

Stated Briefly

Teacher demographics and evaluation: A descriptive study in a large urban district



Stated Briefly

Jessica Bailey

Education Development Center

Candice Bocala

WestEd

Karen Shakman

Jacqueline Zweig

Education Development Center

In collaboration with the Northeast Educator Effectiveness Research Alliance

This study examined the relationship between teacher demographic characteristics (race/ethnicity, age, and gender) and teacher summative performance evaluation ratings in a large urban district over three school years—2012/13–2014/15. In all three years a disproportionately large percentage of Black teachers, teachers age 50 and older, and male teachers were rated below proficient compared with their representation in the overall population of teachers who had a summative rating. However, from 2012/13 to 2014/15 the percentage of teachers whose ratings improved did not vary by teacher characteristics.

Why this study?

Nationwide, the prevalence of new educator evaluation systems has increased since the inception of federal initiatives such as the Race to the Top grant competition in 2010. Yet limited empirical research has examined teacher demographic characteristics and their relationship to teacher evaluation outcomes,

This brief summarizes the findings of Bailey, J., Bocala, C., Shakman, K., & Zweig, J. (2016). *Teacher characteristics and evaluation: A descriptive study in a large urban district* (REL 2017–189). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. That report is available at <http://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=4477>.

such as teacher evaluation ratings. Previous research has examined teacher characteristics and evaluation outcomes, but largely in terms of teachers' credentials, such as certification (Ballou, 1996; Clotfelter, Ladd, & Vigdor, 2010; Goldhaber & Brewer, 2000), or personality characteristics such as enthusiasm, caring, or intelligence as perceived by the principal (Harris & Sass, 2009; Harris, Ingle, & Rutledge, 2014; Master, 2014).

Using data from one urban public school district in the Regional Educational Laboratory Northeast & Islands Region, this study examined teacher summative performance evaluation ratings disaggregated by teacher demographic characteristics, including race/ethnicity, age, and gender. This information is of particular interest to officials in the district because in 2012/13 it instituted a new, more rigorous educator evaluation system (see box 1 for an overview of the system) that provoked public concern that racial/ethnic minority teachers might be more likely to receive lower ratings than non-racial/ethnic minority teachers and thus be identified for possible dismissal. This study may also be of broad interest to other districts and states as they roll out new evaluation systems and address similar issues related to the distribution of ratings across teachers with different demographic characteristics.

What the study examined

This study addressed two research questions:

1. Do the percentages of teachers with a below proficient summative performance rating vary by teacher demographic characteristics?
2. Do the percentages of teachers who improved their summative performance rating over three years vary by teacher demographic characteristics?

The data for this study were collected by the district as part of its online educator evaluation system. The dataset included demographic data for the full population of teachers—including race/ethnicity, age, and gender—and summative performance ratings for the full population of teachers eligible to receive a

Box 1. Educator evaluation in the study district

In the district's educator evaluation system, teachers are assessed using a rubric with four standards of professional practice:

- Curriculum, Planning, and Assessment (Standard I).
- Teaching All Students (Standard II).
- Family and Community Engagement (Standard III).
- Professional Culture (Standard IV).

The rubric yields a rating for each of the four standards, and the ratings are then used to generate the summative performance rating. Both the standards and summative performance ratings use four categories: exemplary, proficient, needs improvement, and unsatisfactory. In lieu of a formula for calculating a rating, evaluators use their professional judgment and minimum threshold criteria to determine the summative performance rating. The minimum threshold criteria specify that a teacher must be rated proficient or exemplary on both Standard I and Standard II to receive an overall summative rating of proficient or exemplary.

Teachers may be placed on a one- or two-year evaluation plan depending on their employment status and previous rating. Teachers with tenure and a summative performance rating of proficient or exemplary may be eligible for a two-year evaluation plan. Teachers on a one-year evaluation plan receive a summative performance rating at the end of the year, and teachers on a two-year evaluation plan receive a formative evaluation rating at the end of the first year and a summative performance rating at the end of the second year. A formative evaluation is used to determine the evaluation plan for the second year.

summative performance rating in one or more of the three school years (2012/13–2014/15). The number of teachers with a summative performance rating varied by year because some teachers entered the district, other teachers left, and many moved from a one-year evaluation plan to a two-year plan and therefore received a formative evaluation rating during the first year of the two-year plan. The total number of teachers with a summative performance rating was 3,287 for 2012/13, 2,930 for 2013/14, and 2,615 for 2014/15. For more on the sample and analyses, see Bailey, Bocala, Shakman, & Zweig (2016).

The summative performance ratings were the focus of this study, regardless of whether a teacher received them at the end of a one-year plan or a two-year plan. Because the number of teachers who were rated unsatisfactory was small for research question 1 (on teacher characteristics) the ratings are categorized as below proficient (needs improvement and unsatisfactory) and at least proficient (proficient and exemplary). The percentage of teachers in the below proficient category ranged from 6 percent to 8 percent across the three years.

For research question 2 (on teacher improvement) analyses focused on teachers who had the opportunity to improve between 2012/13 and 2014/15, meaning that they met the following criteria: taught in the district and received summative performance ratings in both years and had a rating of unsatisfactory, needs improvement, or proficient in the first year. Teachers who were rated exemplary in the first year were included in the sample only if their rating declined between 2012/13 and 2014/15. Teachers whose rating remained exemplary had no improvement to make and were therefore excluded from analyses.

To address research question 1, the study team compared the percentage of teachers within each teacher characteristic category (race/ethnicity, age, and gender) who received a below proficient summative performance rating in all three years. Specifically, the percentage of teachers in each rating category was calculated separately for each characteristic, for combinations of characteristics, and for each year using the full population of teachers with a summative performance rating for that year.

To address research question 2 the study team compared the percentage of teachers within each teacher characteristic category who had a summative performance rating and whose rating improved between 2012/13 and 2014/15. The full report includes comparisons over two one-year periods (2012/13–2013/14 and 2013/14–2014/15) as well as over the three-year period; see Bailey et. al. (2016) for the full set of findings.

The demographic composition of the population of teachers with summative performance ratings provides the overall context for the study. Across all three years, approximately 60 percent of teachers with summative performance ratings were White, 55 percent were ages 30–49, and 75 percent were female, with some variation over time (see Bailey et al., 2016, for details for each year).

What the study found

This section summarizes several key findings from the study (see Bailey et al., 2016, for the full set of findings). The first four focus on the teachers who received a below proficient summative performance rating in each year by teacher characteristic (research question 1). The fifth focuses on improvement in ratings, by teacher characteristic, for teachers who received a summative rating in 2012/13 and 2014/15 and had the opportunity to improve during the time period (research question 2).

In all three years a disproportionately large percentage of Black teachers, teachers age 50 and older, and male teachers were rated below proficient compared with their representation in the population of teachers with a summative performance rating

Black teachers accounted for 22–23 percent of teachers with a summative performance rating in each year but 35–43 percent of teachers rated below proficient each year (table 1; see also appendix B in Bailey

Table 1. Demographic characteristics of teachers with a below proficient summative performance rating, 2012/13–2014/15

Characteristic	2012/13	2013/14	2014/15
	Percent	Percent	Percent
Race/ethnicity^a			
Black	42.6	34.7	36.6
White	40.6	46.8	41.6
Other	16.9	18.5	21.7
Age			
Younger than 30	12.4	13.4	14.3
30–49	38.2	41.2	43.5
50 and older	49.4	45.4	42.2
Gender			
Male	37.3	36.6	36.0
Female	62.7	63.4	64.0

Note: N = 249 for 2012/13, 216 for 2013/14, and 161 for 2014/15.

a. Black includes African American and other includes American Indian or Alaska Native, Asian, Hispanic or Latino, and Native Hawaiian or other Pacific Islander.

Source: Authors' analysis based on district data for 2012/13–2014/15.

et al., 2016). Teachers age 50 and older accounted for 23–27 percent of teachers with a summative performance rating in each year but 42–49 percent of teachers rated below proficient each year. And male teachers accounted for 25–27 percent of teachers with a summative performance rating in each year but 36–37 percent of teachers rated below proficient each year.

In all three years the percentage of teachers with a summative performance rating who were rated below proficient was higher among Black teachers than among White teachers, although the gap was smaller in 2013/14 and 2014/15

In 2012/13, 15 percent of Black teachers with a summative rating were rated below proficient, compared with 5 percent of White teachers and 8 percent of other racial/ethnic minority teachers (figure 1). In 2013/14 and 2014/15 the difference between groups was not as pronounced. Between 2012/13 and 2014/15 the gap between Black teachers and White teachers had narrowed from 10 percentage points to 6. The number of teachers receiving summative performance ratings in each year was different and decreased over time across all race/ethnicity categories (see appendix B of Bailey et al., 2016, for details). Therefore, reported percentages of teachers by characteristic might differ due to the changing population of teachers examined.

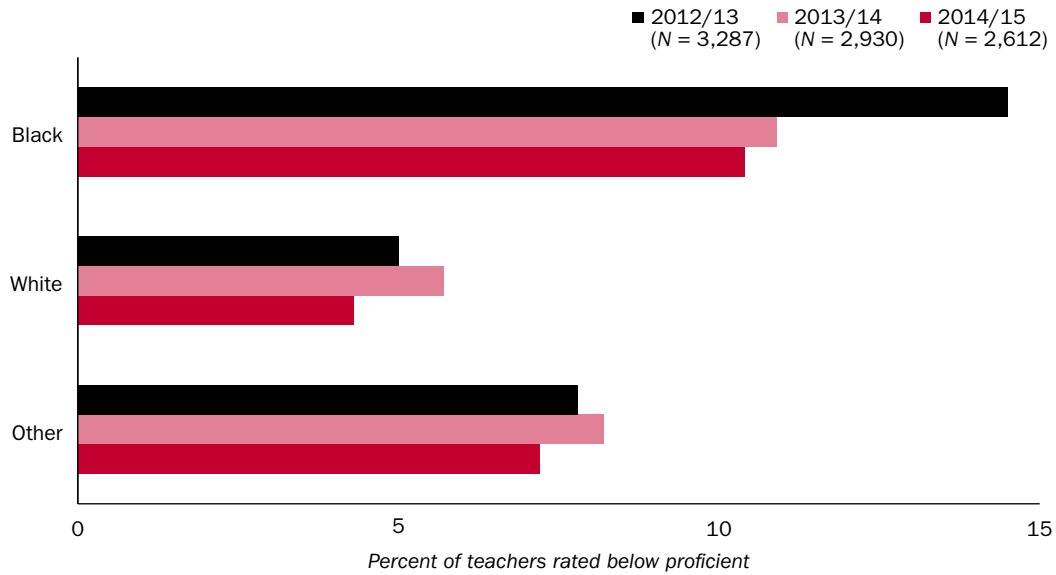
In all three years the percentage of teachers with a summative performance rating who were rated below proficient was higher among teachers age 50 and older than among teachers younger than 50

Between 2012/13 and 2014/15 the percentage of teachers age 50 and older receiving a below proficient rating ranged from 11 percent to 14 percent, compared with 4 percent to 5 percent for teachers younger than age 30 and 5 percent to 6 percent for teachers age 30–49 (figure 2). The gap between teachers age 50 and older and their younger counterparts ranged 9 percent in 2012/13 to 6 percent in 2014/15.

In all three years the difference in the percentages of male and female teachers with a summative performance rating who were rated below proficient was approximately 5 percentage points or less

A higher percentage of male teachers than of female teachers with a summative performance rating received a below proficient rating in each year, though the difference was approximately 5 percentage points or less. In 2012/13 the percentage was 11 percent among male teachers, compared with 6 percent among female

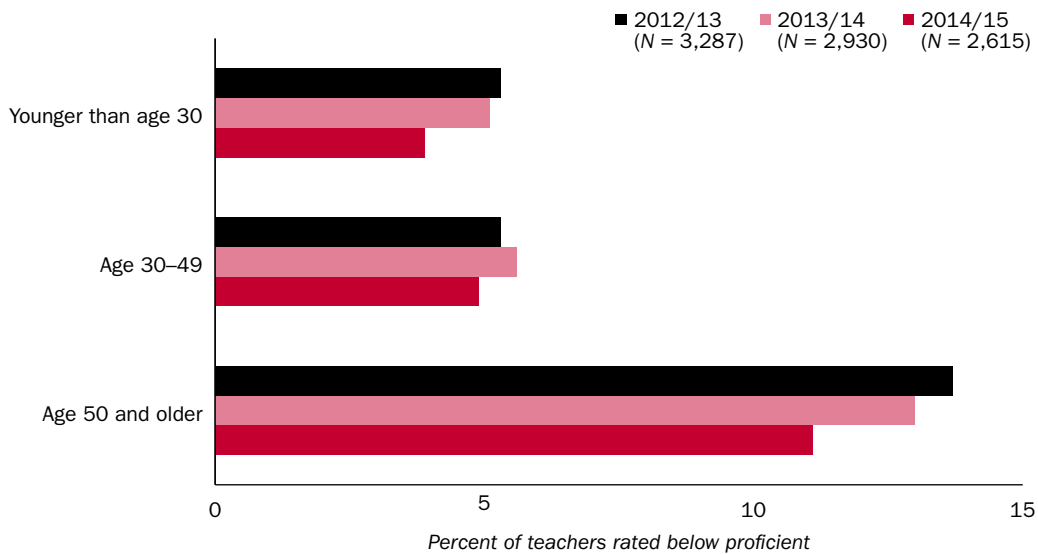
Figure 1. In 2012/13–2014/15 the percentage of teachers with a summative performance rating who were rated below proficient was higher among Black teachers than among White teachers



Note: Black includes African American, and other includes American Indian or Alaska Native, Asian, Hispanic or Latino, and Native Hawaiian or other Pacific Islander.

Source: Authors' analysis based on district data for 2012/13–2014/15.

Figure 2. In 2012/13–2014/15 the percentage of teachers with a summative performance rating who were rated below proficient was higher among teachers age 50 and older than among teachers younger than 50



Source: Authors' analysis based on district data for 2012/13–2014/15.

teachers; in 2013/14 it was 10 percent among male teachers, compared with 6 percent among female teachers; and in 2014/15 it was 8 percent among male teachers, compared with 5 percent among female teachers.

The percentage of teachers who improved their summative performance rating between 2012/13 and 2014/15 did not vary by race/ethnicity, age, or gender

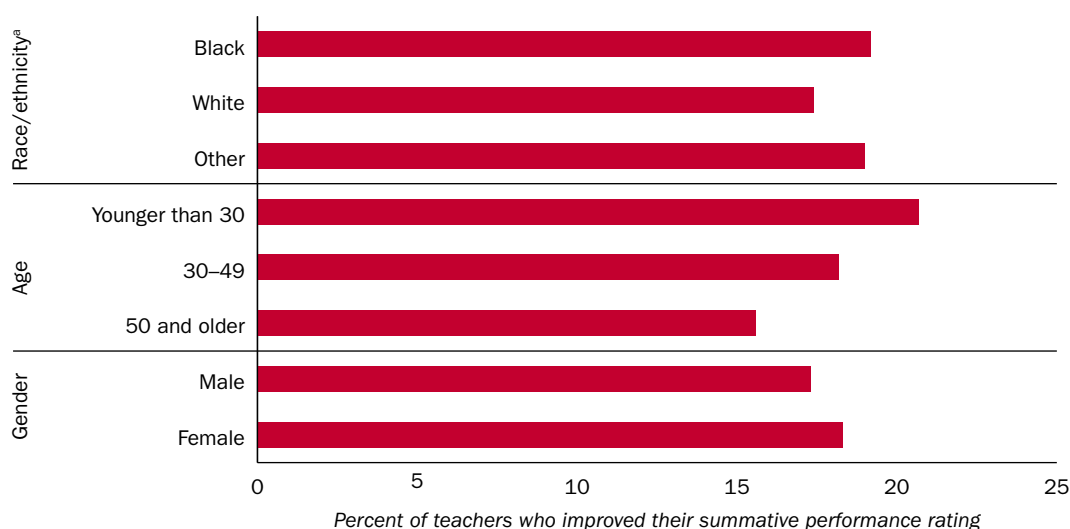
The difference in the percentage of teachers who improved their summative performance rating from 2012/13 to 2014/15 was 5 percentage points or less across the age categories (figure 3). The difference was less than 2 percentage points between Black and other racial/ethnic minority teachers, between Black and White teachers, and between female teachers and male teachers. (See Bailey et al., 2016, for details on the changes over a one-year period.)

Implications of the study findings

Examining the data over three years revealed that a disproportionately large percentage of Black teachers, older teachers, and male teachers were rated below proficient compared with their representation in the population of teachers with summative performance ratings. While the percentage of Black teachers and older teachers who were rated below proficient decreased over time in some cases, gaps between their ratings and the ratings of their White and younger counterparts persisted. Thus, the district may want to consider what programs or policies aimed specifically at these teachers and their evaluators may increase their chances for improvement and reduce the gaps. Additional research may help uncover why these patterns persist. For example, additional research might examine:

- Other contextual factors that may contribute to the teachers’ ratings, such as the characteristics of schools in which they teach, evaluator demographic characteristics (such as race/ethnicity, age, and gender), and demographic match between teachers and evaluators.

Figure 3. The difference in the percentage of teachers who improved their summative performance rating from 2012/13 to 2014/15 was 5 percentage points or less across age categories, 2 percentage points between Black teachers and other racial/ethnic minority teachers, and 1 percentage point between female teachers and male teachers



Note: N = 280.

a. Black includes African American, and other includes American Indian or Alaska Native, Asian, Hispanic or Latino, and Native Hawaiian or other Pacific Islander.

Source: Authors’ analysis based on district data for 2012/13–2014/15.

- Patterns of mobility and retention of teachers, by characteristic, over time. The changing teaching population, specifically changes within the groups of teachers who receive a summative evaluation each year, may have contributed to the patterns observed in this study, which may warrant further examination.
- The extent to which support targeted to specific subgroups of teachers (such as a district's effort to provide support to male racial/ethnic minority teachers) may help improve teacher evaluation outcomes.

For the district that provided the data, this study is part of a larger effort to create a human capital system that identifies teachers' needs, provides teachers with targeted professional development and support, monitors their progress, and ultimately achieves the larger objective of improving teaching and learning across the district. Understanding patterns in the distribution of teachers' ratings helps ensure that the system is meeting its human capital goals. For example, the district may use this descriptive information to provide targeted professional development and feedback to teachers who are most in need of improvement as well as to hypothesize what factors may influence the patterns of ratings and improvement identified in the study. This information may also be useful to other districts interested in examining patterns in evaluation ratings and teacher characteristics.

Limitations of the study

The primary limitation of this study is that although the data are drawn from three years, the findings are based on different populations of teachers each year. For example, each year some teachers left the district, and others entered. And teachers might not have received a summative performance rating for one year for various reasons, including being placed on a two-year evaluation cycle, which provides summative performance ratings only every other year.

To address this concern, missing-data analyses were conducted for each year to examine the differences between teachers who had a rating and teachers who did not and between teachers for whom improvement over time could be calculated and teachers who were missing a rating in a subsequent year and were thus excluded from the improvement analyses. For each of the three years the missing-data analyses revealed statistically significant differences between teachers with summative evaluation ratings and the larger population of teachers in the district, by race/ethnicity and age. Therefore, some of the findings may be influenced by the fact that the number of teachers in the analysis changed each year.

Additional missing-data analyses indicated statistically significant differences by race/ethnicity and age between teachers who received a summative rating in both years and teachers who were missing a rating in the second year. This suggests that reported percentages of teachers who improved their rating examined in research question 2 may be due partially to missing data in the population of teachers examined. (See appendix A in Bailey et al., 2016, for a discussion of the missing-data analyses.)

References

- Bailey, J., Bocala, C., Shakman, K., & Zweig, J. (2016). *Teacher demographics and evaluation: A descriptive study in a large urban district* (REL 2017–189). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. <http://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=4477>.
- Ballou, D. (1996). Do public schools hire the best applicants? *The Quarterly Journal of Economics*, 111(1), 97–133.
- Clotfelter, C., Ladd, H., & Vigdor, J. (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. <http://eric.ed.gov/?q=EJ889247>
- Goldhaber, D., & Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129–145. Retrieved May 12, 2016, from <http://epa.sagepub.com/content/22/2/129.abstract>.
- Harris, D., & Sass, T. (2009). *What makes for a good teacher and who can tell?* Washington, DC: CALDER Urban Institute National Center for Analysis of Longitudinal Data in Education Research. Retrieved August 28, 2015, from <http://www.urban.org/sites/default/files/alfresco/publication-pdfs/1001431-What-Makes-for-a-Good-Teacher-and-Who-Can-Tell-.PDF>.
- Harris, D., Ingle, W., & Rutledge, S. (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. <http://eric.ed.gov/?q=EJ1018910>
- Master, B. (2014). Staffing for success: Linking teacher evaluation and school personnel management in practice. *Educational Evaluation and Policy Analysis*, 36(2), 207–227. <http://eric.ed.gov/?q=EJ1024230>

REL 2017–215

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

January 2017

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0009 by Regional Educational Laboratory Northeast & Islands, administered by Education Development Center, Inc. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Bailey, J., Bocala, C., Shakman, K., & Zweig, J. (2017). *Stated Briefly: Teacher characteristics and educator evaluation: A descriptive study in a large urban district* (REL 2017–215). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

This report is available on the Regional Educational Laboratory website at <http://ies.ed.gov/ncee/edlabs>.

The Regional Educational Laboratory Program produces 7 types of reports



Making Connections

Studies of correlational relationships



Making an Impact

Studies of cause and effect



What's Happening

Descriptions of policies, programs, implementation status, or data trends



What's Known

Summaries of previous research



Stated Briefly

Summaries of research findings for specific audiences



Applied Research Methods

Research methods for educational settings



Tools

Help for planning, gathering, analyzing, or reporting data or research