & Connor, R. (2009). What the national data tell us about minority and Black teacher turnover. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA, April.

Ingersoll, R. M., & May, H. (2011). The minority teacher shortage: Fact or fable?. *Phi Delta Kappan*, *93*(1), 62-65.

Ingersoll, R., & May, H. (2016). Minority Teacher Recruitment, Employment and Retention: 1987 to 2013. *Learning Policy Institute, Stanford, CA.*

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of labor Economics*, *26*(1), 101-136.

Johnson, R. L., Penny, J., & Gordon, B. (2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. *Applied Measurement in Education*, *13*(2), 121-138.

Kinsler, J. (2011). Understanding the Black–White school discipline gap. *Economics of Education Review*, *30*(6), 1370-1383.

Koedel, C., Li, J., Springer, M. G., & Tan, L. (2017). The impact of performance ratings on job satisfaction for public school teachers. *American Educational Research Journal*, *54*(2), 241-278.

Kraiger, K., & Ford, J. K. (1985). A meta-analysis of ratee race effects in performance ratings. Journal of applied psychology, 70(1), 56.

Ladd, H. F. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, 33(2), 235-261.

LaRocque, M., Kleiman, I., & Darling, S. M. (2011). Parental involvement: The missing link in school achievement. *Preventing School Failure*, *55*(3), 115-122.

Lindsay, C. A., & Hart, C. M. (2017). Exposure to Same-Race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina. *Educational Evaluation and Policy Analysis*, 0162373717693109.

Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education*, *80*(3), 44-70.

Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? Language Testing, 19, 246–277.

Mercer, W. A., & Mercer, M. M. (1986). Standardized testing: Its impact on Blacks in Florida's education system. Urban Educator, 8(1), 105–113.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, *79*(4), 33-53.

Morris, E. W., & Perry, B. L. (2016). The punishment gap: School suspension and racial disparities in achievement. *Social Problems*, *63*(1), 68-86.

Ouazad, A. (2014). Assessed by a teacher like me: Race and teacher assessments. *Education Finance and Policy*, *9*(3), 334-372.

Poon, J. M. (2004). Effects of performance appraisal politics on job satisfaction and turnover intention. *Personnel review*, *33*(3), 322-334.

Putnam, H., Hansen, M., Walsh, K., & Quintero, D. (2016). High hopes and harsh realities: The real challenges of building a diverse workforce. Washington, DC: Brown Center on Education Policy, Brookings Institution.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The urban review*, *3*(1), 16-20.

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education*, *4*(4), 537-571.

Scafidi, B., Sjoquist, D. L., & Stinebrickner, T. R. (2007). Race, poverty, and teacher mobility. Economics of Education Review, 26, 145–159.

Simon, N., & Johnson, S. M. (2015). Teacher turnover in high-poverty schools: What we know and can do. *Teachers College Record*, 117(3), 1-36.

Stauffer, J. M., & Buckley, M. R. (2005). The existence and nature of racial bias in supervisory ratings. Journal of applied psychology, 90(3), 586.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology*, *69*(5), 797.

Stewart, J., Meier, K. J., & England, R. E. (1989). In quest of role models: Change in Black teacher representation in urban school districts, 1968-1986. *The Journal of Negro Education*, *58*(2), 140-152.

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, *102*(7), 3628-51.

The Century Foundation (2017). The Benefits of Socioeconomically and Racially Integrated Schools and Classrooms. https://tcf.org/content/facts/the-benefits-of-socioeconomically-and-racially-integrated-schools-and-classrooms/

Torres, J., Santos, J., Peck, N. L., & Cortes, L. (2004). Minority Teacher Recruitment, Development, and Retention. *Northeast and Islands Regional Educational Laboratory LAB*.

Villegas, A. M., & Irvine, J. J. (2010). Diversifying the teaching force: An examination of major arguments. *The Urban Review*, *42*(3), 175-192.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. *New Teacher Project*.

Whitehurst, G. J. R., Chingos, M. M., & Lindquist, K. M. (2014). Evaluating teachers with classroom observations. *Brown Center on Education Policy: Brookings Institution*.

Figure 1: The Composition of Tennessee's Teacher Workforce by Race and Gender

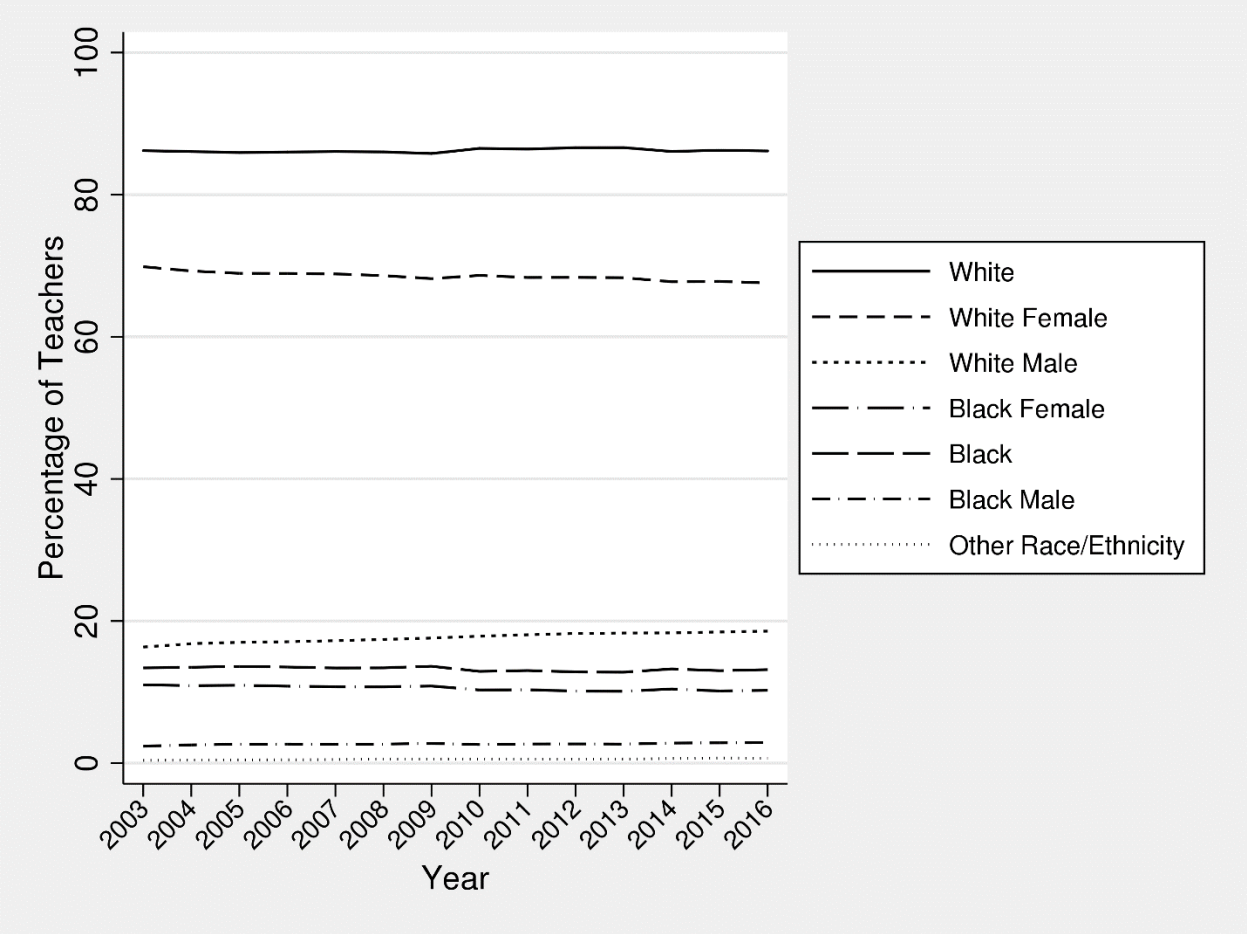


Figure 2: Teacher Turnover in Tennessee by Race and Gender

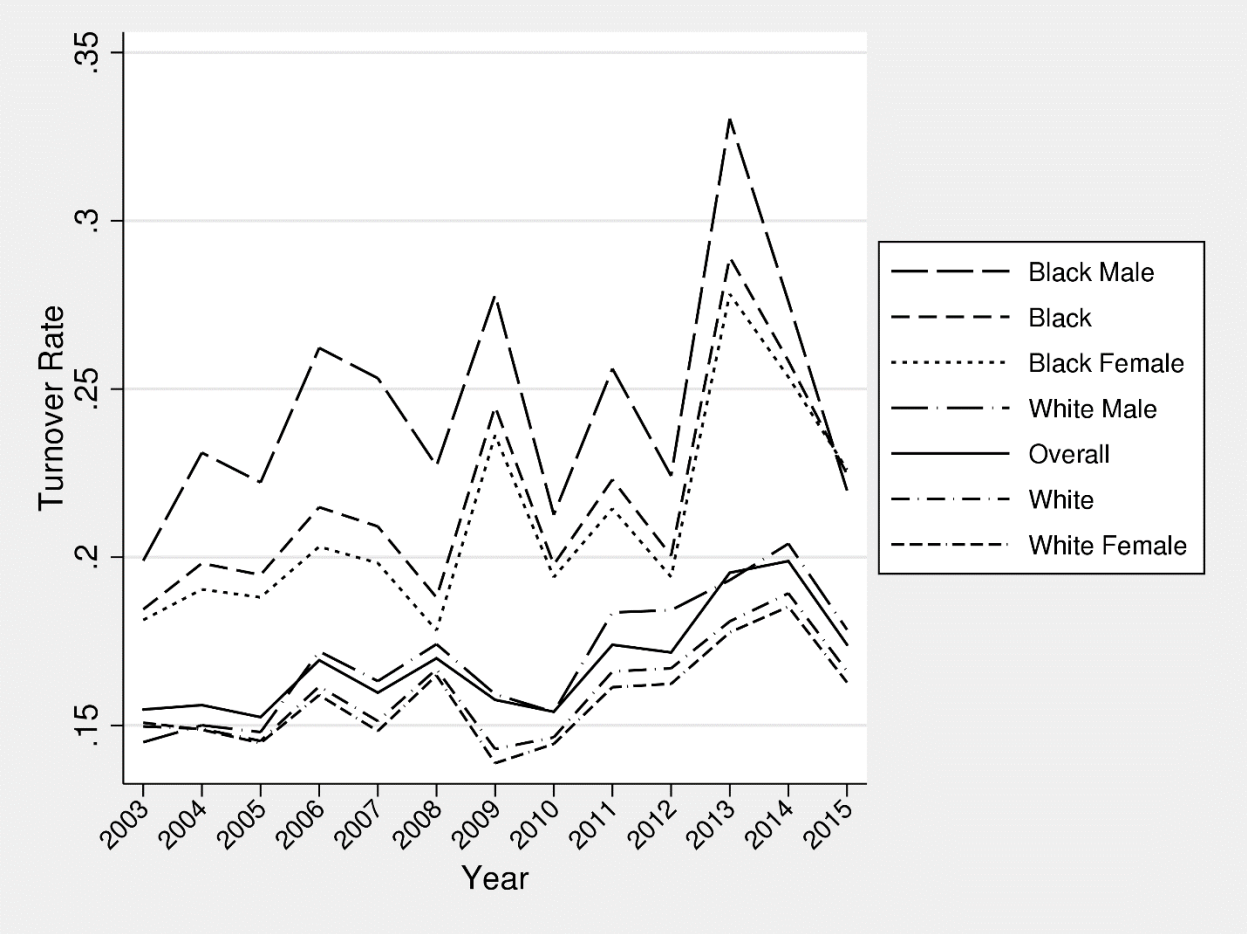


Figure 3: Types of Teacher Turnover in Tennessee by Race and Gender

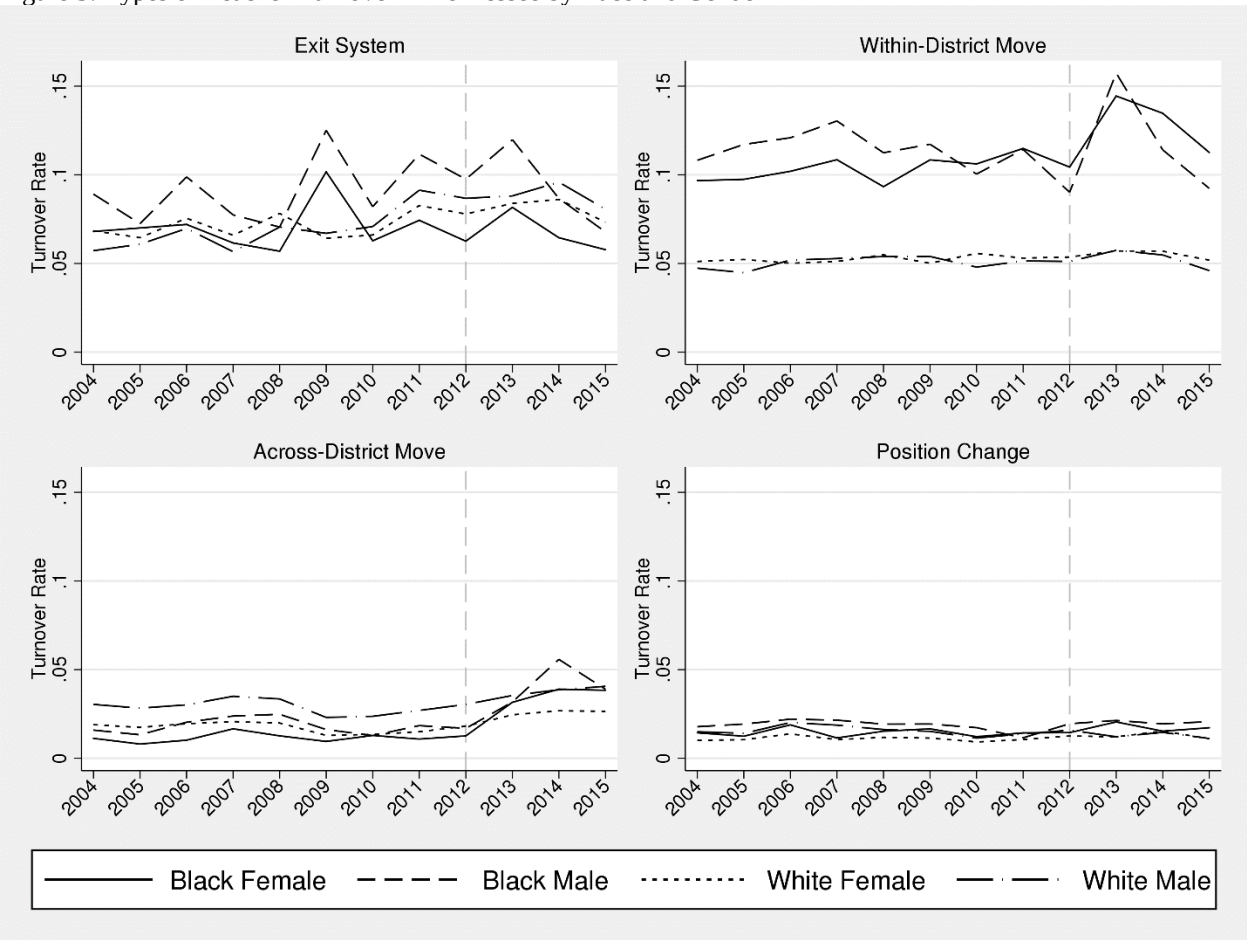
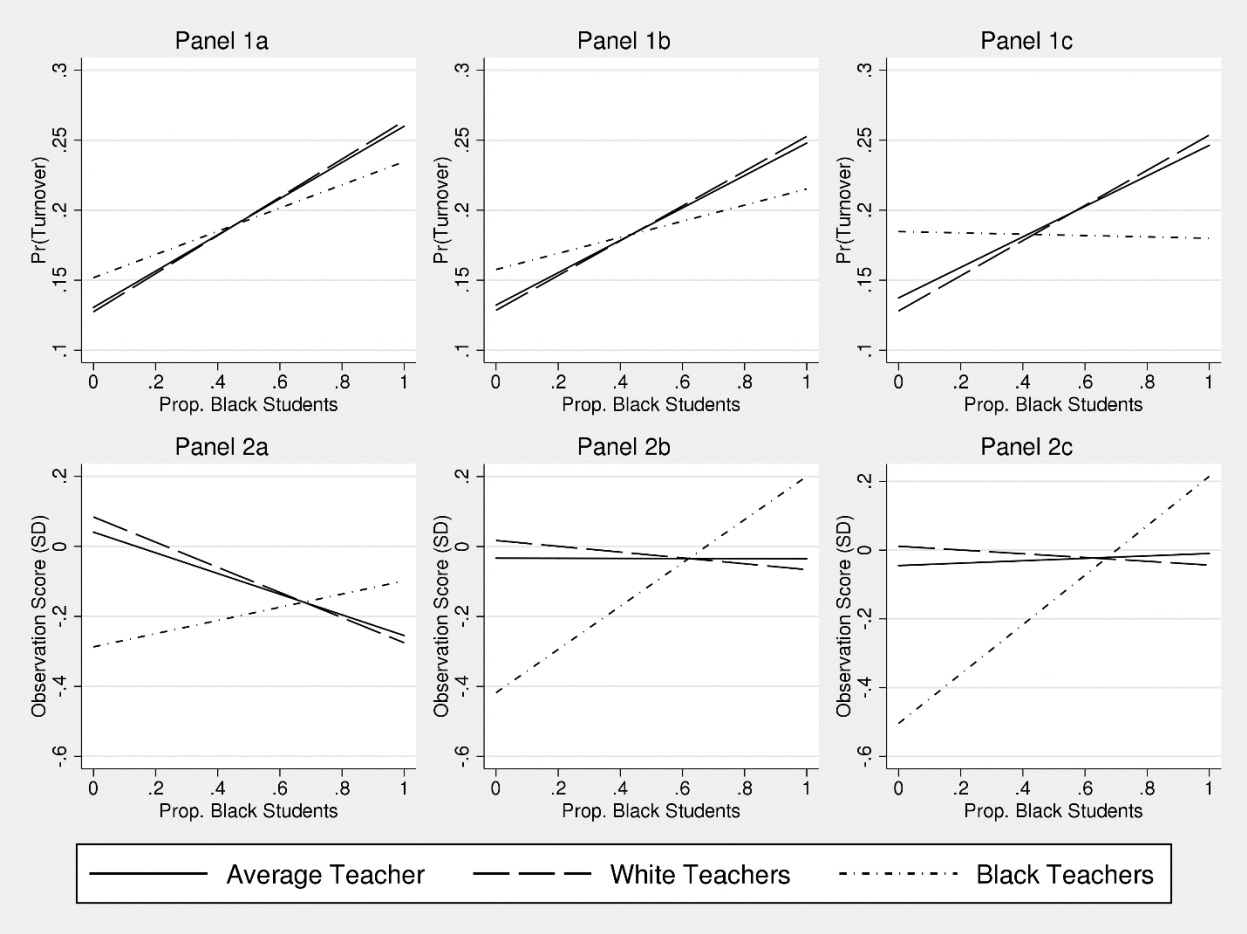
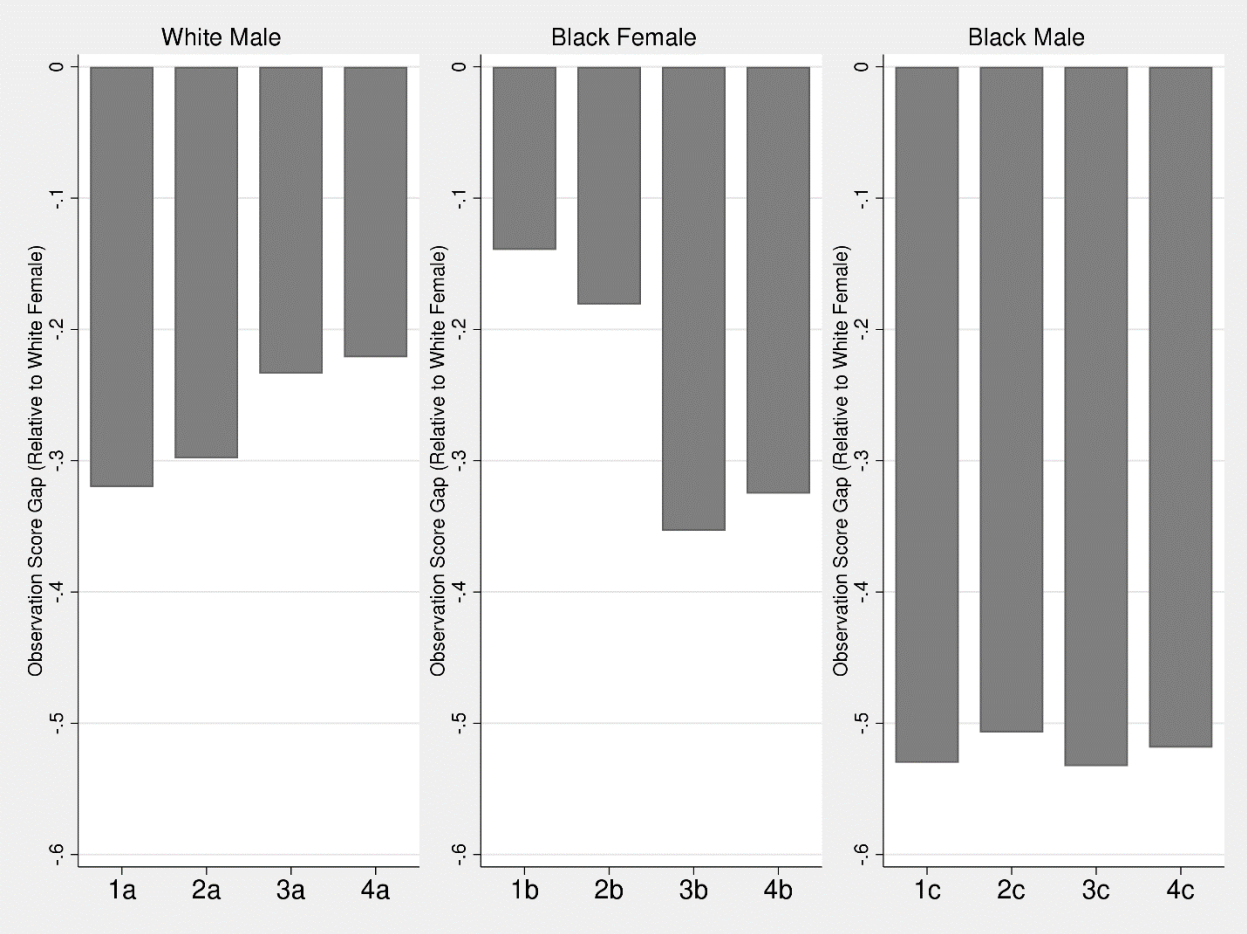


Figure 4: Differential Responsiveness of Black Teachers to School Context



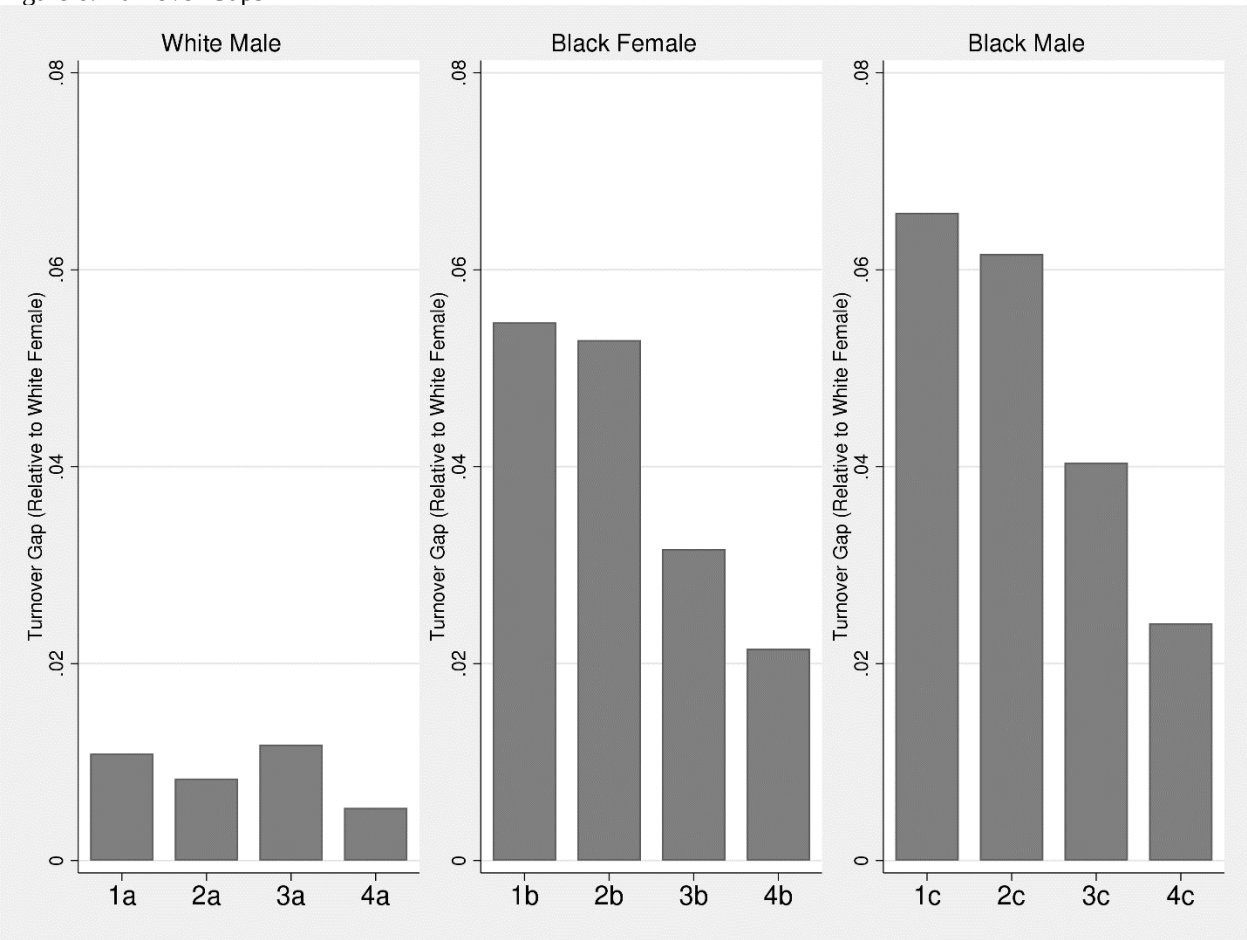
Notes: Plots are model-based predictions that include controls for year fixed effects, teacher characteristics, and the proportion of Black students in school (interacted with whether the teacher is Black). Panels 1a/2a include no other school controls, 1b/2b include the remaining set of school controls, and 1c/2c include school controls and the full set of interactions.

Figure 5: Observation Score Gaps



Notes: Each bar represents the estimated contrast between the given group and White females. Model 1 includes no covariates. Model 2 includes teacher covariates. Model 3 includes teacher covariates, school covariates, and interactions between school covariates and the race-by-gender group. Model 4 includes teacher covariates, school covariates, the aforementioned interactions, and the teacher's individual TVAAS composite score.

Figure 6: Turnover Gaps



Notes: Each bar represents the estimated contrast between the given group and White females. Model 1 includes no covariates. Model 2 includes teacher covariates. Model 3 includes teacher covariates, school covariates, and interactions between school covariates and the race-by-gender group. Model 4 includes teacher covariates, school covariates, the aforementioned interactions, and a cubic polynomial of the teacher's average observation score.

Table 2: Differences in Individual and School Contextual Characteristics by Teacher Race and Gender

	Teacher Race		Teacher Gender		Teacher Race x Gender			
	White	Black	Female	Male	White Female	White Male	Black Female	Black Male
Teacher Characteristics								
Age	42.7	43.9	42.8	43.2	42.6	43.1	44.0	43.5
Experience	12.1	12.1	12.3	11.4	12.3	11.6	12.6	10.4
Tenure	5.3	4.1	5.2	4.8	5.4	4.9	4.3	3.6
MA or above	0.56	0.67	0.58	0.53	0.57	0.53	0.71	0.53
School Characteristics								
Achievement Index (SDs)	0.23	-0.68	0.11	0.13	0.23	0.23	-0.69	-0.62
Enrollment (100s)	8.36	8.29	7.93	9.92	7.92	10.00	8.00	9.34
Proportion Black	0.18	0.67	0.24	0.25	0.18	0.19	0.67	0.66
Proportion Hispanic	0.08	0.10	0.08	0.07	0.08	0.07	0.10	0.09
Proportion Gifted	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Proportion w/ Disabilities	0.15	0.15	0.15	0.14	0.15	0.14	0.14	0.15
Proportion FRPL	0.57	0.77	0.60	0.57	0.57	0.54	0.77	0.74
Elementary	0.51	0.48	0.59	0.20	0.60	0.20	0.55	0.20
Middle	0.18	0.24	0.18	0.22	0.17	0.21	0.22	0.29
High	0.26	0.25	0.19	0.52	0.19	0.53	0.20	0.46
Other	0.04	0.03	0.04	0.06	0.04	0.06	0.03	0.05
Urban	0.26	0.75	0.32	0.32	0.26	0.27	0.76	0.73
Suburban	0.20	0.09	0.19	0.19	0.21	0.20	0.09	0.08
Town	0.18	0.05	0.16	0.16	0.18	0.18	0.05	0.06
Rural	0.35	0.10	0.32	0.33	0.35	0.36	0.10	0.12

Notes: Table includes all Black and White teachers in Tennessee from the 2011–12 through 2015–2016 school years.

Table 3: Differences in Evaluation Measures by Teacher Race and Gender

	Teacher Race		Teacher Gender		Teacher Race x Gender			
	White	Black	Female	Male	White Female	White Male	Black Female	Black Male
Level of Effectiveness (LOE)	3.84	3.63	3.85	3.66	3.88	3.69	3.68	3.41
Average Observation Score	3.89	3.80	3.92	3.73	3.93	3.74	3.85	3.62
Growth Score	3.46	3.33	3.47	3.34	3.49	3.36	3.36	3.19
Individual Growth Score	3.36	3.17	3.37	3.18	3.39	3.21	3.21	2.96
Achievement Score	4.12	3.92	4.10	4.09	4.12	4.12	3.94	3.86

Table 4: Modeling teacher turnover

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Contrasts						
Black vs. White	0.062***	0.058***	-0.010***	-0.009***	0.013***	0.025***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
Panel B: Model Coefficients						
Teacher Characteristics						
Black		0.058***	-0.010***	-0.009***	0.030***	0.089***
		(0.004)	(0.002)	(0.002)	(0.004)	(0.014)
Male		0.014***	0.009***	0.010***	0.009***	0.009***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Teacher Education						
MA Degree		0.011***	0.011***	0.010***	0.011***	0.011***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
MA+ Degree		0.024***	0.014***	0.019***	0.014***	0.015***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
EDS Degree		0.015***	0.020***	0.027***	0.021***	0.022***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PhD Degree		0.064***	0.055***	0.052***	0.054***	0.054***
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Teacher Age						
30-39		-0.008***	-0.008***	-0.007***	-0.008***	-0.008***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
40-49		-0.021***	-0.023***	-0.021***	-0.023***	-0.022***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
50-59		-0.017***	-0.024***	-0.024***	-0.024***	-0.024***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
60 and above		0.092***	0.081***	0.079***	0.081***	0.081***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Teacher Experience						
0-4 years		0.110***	0.099***	0.090***	0.099***	0.098***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
5-14 years		0.033***	0.030***	0.028***	0.030***	0.030***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
25-39 years		0.032***	0.037***	0.038***	0.036***	0.036***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
40+ years		0.095***	0.103***	0.104***	0.104***	0.105***
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
School Characteristics						
Achievement Index			-0.013***		-0.014***	-0.011***
			(0.002)		(0.002)	(0.002)
x Black						-0.021***
						(0.004)
Enrollment (100s)			-0.001***		-0.001***	-0.001***
			(0.000)		(0.000)	(0.000)
x Black						-0.002***
						(0.001)
Proportion Black			0.113***		0.127***	0.128***
			(0.005)		(0.005)	(0.005)
x Black					-0.069***	-0.132***
					(0.007)	(0.014)
Proportion Hispanic/Latino			0.047***		0.026**	0.040***
			(0.013)		(0.013)	(0.013)
x Black						-0.131***
						(0.028)

Proportion Gifted			-0.154***		-0.175***	-0.138***
			(0.027)		(0.027)	(0.029)
x Black						-0.114**
						(0.056)
Proportion w/ Disabilities			0.062***		0.060***	0.052***
			(0.020)		(0.020)	(0.020)
x Black						-0.040
						(0.042)
Proportion FRPL			-0.047***		-0.046***	-0.041***
			(0.006)		(0.006)	(0.005)
x Black						0.006
						(0.017)
School Level						
Middle			0.032***		0.032***	0.027***
			(0.002)		(0.002)	(0.002)
x Black						0.030***
						(0.006)
High			0.014***		0.014***	0.010***
			(0.003)		(0.003)	(0.003)
x Black						0.035***
						(0.008)
Other			0.016***		0.016***	0.015***
			(0.006)		(0.006)	(0.006)
x Black						0.015
						(0.016)
Urbanicity						
Urban			-0.001		-0.001	0.001
			(0.003)		(0.003)	(0.003)
x Black						-0.017**
						(0.007)
Town			-0.006**		-0.007**	-0.004
			(0.003)		(0.003)	(0.003)
x Black						-0.039***
						(0.010)
Rural			-0.000		-0.001	0.000
			(0.002)		(0.002)	(0.002)
x Black						-0.004
						(0.009)
School Fixed Effects	No	No	No	Yes	No	No
Observations	625065	596251	562729	562984	562729	562729
R^2	0.005	0.024	0.032	0.048	0.032	0.033

School-by-year clustered standard errors in parentheses. Models include year fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Modeling multinomial teacher turnover

	Transfers		Exits		Position Changes	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Contrasts						
Black vs. White	0.066***	0.018***	0.004**	0.008***	0.006***	0.004***
	(0.004)	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)
Panel B: Model Coefficients						
Teacher Characteristics						
Black		0.045***		0.063***		0.011***
		(0.012)		(0.009)		(0.004)
Male		0.009***		-0.001		0.004***
		(0.001)		(0.001)		(0.001)
Teacher Education						
MA Degree		0.009***		-0.007***		0.012***
		(0.001)		(0.001)		(0.000)
MA+ Degree		0.014***		-0.015***		0.020***
		(0.002)		(0.002)		(0.001)
EDS Degree		0.012***		-0.018***		0.035***
		(0.002)		(0.001)		(0.001)
PhD Degree		0.030***		0.000		0.044***
		(0.005)		(0.005)		(0.004)
Teacher Age						
30-39		-0.007***		-0.004**		0.001**
		(0.002)		(0.002)		(0.001)
40-49		-0.012***		-0.012***		-0.003***
		(0.002)		(0.002)		(0.001)
50-59		-0.020***		-0.001		-0.007***
		(0.002)		(0.002)		(0.001)
60 and above		-0.028***		0.125***		-0.012***
		(0.002)		(0.003)		(0.001)
Teacher Experience						
0-4 years		0.059***		0.064***		-0.007***
		(0.002)		(0.002)		(0.001)
5-14 years		0.018***		0.015***		0.001
		(0.001)		(0.001)		(0.001)
25-39 years		-0.010***		0.051***		-0.001**
		(0.001)		(0.001)		(0.001)
40+ years		-0.021***		0.131***		-0.005***
		(0.004)		(0.008)		(0.002)
School Characteristics						
Achievement Index		-0.010***		-0.003***		-0.000
		(0.001)		(0.001)		(0.000)
x Black		-0.018***		-0.011***		-0.001
		(0.004)		(0.003)		(0.001)
Enrollment (100s)		-0.000*		-0.000		-0.000***
		(0.000)		(0.000)		(0.000)
x Black		-0.002***		-0.001*		-0.000**
		(0.001)		(0.000)		(0.000)
Proportion Black		0.094***		0.056***		0.006***
		(0.004)		(0.004)		(0.001)
x Black		-0.075***		-0.088***		-0.002
		(0.011)		(0.009)		(0.004)
Proportion Hispanic/Latino		0.034***		0.009		0.000
		(0.010)		(0.009)		(0.003)
x Black		-0.073***		-0.082***		0.002

		(0.024)		(0.018)		(0.008)
Proportion Gifted		-0.122***		-0.040**		-0.010
		(0.022)		(0.019)		(0.008)
x Black		0.041		-0.132***		-0.022
		(0.047)		(0.039)		(0.018)
Proportion w/ Disabilities		0.032**		0.028*		0.013***
		(0.016)		(0.016)		(0.004)
x Black		0.064*		-0.095***		-0.021**
		(0.038)		(0.027)		(0.010)
Proportion FRPL		-0.027***		-0.019***		-0.003*
		(0.004)		(0.004)		(0.001)
x Black		0.020		-0.024**		-0.000
		(0.014)		(0.012)		(0.005)
School Level						
Middle		0.019***		0.010***		0.003***
		(0.002)		(0.001)		(0.001)
x Black		0.021***		0.011***		0.008***
		(0.006)		(0.004)		(0.002)
High		-0.002		0.009***		0.004***
		(0.002)		(0.002)		(0.001)
x Black		0.027***		0.016***		0.006***
		(0.007)		(0.005)		(0.002)
Other		0.006		0.007*		0.005**
		(0.004)		(0.004)		(0.002)
x Black		-0.007		0.024*		0.005
		(0.012)		(0.013)		(0.004)
Urbanicity						
Urban		-0.001		0.001		0.001
		(0.002)		(0.002)		(0.001)
x Black		-0.013**		-0.001		-0.007***
		(0.006)		(0.005)		(0.002)
Town		0.002		-0.010***		0.002***
		(0.003)		(0.002)		(0.001)
x Black		-0.037***		-0.007		-0.003
		(0.008)		(0.007)		(0.003)
Rural		0.002		-0.004**		0.002***
		(0.002)		(0.002)		(0.001)
x Black		-0.014**		0.009		0.001
		(0.007)		(0.006)		(0.003)
Observations	566266	514857	561407	511234	522057	479066
R^2	0.007	0.030	0.001	0.044	0.000	0.011

School-by-year clustered standard errors in parentheses. Models include year fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Modeling Observation Scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Contrasts							
Black vs. White	-0.150***	-0.182***	-0.057***	-0.165***	-0.133***	-0.340***	-0.318***
	(0.015)	(0.014)	(0.012)	(0.009)	(0.013)	(0.020)	(0.027)
Panel B: Model Coefficients							
TVAAS Composite					0.320***		0.319***
					(0.004)		(0.004)
Teacher Characteristics							
Black	-0.150***	-0.186***	-0.057***	-0.165***	-0.133***	-0.534***	-0.455***
	(0.015)	(0.014)	(0.012)	(0.009)	(0.013)	(0.099)	(0.132)
Male		-0.302***	-0.258***	-0.266***	-0.263***	-0.255***	-0.236***
		(0.008)	(0.006)	(0.005)	(0.007)	(0.006)	(0.008)
Teacher Education							
MA Degree		0.124***	0.114***	0.120***	0.102***	0.114***	0.101***
		(0.005)	(0.005)	(0.004)	(0.006)	(0.005)	(0.007)
MA+ Degree		0.225***	0.217***	0.143***	0.114***	0.213***	0.171***
		(0.011)	(0.010)	(0.008)	(0.012)	(0.010)	(0.013)
EDS Degree		0.209***	0.231***	0.183***	0.140***	0.220***	0.184***
		(0.011)	(0.011)	(0.008)	(0.011)	(0.011)	(0.014)
PhD Degree		0.335***	0.331***	0.298***	0.274***	0.337***	0.325***
		(0.024)	(0.023)	(0.021)	(0.028)	(0.023)	(0.032)
Teacher Age							
30-39		0.018**	0.013*	0.012*	0.027***	0.014*	0.021**
		(0.007)	(0.007)	(0.006)	(0.009)	(0.007)	(0.010)
40-49		-0.030***	-0.035***	-0.050***	-0.031***	-0.037***	-0.022*
		(0.008)	(0.008)	(0.007)	(0.010)	(0.008)	(0.011)
50-59		-0.153***	-0.142***	-0.156***	-0.133***	-0.139***	-0.120***
		(0.010)	(0.010)	(0.008)	(0.012)	(0.010)	(0.014)
60 and above		-0.303***	-0.293***	-0.303***	-0.256***	-0.289***	-0.252***
		(0.013)	(0.013)	(0.011)	(0.016)	(0.013)	(0.018)
Teacher Experience							
0-4 years		-0.492***	-0.457***	-0.422***	-0.356***	-0.453***	-0.403***
		(0.009)	(0.009)	(0.007)	(0.010)	(0.009)	(0.012)
5-14 years		-0.095***	-0.081***	-0.078***	-0.061***	-0.082***	-0.074***
		(0.006)	(0.006)	(0.005)	(0.008)	(0.006)	(0.009)
25-39 years		0.133***	0.129***	0.119***	0.113***	0.132***	0.115***
		(0.009)	(0.009)	(0.007)	(0.011)	(0.009)	(0.013)
40+ years		0.156***	0.160***	0.150***	0.131***	0.158***	0.132***
		(0.029)	(0.029)	(0.026)	(0.039)	(0.029)	(0.045)
School Characteristics							
Achievement Index			0.142***	0.020	-0.029	0.153***	0.102***
			(0.012)	(0.017)	(0.021)	(0.012)	(0.014)
x Black						-0.011	-0.018
						(0.023)	(0.031)
Enrollment (100s)			0.004*	0.020**	0.009	0.005**	-0.001
			(0.002)	(0.009)	(0.010)	(0.002)	(0.002)
x Black						-0.003	-0.004
						(0.004)	(0.005)
Proportion Black			0.067*	0.331	0.378	-0.048	-0.139***
			(0.039)	(0.313)	(0.369)	(0.041)	(0.045)
x Black						0.769***	0.680***
						(0.082)	(0.104)
Proportion Hispanic/Latino			-0.229**	-1.078***	-1.088***	-0.121	-0.110
			(0.094)	(0.367)	(0.411)	(0.092)	(0.112)

x Black						0.699***	0.643***
						(0.172)	(0.218)
Proportion Gifted			1.394***	0.268	0.314	1.320***	0.990***
			(0.269)	(0.710)	(0.749)	(0.273)	(0.284)
x Black						1.615***	1.851***
						(0.373)	(0.423)
Proportion w/ Disabilities			-0.097	0.231	-0.193	-0.056	-0.342*
			(0.143)	(0.280)	(0.330)	(0.150)	(0.193)
x Black						0.657**	0.687*
						(0.280)	(0.384)
Proportion FRPL			-0.146***	-0.172***	-0.198***	-0.145***	-0.150***
			(0.052)	(0.052)	(0.059)	(0.051)	(0.057)
x Black						-0.225*	-0.235
						(0.122)	(0.152)
School Level							
Middle			-0.164***			-0.147***	-0.172***
			(0.018)			(0.019)	(0.020)
x Black						-0.116***	-0.185***
						(0.035)	(0.040)
High			-0.174***			-0.174***	-0.104***
			(0.022)			(0.022)	(0.024)
x Black						-0.013	-0.035
						(0.045)	(0.054)
Other			-0.053			-0.058	-0.033
			(0.041)			(0.040)	(0.039)
x Black						0.100	-0.020
						(0.110)	(0.117)
Urbanicity							
Urban			0.033			0.019	0.016
			(0.024)			(0.024)	(0.027)
x Black						0.110***	0.175***
						(0.038)	(0.050)
Town			0.073***			0.079***	0.080***
			(0.025)			(0.024)	(0.026)
x Black						0.015	0.046
						(0.058)	(0.072)
Rural			0.014			0.021	0.033
			(0.022)			(0.022)	(0.024)
x Black						-0.048	-0.037
						(0.047)	(0.059)
School Fixed Effects	No	No	No	Yes	Yes	No	No
Observations	250240	244600	232813	232892	97681	232813	97652
R^2	0.002	0.071	0.101	0.276	0.387	0.106	0.212

School-by-year clustered standard errors in parentheses. Models include year fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Contrasts in Other Evaluation Measures

	(1)	(2)	(3)
Panel A: Growth Score			
Black vs. White	-0.090***	-0.012	-0.031
	(0.018)	(0.007)	(0.019)
	[252281]	[234979]	[234900]
Panel B: Individual Growth Score			
Black vs. White	-0.127***	-0.093***	-0.120***
	(0.013)	(0.013)	(0.021)
	[105094]	[100763]	[100734]
Panel C: Achievement Score			
Black vs. White	-0.171***	-0.033***	-0.109***
	(0.019)	(0.008)	(0.022)
	[248376]	[231327]	[231247]
Panel D: Level of Effectiveness			
Black vs. White	-0.201***	-0.067***	-0.146***
	(0.016)	(0.008)	(0.019)
	[254152]	[236180]	[236100]
Teacher Controls	No	Yes	Yes
School Controls	No	Yes	Yes
School Fixed Effects	No	Yes	No
Interactions with School Controls	No	No	Yes

Each cell shows the estimated contrast between Black and White teachers from a different regression model. Standard errors in parentheses, number of observations in brackets. In each model, standard errors are clustered at the school-by-year level. Models include year fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Estimated Turnover Gaps with and without Controls for Different Evaluation Measures

	All Turnover		Transfers		Exits		Position Changes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Observation Score								
Black vs. White	0.030***	0.020***	0.026***	0.020***	0.004	-0.004	0.007***	0.008***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)
Panel B: Growth Score								
Black vs. White	0.030***	0.030***	0.026***	0.026***	0.005	0.004	0.006***	0.006***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)
Panel C: Individual Growth Score								
Black vs. White	0.036***	0.034***	0.032***	0.031***	0.009*	0.007	0.003	0.003
	(0.008)	(0.008)	(0.007)	(0.007)	(0.005)	(0.005)	(0.002)	(0.002)
Panel D: Achievement Score								
Black vs. White	0.030***	0.029***	0.026***	0.025***	0.004	0.003	0.006***	0.006***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)
Panel E: Level of Effectiveness								
Black vs. White	0.029***	0.026***	0.025***	0.023***	0.004	0.001	0.006***	0.007***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)
Control for Evaluation Score	No	Yes	No	Yes	No	Yes	No	Yes

Each cell shows the estimated contrast between Black and White teachers from a different regression model. Standard errors in parentheses, number of observations in brackets. In each model, standard errors are clustered at the school-by-year level. Models include teacher controls, school controls, interactions between teacher race and school controls, and year fixed effects. Odd numbered columns are restricted to teachers with a non-missing evaluation measure for the corresponding row.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Estimated Satisfaction Gaps with and without Controls for Different Evaluation Measures

	(1)	(2)
Panel A: Observation Score		
Black vs. White (SD units)	0.041*	0.092***
	(0.024)	(0.023)
Panel B: Level of Effectiveness		
Black vs. White (SD units)	0.042*	0.054**
	(0.023)	(0.023)
Panel C: Growth Score		
Black vs. White (SD units)	0.039*	0.040*
	(0.024)	(0.023)
Panel D: Individual Growth Score		
Black vs. White (SD units)	0.028	0.032
	(0.035)	(0.035)
Panel E: Achievement Score		
Black vs. White (SD units)	0.040*	0.042*
	(0.024)	(0.024)
Control for Evaluation Score	No	Yes

Each cell shows the estimated contrast between Black and White teachers from a different regression model. Standard errors in parentheses, number of observations in brackets. In each model, standard errors are clustered at the school-by-year level. Models include teacher controls, school controls, interactions between teacher race and school controls, and year fixed effects. Column 1 is restricted to teachers with a non-missing evaluation measure for the corresponding row.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$