



What's Happening

October 2016

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

Jessica Bailey

Education Development Center, Inc.

Candice Bocala

WestEd

Karen Shakman

Jacqueline Zweig

Education Development Center, Inc.



ies NATIONAL CENTER FOR
EDUCATION EVALUATION
AND REGIONAL ASSISTANCE

Institute of Education Sciences
U.S. Department of Education

REL
NORTHEAST
& ISLANDS

Regional Educational Laboratory
At Education Development Center, Inc.

U.S. Department of Education

John B. King, Jr., *Secretary*

Institute of Education Sciences

Ruth Neild, *Deputy Director for Policy and Research*
Delegated Duties of the Director

National Center for Education Evaluation and Regional Assistance

Joy Lesnick, *Acting Commissioner*
Amy Johnson, *Action Editor*
Elizabeth Eisner, *Project Officer*

REL 2017–189

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.

October 2016

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. (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. 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>.

Summary

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. Yet limited empirical research examines teacher demographic characteristics and their relationship to teacher evaluation outcomes, 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, Ingle, & Rutledge, 2014; Harris & Sass, 2009; Master, 2014). No recently published studies examine teachers' demographic characteristics as they relate to educator evaluation outcomes.

Using data from one urban public school district in the Regional Educational Laboratory Northeast & Islands Region that implemented a new educator evaluation system in 2012/13, this study examines teacher summative performance evaluation ratings—ratings that teachers on a one-year, summative evaluation plan receive annually, which have four categories: exemplary, proficient, needs improvement, and unsatisfactory—disaggregated by teacher characteristics, including race/ethnicity, age, and gender. Some public concern had been expressed in this district—and in others across the country—that racial/ethnic minority teachers may be more likely than other teachers to be identified for possible dismissal based on a lower performance rating. As a result the district wanted to examine more closely any patterns in performance ratings over time by teacher demographic characteristics. As other districts and states roll out new evaluation systems, they may face similar concerns related to the distribution of ratings across teachers with different demographic characteristics.

This study addressed two topics: whether the percentage of teachers with a below proficient summative performance rating varies by teacher characteristics and whether the percentage of teachers who improved their summative performance rating over three years varies by teacher characteristics. Tests of statistical significance were not conducted because the study examined a population of teachers. Differences greater than 5 percentage points were deemed to be substantively meaningful and are highlighted in this report.

Key findings based on an analysis of teacher ratings from three years (2012/13–2014/15) include:

- The characteristics of teachers in each of the three years included a disproportionately large percentage of Black teachers, teachers age 50 and older, and male teachers rated below proficient compared with the percentage of those demographics in the population of teachers with a summative performance rating.
- 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 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 age 50.
- In all three years the difference in the percentage of male and female teachers with a summative performance rating who were rated below proficient was approximately 5 percentage points or less.

- The percentage of teachers who improved their rating during all three year-to-year comparisons did not vary by race/ethnicity, age, or gender.

The findings provide an overview of teachers' summative performance ratings and their improvement in ratings over time, by teacher characteristics.¹ This report describes patterns in summative performance ratings over three years but does not explain why the patterns exist or to what they may be attributed. Instead, the findings suggest the need for further research on the potential causes of the gaps identified, as well as strategies for ameliorating them.

Contents

Summary	i
Why this study?	1
What the study examined	2
What the study found	4
Teacher demographics and ratings	5
Teacher characteristics and improvement in ratings	9
Implications of the study findings	11
Limitations of the study	12
Appendix A. Study data and methodology	A-1
Appendix B. Supplementary tables	B-1
Notes	Notes-1
References	Ref-1
Boxes	
1 Key terms	2
2 Data and methods	3
Figures	
1 In 2012/13–2014/15, approximately 60 percent of teachers in the study district who had a summative performance rating were White, 55 percent were age 30–49, and 75 percent were female	5
2 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	6
3 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	7
4 In 2012/13 and 2013/14 the percentage of teachers age 50 and older with a summative performance rating who were rated below proficient was higher among Black teachers than among White teachers	8
5 In 2012/13–2014/15 the difference in the percentage of male and female teachers with a summative performance rating who were rated below proficient was less than 5 percentage points	8
6 In all three year-to-year comparisons the percentages of Black, White, and other racial/ethnic minority teachers who improved their summative performance rating were within 5 percentage points of each other	10

Tables

A1	Data elements for research questions	A-1
A2	Total number of teachers and number of teachers with a summative performance rating, 2012/13–2014/15	A-2
A3	Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating, 2012/13	A-3
A4	Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating, 2013/14	A-4
A5	Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating, 2014/15	A-4
A6	Overview of missing data, 2012/13–2014/15	A-5
A7	Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating in both 2012/13 and 2013/14	A-6
A8	Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating in both 2013/14 and 2014/15	A-6
A9	Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating in both 2012/13 and 2014/15	A-7
A10	Research question 2 example variable values	A-8
B1	Summative performance ratings, 2012/13–2014/15	B-1
B2	Summative performance ratings using all four rating categories, 2012/13–2014/15	B-1
B3	Demographic characteristics of all teachers with a summative performance rating, 2012/13–2014/15	B-2
B4	Demographic characteristics of teachers with a below proficient summative performance rating, 2012/13–2014/15	B-2
B5	Summative performance ratings, by teacher characteristics, 2012/13	B-3
B6	Summative performance ratings, by teacher characteristics, 2013/14	B-3
B7	Summative performance ratings, by teacher characteristics, 2014/15	B-4
B8	Percentage of teachers with a below proficient summative performance rating, by race/ethnicity and age, 2012/13–2014/15	B-4
B9	Number and percentage of teachers with a summative performance evaluation rating who left the district, by age and race/ethnicity, 2012/13–2014/15	B-5
B10	Percentage of teachers receiving a below proficient summative performance rating, by age and gender, 2012/13–2014/15	B-5
B11	Percentage of teachers receiving a below proficient summative performance rating, by gender and race/ethnicity, 2012/13–2014/15	B-6
B12	Percentage of teachers whose summative performance ratings improved, by teacher characteristics, 2012/13–2014/15	B-6
B13	Percentage of teachers whose summative performance ratings improved, by race/ethnicity and age, 2012/13–2014/15	B-7
B14	Percentage of teachers whose summative performance ratings improved, by age and gender, 2012/13–2014/15	B-7
B15	Percentage of teachers whose summative performance ratings improved, by gender and race/ethnicity, 2012/13–2014/15	B-8
B16	Characteristics of teachers on one-year and two-year evaluation plans, 2012/13	B-8
B17	Characteristics of teachers on one-year and two-year evaluation plans, 2013/14	B-9
B18	Characteristics of teachers on one-year and two-year evaluation plans, 2014/15	B-9

Why this study?

Since the inception of federal initiatives such as the Race to the Top grant competition, as well as state waivers from the federal Elementary and Secondary Education Act (No Child Left Behind Act of 2001), educator evaluation systems have rapidly transformed nationwide. Although research has assessed various aspects of the evaluation systems, including the reliability of specific measures used in the new systems (Ho & Kane, 2013; Sartain, Stoelinga, & Brown, 2011) and implementation in some states and districts (Curtis, 2012; Riordan, Lacireno-Paquet, Shakman, Bocala, & Chang, 2015), limited empirical work examines teacher characteristics and their relationship to teacher evaluation outcomes in the new evaluation systems. To the extent that research has investigated the relationship between teacher characteristics and aspects of educator evaluation, such as teachers' value added to student achievement or principal tenure decisions, teacher characteristics have been defined largely as either teacher credentials—such as certification and licensure, teacher university preparation, or graduate degree (Ballou, 1996; Clotfelter, Ladd, & Vigdor, 2010; Goldhaber & Brewer, 2000)—or personality characteristics—such as enthusiasm, caring, or intelligence, as perceived by the principal (Harris, Ingle, & Rutledge, 2014; Harris & Sass, 2009; Master, 2014).

This study was conducted to address the lack of empirical work examining teacher characteristics and their relationship to teacher evaluation outcomes

This study was conducted to address that gap in the literature and in response to a request by the Regional Educational Laboratory (REL) Northeast & Islands Northeast Educator Effectiveness Research Alliance, an alliance of state and district leaders focused on educator effectiveness and the use of new educator evaluation systems to promote improvements in teaching and learning. One of its goals is to use research to support states' and districts' educator evaluation systems and to build their capacity to evaluate their own systems.

This report provides findings about the improvement of teacher evaluation ratings over time by teacher demographic characteristics, which include teachers' race/ethnicity, age, and gender, in a large urban school district in the REL Northeast & Islands Region. This information is of particular interest to officials in the district because in 2012/13 it implemented a new, more rigorous evaluation system with a new rubric for assessing teacher practice. In addition, some people in the district had expressed concern that racial/ethnic minority teachers may be more likely than other teachers to be identified for possible dismissal as a result of their evaluation rating. Therefore, the district was interested in better understanding patterns in evaluation outcomes by teacher characteristics, in particular for teachers rated below proficient.

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. The passage of the Every Student Succeeds Act in December 2015 highlights equity as it relates to the teacher workforce and requires states to report data on differences in teachers' professional qualifications between high- and low-poverty schools. These data include the number and percentage of teachers who are inexperienced, hold emergency or provisional certification, and teach outside their field. States will also need to describe the measures they will use to ensure that low-income and racial/ethnic minority students are not served at disproportionate rates by ineffective, out-of-field, or inexperienced teachers.

Although this study reports on patterns in evaluation ratings by teacher demographic characteristics, it does not provide explanations for the patterns, such as whether a particular training or support initiative led to changes in summative ratings. It also does not make claims regarding the presence of bias or inequity within the evaluation system or by the evaluators. The study does not examine evaluator characteristics or any patterns between ratings and the evaluators. The focus is explicitly on teacher characteristics and teacher ratings. The study provides information about what patterns exist—by teacher characteristics—that may lead to future research to explore the patterns in greater depth.

What the study examined

The study district introduced a new educator evaluation system districtwide in 2012/13 that includes a professional practice rubric with four standards to be used in the evaluation:

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

Evaluators use the rubric to assign standard-level and summative performance ratings to the teacher. The rubric yields a rating for each of the four standards, and the ratings are then used to generate the final summative performance rating (see box 1 for definitions of key terms). Both the standard and summative performance ratings use four categories: exemplary, proficient, needs improvement, and unsatisfactory. In lieu of a formula for calculating a final rating, evaluators use their professional judgment and minimum threshold

Although this study reports on patterns in evaluation ratings by teacher demographic characteristics, it does not provide explanations for the patterns, such as whether a particular training or support initiative led to changes in summative ratings

Box 1. Key terms

Age. Age is reported in three categories: younger than 30, 30–49, and 50 and older.

At least proficient. Summative performance ratings of either proficient or exemplary.

Below proficient. Summative performance ratings of either needs improvement or unsatisfactory.

Formative evaluation ratings. Ratings that teachers receive at the conclusion of the first year of a two-year evaluation plan and that are used to determine the evaluation plan for the second year.

Gender. Two categories were used: male and female.

One-year evaluation plan. An evaluation plan that results in an annual summative performance rating.

Race/ethnicity. Three categories of race/ethnicity were used: White, Black (includes African American), and other (includes American Indian or Alaska Native, Asian, Hispanic or Latino, and Native Hawaiian or other Pacific Islander). These categories were used because they reflected the largest groups in the district. The district data did not report a category for two or more races/ethnicities.

Summative performance ratings. Ratings that teachers receive annually, which have four categories: exemplary, proficient, needs improvement, and unsatisfactory.

Two-year evaluation plan. An evaluation plan that results in a formative evaluation rating in the first year and a summative performance rating in the second year; this plan is available only to teachers who were rated at least proficient in the previous year.

criteria to determine the summative performance rating. Minimum threshold criteria specify that a teacher must be rated exemplary or proficient on both Standard I and Standard II to receive an overall summative rating of exemplary or proficient. Teachers may be placed on a one- or two-year evaluation plan depending on their employment status and previous rating.² 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.

The summative performance ratings were the focus of this study, regardless of whether a teacher received them at the end of a one-year or a two-year evaluation plan. Because the number of teachers who were rated unsatisfactory was small, the ratings are categorized in this report as below proficient (needs improvement and unsatisfactory) and at least proficient (proficient and exemplary).

The summative performance ratings were the focus of this study and are categorized in this report as below proficient (needs improvement and unsatisfactory) and at least proficient (proficient and exemplary)

Using data from three school years, 2012/13–2014/15, the study addressed two research questions:

1. Do the percentages of teachers with a below proficient summative performance rating vary by teacher demographic characteristics (race/ethnicity, age, and gender)?
2. Do the percentages of teachers who improved their summative performance rating over three years vary by teacher demographic characteristics?

See box 2 for a summary of the data and methods and appendix A for more details.

Box 2. Data and methods

Data

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 summative performance rating in one or more of the three years. 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 2014/15.¹ The decrease in the number of teachers with a summative performance rating over time generally reflects an increase in the number of teachers placed on two-year evaluation plans and who therefore did not receive a summative performance rating in the first year of the two-year evaluation plan. Longitudinal data for teachers with a summative performance rating for at least two years (for example, for 2013/14 and 2014/15) were used for the year-to-year comparisons (for research question 2). The total numbers varied for those comparisons, ranging from 1,198 to 1,697 (see table A6 in appendix A).

(continued)

Box 2. Data and methods *(continued)*

Methods

Descriptive analyses of frequencies were conducted to examine the characteristics, summative performance ratings, and improvement on ratings over time of teachers in the district. Specifically, crosstabulations were used in two ways: in most instances the calculations examined differences in percentages of teachers with a below proficient summative performance rating using a single or combination of teacher characteristics as the denominator; in other instances the calculations examined differences among the teacher characteristics using summative performance rating as the denominator. In all instances the calculations were conducted separately for each year using the full population of teachers with a summative performance rating for that year. The percentage of teachers improving their summative performance rating from one year to the next was calculated separately for each characteristic and for combinations of characteristics using the same teachers for each year-to-year comparison. These samples include only teachers who had an opportunity to improve their rating, either from one year to the next or over two years (if they were on a two-year evaluation plan starting in 2012/13). Thus, teachers who did not have a summative rating because they left (or entered) between the first and third years of the study or who had only a formative evaluation as a result of being placed on a two-year evaluation plan are not included in the analyses relating to improvement over time. Appendix A includes a detailed description of the methods, including results of missing-data analyses, for research questions 1 and 2.

Ratings for each of the four standards on the professional practice rubric also were examined, but no consistent patterns were apparent by teacher characteristics or that reflected new findings not revealed by the summative performance ratings analyses. In addition, although school level taught (elementary, middle, or high school) was initially considered as a characteristic to examine, the district variable for level taught was coded in such a way that interpretation of the analyses would have been difficult (for example, K–8 schools are neither clearly elementary nor clearly middle schools). Thus, analyses by school level are not included.

Note

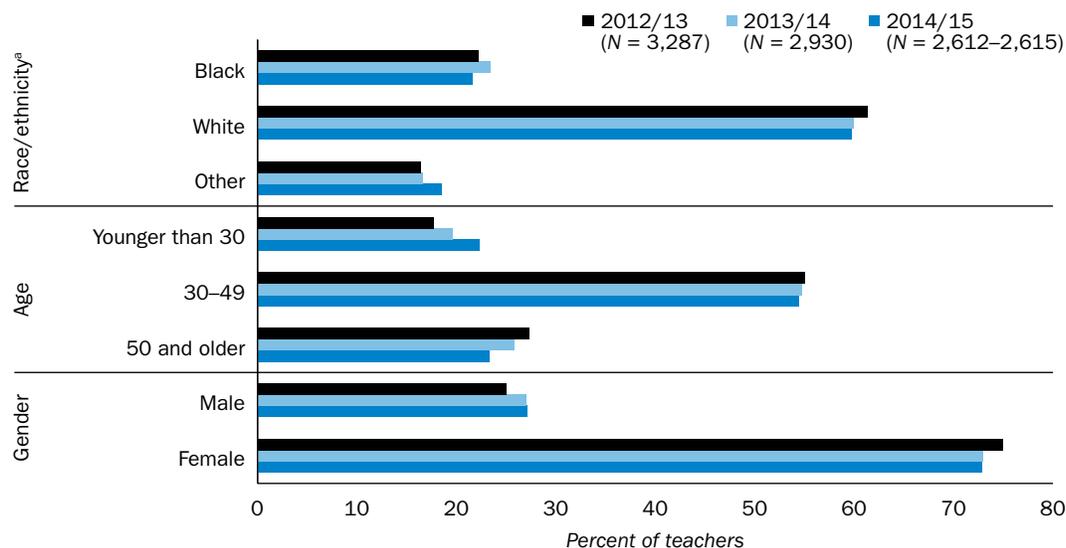
1. The total number of teachers in the district is closer to 4,600, but not all of them had a summative evaluation rating. See appendix A for further discussion and results of missing-data analyses.

The demographic composition of the teacher population with summative performance ratings provides the overall context for this study and its findings (figure 1; see also table B3 in appendix B). Across all three years, approximately 60 percent of teachers who had a summative performance rating were White, 55 percent were age 30–49, and 75 percent were female. From 2012 to 2015 the percentage of teachers younger than age 30 increased 4 percentage points, whereas the percentage of teachers age 50 and older decreased 4 percentage points. The demographic categories for race/ethnicity and gender did not change by more than 2 percentage points during the three-year period.

What the study found

This section presents the findings of the study. The analyses yielded more findings than could be included in this report; thus only differences of more than 5 percentage points between groups are discussed.

Figure 1. In 2012/13–2014/15, approximately 60 percent of teachers in the study district who had a summative performance rating were White, 55 percent were age 30–49, and 75 percent were female



Note: See table B3 in appendix B for corresponding values.

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.

Teacher demographics and ratings

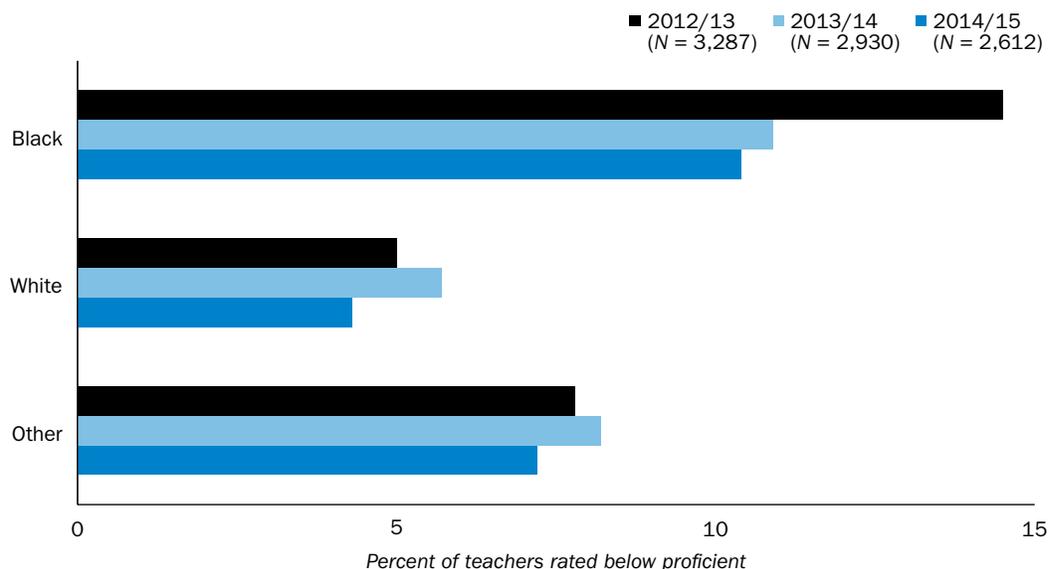
This section focuses on the small percentage of teachers who were rated below proficient in each year. However, the majority of teachers in all three years received at least a proficient summative performance rating (92 percent in 2012/13, 93 percent in 2013/14, and 94 percent in 2014/15; see table B1 in appendix B; see table B2 in appendix B for the number and percentage of teachers in each of the four rating categories).

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 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 (see tables B3 and B4 in appendix B). 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 were rated below proficient, compared with 5 percent of White teachers and 8 percent of other racial/ethnic minority teachers (figure 2; see also tables B5–B7 in appendix B). In 2013/14, 11 percent of Black teachers were rated below proficient, compared

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

Figure 2. 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: See tables B5–B7 in appendix B for corresponding values. 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.

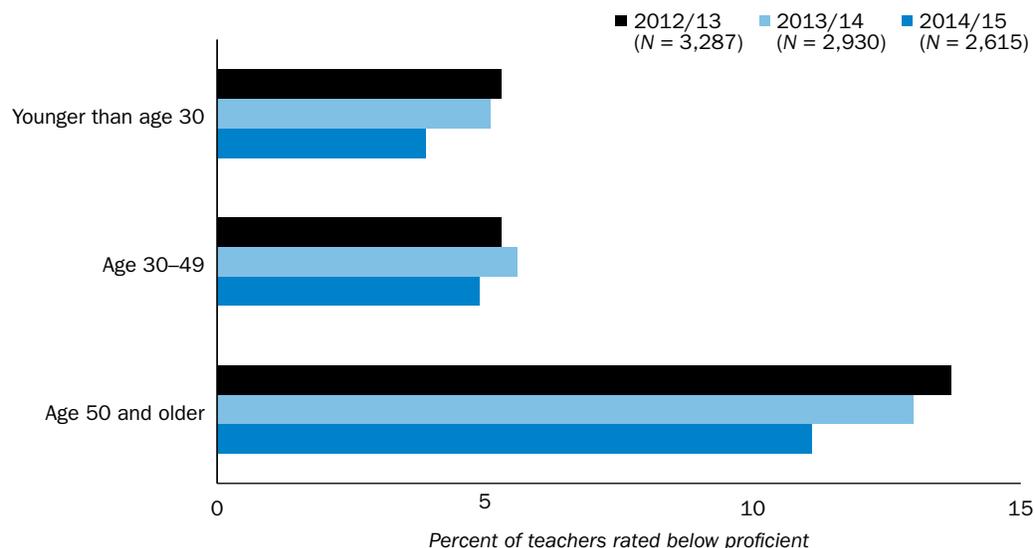
with 6 percent of White teachers and 8 percent of other racial/ethnic minority teachers. In 2014/15, 10 percent of Black teachers, 4 percent of White teachers, and 7 percent of other racial/ethnic minority teachers were rated below proficient. By 2014/15 the gap between Black and White teachers rated below proficient had narrowed from 10 percentage points to 6. The number of teachers with a summative performance rating in each year was different and decreased over time across all race/ethnicity categories (see table B3 in appendix B).

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. In 2012/13, 14 percent of teachers age 50 and older were rated below proficient, compared with 5 percent of teachers younger than age 30 and 5 percent of teachers age 30–49 (figure 3; see also tables B5–B7 in appendix B). In 2013/14, 13 percent of teachers age 50 and older were rated below proficient, compared with 5 percent of teachers younger than age 30 and 6 percent of teachers age 30–49. In 2014/15, 11 percent of teachers age 50 and older were rated below proficient, compared with 4 percent of teachers younger than age 30 and 5 percent of teachers age 30–49.

In 2012/13 and 2013/14 the percentage of teachers with a summative performance rating who were rated below proficient was higher among Black teachers age 50 and older than among White teachers and other racial/ethnic minority teachers age 50 and older. In 2012/13 and 2013/14 Black teachers accounted for the largest percentage of teachers age 50 and older who were rated below proficient: 21 percent in 2012/13 and 18 percent in 2013/14.³ The percentage for White teachers age 50 and older was 11 percent in 2012/13

In 2012/13, 14 percent of teachers age 50 and older were rated below proficient, compared with 5 percent of teachers younger than age 30 and 5 percent of teachers age 30–49

Figure 3. 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



Note: See tables B5–B7 in appendix B for corresponding values.

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

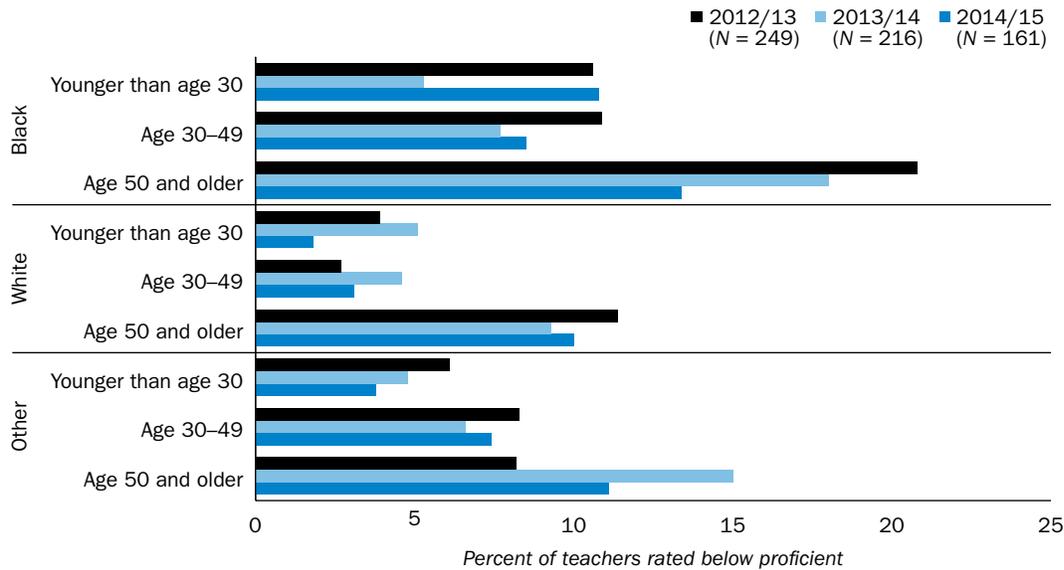
and 9 percent in 2013/14, and the percentage for other racial/ethnic minority teachers age 50 and older was 8 percent in 2012/13 and 15 percent in 2013/14 (figure 4; see also table B8 in appendix B). In 2014/15 the gap between Black and White teachers age 50 and older narrowed to 3 percentage points (13 percent of Black teachers and 10 percent of White teachers).

To put these findings in the context of demographic changes within the district’s teaching population, from 2012/13 to 2014/15 teachers age 50 and older with a summative performance rating had the highest rates of departure across all racial/ethnic groups—and particularly among Black teachers. Over that period the percentage of teachers age 50 and older who left the district was higher among Black teachers (35 percent) than among White teachers (23 percent). Moreover, the percentage of teachers age 50 and older who were rated below proficient who left the district was higher among Black teachers (63 percent) than among White teachers (51 percent; see table B9 in appendix B).

In all three years the difference in the percentage of male and female teachers with a summative performance rating who were rated below proficient was approximately 5 percentage points or less. In 2012/13 the percentage of teachers with a summative performance rating who were rated below proficient was 11 percent among male teachers, compared with 6 percent among female teachers (figure 5; see also tables B5–B7 in appendix B). In 2013/14 it was 10 percent among male teachers, compared with 6 percent among female teachers. In 2014/15 it was 8 percent among male teachers, compared with 5 percent among female teachers. The number of teachers with a summative performance rating decreased among both male and female teachers from 2012/13 to 2014/15 (see table B3 in appendix B).

In 2012/13 the percentage of teachers with a summative performance rating who were rated below proficient was 11 percent among male teachers, compared with 6 percent among female teachers. In 2013/14 it was 10 percent among male teachers, compared with 6 percent among female teachers

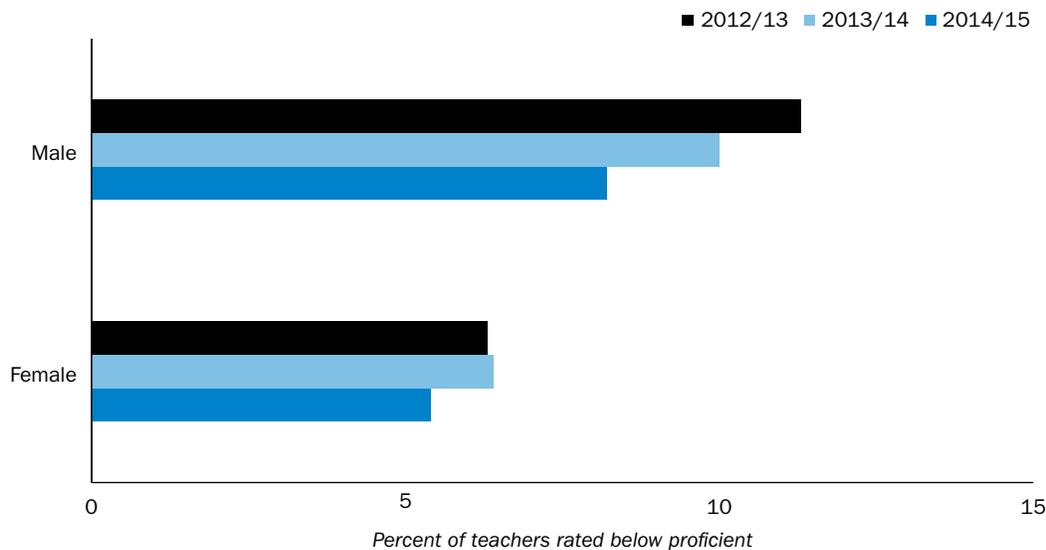
Figure 4. In 2012/13 and 2013/14 the percentage of teachers age 50 and older with a summative performance rating who were rated below proficient was higher among Black teachers than among White teachers



Note: See table B8 in appendix B for corresponding values. Black includes African American and other includes American Indian or Alaska Native, Asian, Hispanic or Latino, and Native Hawaiian or other Pacific Islander. Percentages were calculated using the number of teachers with a below proficient rating within each racial/ethnic group as the denominator, which allows for comparisons across racial/ethnic groups relative to each group's size.

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

Figure 5. In 2012/13-2014/15 the difference in the percentage of male and female teachers with a summative performance rating who were rated below proficient was less than 5 percentage points



Note: N = 3,287 for 2012/13, 2,930 for 2013/14, and 2,614 for 2014/15. See tables B5-B7 in appendix B for corresponding values.

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

In all age categories the percentage of teachers who were rated below proficient was higher among male teachers than among female teachers, except for teachers younger than age 30 in 2013/14 (see table B10 in appendix B).

The percentage of teachers rated below proficient was higher among Black male teachers than among White male teachers in 2012/13 (18 percent versus 8 percent) and in 2014/15 (12 percent versus 5 percent). A similar result was found for Black female teachers and White female teachers (see table B11 in appendix B).

Teacher characteristics and improvement in ratings

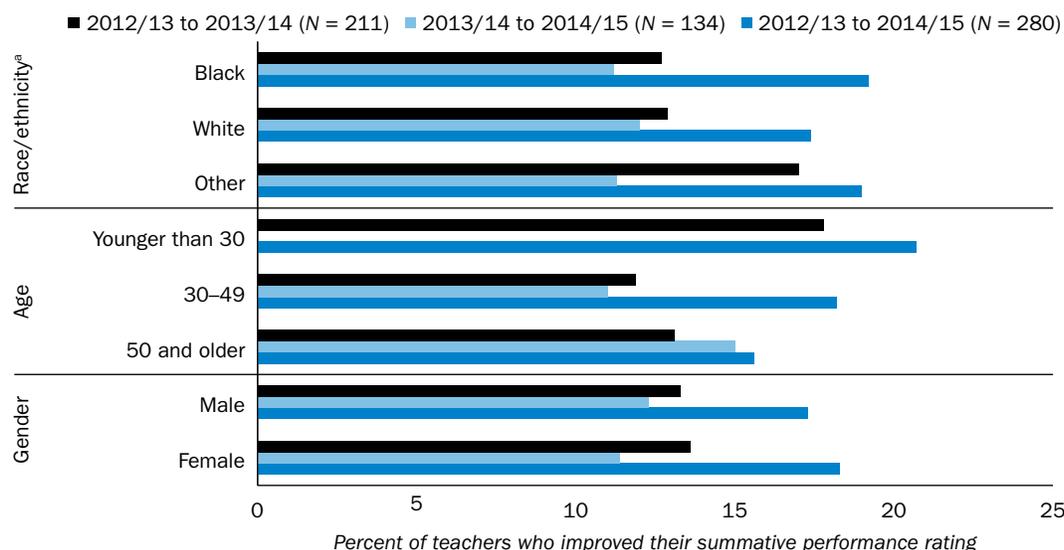
This section presents findings for three year-to-year comparisons of teachers' summative performance ratings: 2012/13 to 2013/14 and 2013/14 to 2014/15, which are both one-year comparisons, and 2012/13 to 2014/15, which is a two-year comparison. These comparisons were conducted using longitudinal samples of teachers so that only the same teachers with a summative performance rating in both years being compared were included. This longitudinal sample is limited in that attrition of the lowest performing teachers could bias the results (see the limitations section for a full discussion of this issue). The number of teachers in each comparison thus varies (see table A6 in appendix A). Whereas the previous section focused on the percentage of teachers who were rated below proficient, this section reports on all teachers who had a summative performance rating and had the opportunity to improve (that is, they had received an unsatisfactory, needs improvement, or proficient rating). Teachers who were rated exemplary were included only if their rating declined from one year to the next. Overall, 13 percent of teachers with a summative performance rating improved their rating from 2012/13 to 2013/14, 11 percent improved their rating from 2013/14 to 2014/15, and 17 percent improved their rating from 2012/13 to 2014/15 (see table B12 in appendix B).

From 2012/13 to 2013/14, 13 percent of Black teachers, 13 percent of White teachers, and 17 percent of other racial/ethnic minority teachers improved their summative performance rating

In all three year-to-year comparisons the percentages of Black, White, and other racial/ethnic minority teachers who improved their summative performance rating were within 5 percentage points of each other. From 2012/13 to 2013/14, 13 percent of Black teachers, 13 percent of White teachers, and 17 percent of other racial/ethnic minority teachers improved their summative performance rating (figure 6; see also table B12 in appendix B). From 2013/14 to 2014/15, 11 percent of Black teachers, 12 percent of White teachers, and 11 percent of other racial/ethnic minority teachers improved their summative performance rating. From 2012/13 to 2014/15, 19 percent of Black teachers, 17 percent of White teachers, and 17 percent of other racial/ethnic minority teachers improved their summative performance rating.

The percentage of teachers who improved their rating from 2012/13 to 2013/14 and from 2012/13 to 2014/15 was higher among teachers younger than age 30 than among teachers age 30 and older. From 2012/13 to 2013/14, 18 percent of teachers younger than age 30 improved their summative performance rating, compared with 12 percent of teachers age 30–49 and 13 percent of teachers age 50 and older (see figure 6 and table B12 in appendix B). From 2013/14 to 2014/15, 10 percent of teachers younger than age 30 improved their rating, compared with 11 percent of teachers age 30–49 and 15 percent of teachers age 50 and older. From 2012/13 to 2014/15, 21 percent of teachers younger than age 30 improved their rating, compared with 18 percent of teachers age 30–49 and 16 percent of teachers age 50 and older.

Figure 6. In all three year-to-year comparisons the percentages of Black, White, and other racial/ethnic minority teachers who improved their summative performance rating were within 5 percentage points of each other



Note: See table B11 in appendix B for corresponding values.

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.

The percentage of Black teachers who improved their summative performance rating from 2013/14 to 2014/15 was higher among teachers age 50 and older than among younger teachers. However, the same pattern did not exist from 2012/13 to 2013/14 or from 2012/13 to 2014/15. From 2013/14 to 2014/15 the percentage of Black teachers who improved their summative performance rating was higher among teachers age 50 and older (22 percent) than among teachers age 30–49 (8 percent) and among teachers younger than age 30 (4 percent; see table B13 in appendix B). In addition, the percentage of teachers age 50 and older who improved their rating was higher among Black teachers (22 percent) than among White teachers (11 percent) and other racial/ethnic minority teachers (13 percent). From 2012/13 to 2013/14, the percentage of Black teachers who improved their rating was lower among teachers age 50 and older (11 percent) than among teachers age 30–49 (13 percent) and among teachers younger than age 30 (17 percent). From 2012/13 to 2014/15, the percentage of Black teachers who improved their rating was lower among teachers age 50 and older (19 percent) and among teachers age 30–49 (19 percent) than among teachers younger than age 30 (21 percent).

The percentage of teachers age 50 and older who improved their rating was higher among Black teachers (22 percent) than among White teachers (11 percent) and other racial/ethnic minority teachers (13 percent)

In all three year-to-year comparisons the percentages of male and female teachers who improved their summative performance rating were within 1 percentage point of each other. From 2012/13 to 2013/14, 14 percent of female teachers improved their summative performance rating, compared with 13 percent of male teachers (see table B12 in appendix B). From 2013/14 to 2014/15, 11 percent of female teachers improved their rating, compared with 12 percent of male teachers. From 2012/13 to 2014/15, 18 percent of female teachers improved their rating, compared with 17 percent of male teachers. In all three year-to-year comparisons within each age group the percentages of male and female teachers who improved their rating were within 1–2 percentage points of each other, with one

exception. From 2013/14 to 2014/15, 21 percent of male teachers age 50 and older improved their rating, compared with 13 percent of female teachers age 50 and older (see table B14 in appendix B). From 2012/13 to 2013/14 and from 2012/13 to 2014/15 the percentage of teachers who improved their rating was less than 2 percentage points higher among Black and White female teachers than among Black and White male teachers (see table B15 in appendix B). The percentage of teachers in the other race/ethnicity category who improved their rating was less than 5 percentage points among male teachers and female teachers across all three year-to-year comparisons.

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. Moreover, while the percentage of Black teachers and older teachers who were rated below proficient decreased over time in some cases, gaps between their rating and the ratings of their White and younger counterparts persisted. Conducting the year-to-year analyses also revealed that the percentage of teachers making improvements in their rating was fairly consistent across these race/ethnicity, age, and gender categories. These findings confirm the district's concerns that motivated this study. Thus, the district may want to examine the root cause of these disproportionalities and consider what programs or policies aimed specifically at those teachers and their evaluators may increase their chances for improvement and reduce the gaps.

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

Additionally, the district may want to examine other factors that could be related to those teachers' ratings, including the characteristics of the schools and classes in which they teach (such as grade level, subject taught, student demographics, and others) and the professional training that teachers received, as well as the supports available to them. However, the district will need to ensure that these characteristics are recorded in their data management system with high enough quality to warrant inclusion in the analyses. Future research could investigate the relationship between teachers' demographic characteristics and school and class characteristics to further understand the nature of the disparity observed in this study. Related to this, a future area of research is to further examine whether the patterns observed using the summative performance ratings are consistent by standard or if the patterns are not observed or are more dramatic in some standards, particularly over time. This awareness could be useful in understanding the underlying causes of the gaps between groups.

Given that some of the gaps declined (for example, the gap between Black and White teachers rated below proficient declined by almost half) but some persisted (for example, the percentages of Black and White teachers who improved were similar), further research is needed to understand what accounts for those patterns. Additional analyses may help confirm whether the patterns reported in the first research question are an artifact of the changes in the district's teaching population over time—in that many teachers did not have a summative rating in one or more of the years of the study—or a change related to contextual factors within the district (see appendix A for analysis of the study population over time). Patterns of mobility and retention of teachers from various subgroups may be worthy areas of investigation.

Further research is needed to understand patterns in ratings over time and whether the results might be biased because the sample included only teachers with below proficient summative performance ratings (and thus, with the most room for improvement). This research might examine patterns in ratings for teachers rated proficient to see whether the patterns are consistent or differ for those teachers moving from proficient to below proficient.

In addition, longitudinal research is needed to examine whether the patterns persist over time or whether district-level interventions and supports might reduce the gap or otherwise address the disproportionate below proficient ratings among teachers in certain groups. For example, the district has invested in providing additional support for male teachers who are Black or other race/ethnicity, and this kind of targeted support could be studied to examine the effect on teachers' subsequent ratings. Moreover, the district has focused some attention on addressing training and support for evaluators related to potential biases that they may bring to their evaluation practice. Research might examine the effect of this type of support on evaluators' rating of teachers from diverse backgrounds.

Lastly, this study's focus on studying improvement in ratings for teachers who were initially rated below proficient may lead to biased results in that these teachers had the most room for improvement.

This study did not investigate evaluator characteristics, nor did it address the extent to which evaluator characteristics or the pairing of evaluator and teacher characteristics may relate to teacher evaluation outcomes. Further research may also investigate how evaluator characteristics relate to teacher evaluation outcomes and whether evaluators with certain characteristics tend to rate teachers of varying characteristics differently.

Although the study findings were specific to the district, this report may be useful to other states and districts that are designing and implementing new evaluation systems. Specifically, these analyses may serve as a model to help other districts examine patterns in evaluation ratings and teacher demographics. For the district, this study is part of a larger goal 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 is part of the district's effort to ensure that the system is meeting its human capital goals.

Limitations of the study

Although the study draws on a large urban district's teacher evaluation data over three years, several limitations inherent to the data and the design warrant caution in interpreting the results.

First, the study is purely descriptive in design rather than causal. The study was not designed to demonstrate whether teacher demographic characteristics were the cause for any variation in evaluation outcomes.

Second, 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,

Further research is needed to examine whether the patterns persist over time or whether district-level interventions and supports might reduce the gap or otherwise address the disproportionate below proficient ratings among teachers in certain groups

and others entered; those who left may have been the district's lowest performing teachers. Teachers might not have received a summative performance rating for one year for various reasons, including being placed on a two-year evaluation plan, which provides a summative performance rating only every other year. To address this concern, the report includes the percentage of teachers who left the district each year and the percentage on a two-year evaluation plan (see appendix A). In addition, missing-data analyses were conducted to examine the differences between teachers who had a rating and teachers who did not. Missing-data analyses conducted to examine differences between those who had a rating in both years and those who were missing a rating in the second year revealed statistically significant differences by race/ethnicity and age. This indicates that reported percentages of teachers improving their rating, examined in research question 2, may be due partially to missing data in the population of teachers examined. See appendix A for more information about missing data.

Third, the study is based on an educator evaluation system that did not yet include information about teachers' effect on student learning through the examination of student test scores or other evidence of student growth, which is included in the evaluation system as of 2015/16. The findings are based solely on evaluators' ratings within a system that included classroom observations, teachers' presentations of evidence related to standards of teaching practice, and teachers' progress toward their professional and student learning goals. The study does not attempt to speculate as to whether these are objective measures of teacher practice.

The study was not designed to demonstrate whether teacher demographic characteristics were the cause for any variation in evaluation outcomes

Appendix A. Study data and methodology

This appendix describes the data sources and details the study methodology.

Data

The data to address both research questions were derived from the district's online platform designed to support the educator performance evaluation system. According to the district's public website, the purpose of the online system is to allow teachers and administrators to view and record the evaluation process and facilitate communication between teachers and evaluators about opportunities for development. Table A1 presents the data elements used for each research question.

The evaluation ratings in the dataset include summative performance evaluation ratings for the school years 2012/13–2014/15. The data also include teachers' race/ethnicity, age, and gender. The race/ethnicity categories from the district were collapsed into three categories for ease of analysis and reporting, as well as to suppress small cell sizes. Those categories are White, Black (includes African American), and other (includes American Indian or Alaska Native, Asian, Hispanic or Latino, and Native Hawaiian or other Pacific Islander). Age was categorized using the age brackets younger than 30, 30–49, and 50 and older. Gender was categorized as male or female.

Methodology

The following sections describe the analyses conducted for this study, including correlational analyses, missing-data analyses, and the specific analysis used for each research question.

Correlational analyses. Before conducting analyses for either research question, the pairwise correlations of characteristic variables were checked using the Phi coefficient. In other words, the correlation for each combination of teacher characteristics (for example, race/ethnicity and age) was examined separately. This analysis was done to ensure that findings pertaining to one characteristic variable were not moderated by another characteristic variable that could explain the observed finding. For instance, if race/ethnicity and age were correlated, examining differences in ratings by race/ethnicity alone could lead to misinterpreted findings if most of the older teachers were of one race/ethnicity and most of the younger teachers were of another race/ethnicity. In this example, the findings could

Table A1. Data elements for research questions

Data element	Categories
Race/ethnicity	Black (includes African American), White, other (includes American Indian or Alaska Native, Asian, Hispanic or Latino, and Native Hawaiian or other Pacific Islander)
Age	Younger than 30, 30–49, 50 and older
Gender	Male, female
Summative performance rating	Exemplary, proficient, needs improvement, unsatisfactory

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

have been driven by differences due to age (a proxy for experience since the district did not have a reliable variable that measured experience).

Hamilton’s (1990) guidelines for interpreting the correlation coefficient were used, where a correlation of at least ± 0.5 is considered moderate and a correlation of at least ± 0.8 is considered strong. In all of the correlation analyses of combinations of variables (for example, race/ethnicity and age), only small correlations were observed (< 0.15), meaning that the variables were, at most, only slightly related to one another. Therefore, the analyses were conducted using one characteristic variable at a time and interpreted to not be influenced by another variable. Regardless, analyses using combinations of characteristics also were conducted, as described below, because of interest from the district in specific combinations of characteristics.

Missing-data analyses. Two sets of missing-data analyses were conducted: the first set was relevant to research question 1, and the second set was relevant to research question 2. The first set of analyses examined possible differences between the characteristics of teachers who had a summative performance rating in each of the three years and the characteristics of teachers who did not. The second set examined differences between the characteristics of teachers for whom improvement over time could be calculated and the characteristics of teachers who were missing a rating in a subsequent year and thus were excluded from the improvement analyses for research question 2.

Research question 1. The district provided a census dataset that included information for every teacher in the district. The total number of teachers in the dataset differed by year and in each year the number of teachers with a summative performance rating was lower than in the previous year, falling from 72 percent of the total in 2012/13 to 57 percent in 2014/15 (table A2).

The reasons the total number of teachers in the district dataset and the number of teachers with a summative rating differ include:

- Teachers who had a formative evaluation rating (that is, those on two-year evaluation plans) were included in the dataset but were not included in the study because the focus was on teachers who had a summative performance rating. This group explains the largest percentage of teachers missing a summative performance rating from year to year (see table A6 for the corresponding percentages of teachers on two-year evaluation plans from year to year).
- Additional teachers were included in the dataset (for example, related service providers, such as physical and occupational therapists, who are not part of the same teacher evaluation system and thus were not included in the study).

Table A2. Total number of teachers and number of teachers with a summative performance rating, 2012/13–2014/15

Teacher group	2012/13	2013/14	2014/15
Total number of teachers	4,590	4,654	4,627
Teachers with a summative performance rating	3,287	2,930	2,615

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

- Teachers may not have received a summative performance rating for one of several other reasons: the teacher took a leave of absence before receiving a rating, an evaluator did not start a plan for the teacher, an evaluator neglected to provide the teacher with a summative rating, or the teacher left the district before receiving a rating.

Although the percentages of teachers in the first two categories were able to be calculated, the exact percentage of teachers in the last category could not be calculated because a reason for missing a rating was not explicitly identified in the district’s online evaluation data system. However, the study team did examine the extent to which the teachers who had a rating differed from teachers who did not.

A series of chi-square tests were used to determine whether the characteristics of teachers who had a summative performance rating were different from the characteristics of teachers who did not. For the 2012/13 analysis, differences between the two groups in terms of age were statistically significant (table A3).⁴

For the 2013/14 and 2014/15 analyses there were significant differences between the two groups in terms of race/ethnicity, age, and gender (tables A4 and A5).

There were statistically significant differences between teachers with a summative evaluation rating and the larger population of teachers within the district with regard to all the teacher characteristics of interest, particularly age, as those differences were consistent across the three years. These results are important to guide the interpretation of findings; however, the teachers with a summative evaluation rating in each year represent the population of interest for this study.

Table A3. Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating, 2012/13

Characteristic	Had a rating		Did not have a rating	
	Number	Percent	Number	Percent
Total	3,287	71.6	1,303	28.4
Race/ethnicity ^a				
Black	730	71.1	297	28.9
White	2,018	71.6	800	28.4
Other	539	72.3	206	27.7
Age				
Younger than 30	582	77.4	170	22.6
30–49	1,808	71.6	717	28.4
50 and older	897	68.3	416	31.7
Gender				
Male	822	69.4	362	30.6
Female	2,465	72.4	941	27.6

Note: The chi-squared test statistic was 0.34 ($p = .84$) for race/ethnicity, 19.38 ($p < .01$) for age, and 3.75 ($p = .05$) for gender.

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.

Table A4. Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating, 2013/14

Characteristic	Had a rating		Did not have a rating	
	Number	Percent	Number	Percent
Total	2,930	63.0	1,724	37.0
Race/ethnicity ^a				
Black	685	69.1	307	30.9
White	1,758	60.7	1,139	39.3
Other	487	63.7	278	36.3
Age				
Younger than 30	573	71.4	230	28.6
30–49	1,602	62.4	966	37.6
50 and older	755	58.8	528	41.2
Gender				
Male	791	66.3	402	33.7
Female	2,139	61.8	1,322	38.2

Note: The chi-squared test statistic was 22.39 ($p < .01$) for race/ethnicity, 33.96 ($p < .01$) for age, and 7.71 ($p < .01$) for gender.

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 2013/14.

Table A5. Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating, 2014/15

Characteristic	Had a rating		Did not have a rating	
	Number	Percent	Number	Percent
Total	2,615	56.5	2,012	43.5
Race/ethnicity ^a				
Black	566	59.2	390	40.8
White	1,563	54.6	1,299	45.4
Other	483	59.9	323	40.1
Age				
Younger than 30	583	69.3	258	30.7
30–49	1,422	54.9	1,169	45.1
50 and older	610	51.0	585	49.0
Gender				
Male	708	59.4	484	40.6
Female	1,906	55.5	1,526	44.5

Note: The chi-squared test statistic was 10.84 ($p < .05$) for race/ethnicity, 73.49 ($p < .01$) for age, and 5.36 ($p < .05$) for gender. Data on race/ethnicity were missing for three teachers who had a rating, and data on gender were missing for one teacher who had a rating and for two teachers who did not have a rating.

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 2014/15.

The differences between teachers who had and did not have a rating may not be surprising given that many teachers did not receive a summative rating because they were on a two-year evaluation plan. Only teachers rated at least proficient are eligible for a two-year evaluation plan. Chi-square tests were conducted to explore whether teachers on one-year and two-year evaluation plans differed. The results were nearly the same as those in the previous analyses—they showed significant differences by age in each year, by gender and race/ethnicity in 2013/14, and by race/ethnicity in 2014/15 (see tables B16–B18 in appendix B). These results suggest that the characteristics of teachers rated at least proficient—thus eligible for the two-year evaluation plan—differed from the characteristics of teachers evaluated using the one-year plan.

Research question 2 missing-data analyses. Missing-data analyses were conducted to examine differences between teachers for whom improvement over time could be calculated and teachers who were missing a rating in a subsequent year and thus were excluded from the improvement analyses. The first step consisted of calculating the extent of missing data for each of the year-to-year comparisons (table A6). The percentage of missing data was calculated as:

$$100\% - (\% \text{ of teachers with a rating in both years} + \% \text{ on two-year evaluation plan} + \% \text{ departed district after base year}) = \% \text{ missing}$$

To obtain an accurate portrayal of the extent of missing data, the percentages of teachers on a two-year evaluation plan and those that had left the district after the first year were calculated. Approximately 11 percent of teachers who had a summative performance rating in 2012/13 were missing a summative performance rating in 2013/14; the reason why those data were missing is unknown (see table A6). From 2013/14 to 2014/15 the percentage was 16 percent, and over the three-year period it was 12 percent.

The missing-data analyses were conducted to compare data for teachers with a rating in both years to those without a rating, including those on two-year evaluation plans, those who departed the district, and those missing a rating for other reasons. The analyses revealed statistically significant results for the 2012/13 to 2013/14 comparison and the 2013/14 to 2014/15 comparison, specifically for age and race/ethnicity (tables A7 and A8). For the 2012/13 to 2014/15 comparison, the results were significant only for age (table A9). These results indicate that there are differences by age and race/ethnicity between

Table A6. Overview of missing data, 2012/13–2014/15

Reason for missing data	2012/13 to 2013/14		2013/14 to 2014/15		2012/13 to 2014/15	
	Number	Percent	Number	Percent	Number	Percent
Rating in both years	1,614	49.1	1,198	40.9	1,697	51.6
On two-year plan in second year	983	29.8	857	29.2	553	16.8
Departed district after base year	341	10.4	414	14.1	639	19.4
Unknown	349	10.6	461	15.7	398	12.1

Note: Data include only teachers who had a summative rating in the base year; therefore, the actual number of teachers on a two-year evaluation plan or that departed the district after the base year may be higher. Percentages are based on the number of teachers with a summative performance rating in the base year and may not sum to 100 because of rounding. For example, 49.1 percent represents the percentage of teachers with a rating in both 2012/13 and 2013/14, compared with the total number of teachers with a rating in 2012/13.

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

Table A7. Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating in both 2012/13 and 2013/14

Characteristic	Had a rating		Did not have a rating	
	Number	Percent	Number	Percent
Total	1,614	49.1	1,673	50.9
Race/ethnicity ^a				
Black	386	52.9	344	47.1
White	951	47.1	1,067	52.9
Other	277	51.4	262	48.6
Age				
Younger than 30	363	62.4	219	37.6
30–49	860	47.6	948	52.4
50 and older	391	43.6	506	56.4
Gender				
Male	415	50.5	407	49.5
Female	1,199	48.6	1,266	51.4

Note: Data include the number and percentage of teachers in a teacher characteristic category who had a rating in both years and those who were missing a rating in 2013/14. For example, 62.4 percent of teachers younger than age 30 had a rating in both 2012/13 and 2013/14. The chi-squared test statistic was 8.45 ($p < .05$) for race/ethnicity, 53.61 ($p < .01$) for age, 0.84 ($p = .36$) for gender.

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 and 2013/14.

Table A8. Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating in both 2013/14 and 2014/15

Characteristic	Had a rating		Did not have a rating	
	Number	Percent	Number	Percent
Total	1,198	40.9	1,732	59.1
Race/ethnicity ^a				
Black	302	44.1	383	55.9
White	680	38.7	1,078	61.3
Other	215	44.1	272	55.9
Age				
Younger than 30	368	64.2	205	35.8
30–49	582	36.3	1,020	63.7
50 and older	247	32.7	508	67.3
Gender				
Male	338	42.7	453	57.3
Female	859	40.2	1,280	59.8

Note: The data include the number and percentage of teachers in a teacher characteristic category who had a rating in both years and those who were missing a rating in 2014/15. For example, 64.2 percent of teachers younger than age 30 had a rating in both 2013/14 and 2014/15. The chi-squared test statistic was 8.59 ($p < .05$) for race/ethnicity, 163.78 ($p < .01$) for age, and 1.59 ($p = .21$) for gender.

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 2013/14 and 2014/15.

Table A9. Characteristics of teachers who had a summative performance rating and of teachers who did not have a summative performance rating in both 2012/13 and 2014/15

Characteristic	Had a rating		Did not have a rating	
	Number	Percent	Number	Percent
Total	1,697	51.6	1,590	48.4
Race/ethnicity^a				
Black	355	48.6	375	51.4
White	1,054	52.2	964	47.8
Other	288	53.4	251	46.6
Age				
Younger than 30	323	55.5	259	44.5
30–49	961	53.2	847	46.8
50 and older	413	46.0	484	54.0
Gender				
Male	435	52.9	387	47.1
Female	1,262	51.2	1,203	48.8

Note: Data include the number and percentage of teachers in a teacher characteristic category who had a rating in both years and those who were missing a rating in 2014/15. For example, 55.5 percent of teachers younger than age 30 had a rating in both 2012/13 and 2014/15. The chi-squared test statistic was 3.62 ($p = .16$) for race/ethnicity, 16.38 ($p < .01$) for age, and 0.73 ($p = .39$) for gender.

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 and 2014/15.

teachers who had a summative rating in both years and those who were missing a rating in the latter year. This indicates that reported percentages of teachers improving their rating, examined in the second research question, may be due partially to missing data in the population of teachers examined.

Analysis by research question. This section details the analysis for each research question.

Research question 1. To address the question of whether the percentages of teachers who receive a below proficient summative performance rating vary by teacher characteristics, the summative performance ratings were collapsed into two groups: below proficient includes unsatisfactory and needs improvement ratings, and at least proficient includes proficient and exemplary. The ratings were collapsed to suppress the small sample size in the lowest category when analyses were conducted that further disaggregated the group and therefore protect the privacy of the small number of teachers in that category (see table B2 in appendix B).

Frequencies were calculated to display, by teacher characteristic, the number and percentage of teachers who received each category of rating. This research question involved separate analyses of ratings for each of the three years available (2012/13–2014/15). Frequency analyses using combinations of characteristics, such as the combination of race/ethnicity and age and of gender and race/ethnicity, also were calculated.

No tests of statistical significance were conducted, as the data represent the full population of teachers receiving a summative performance rating in each year.

Research question 2. To answer the question of whether the percentages of teachers who improve their summative performance rating over three years vary by teacher characteristics, dichotomous categorical variables were developed to represent whether a teacher improved, declined, or remained at the same rating (variable=1 if teacher improved; 0 otherwise).⁵ This analysis was conducted for the summative performance rating using three years of available data. Three variables were created for this analysis—one per year-to-year comparison. Table A10 provides an example of the variable values assigned to indicate whether the rating improved for the summative performance rating. The summative performance rating may rise to the maximum level of exemplary, followed by proficient, followed by needs improvement.

Cross-tabulations were used to obtain the frequencies of the number and percentage of teachers in the study population who improved their rating compared with those who did not; these data were also disaggregated by race/ethnicity, age and gender. Statistical tests were not necessary because the analysis was based on the full population of teachers who had a summative performance rating from one year to the next.

Table A10. Research question 2 example variable values

Variable	Variable values	Description	Examples
Improvement from 2012/13 to 2013/14	0	No improvement or a decline in rating from previous year	Rating stayed proficient, needs improvement, or unsatisfactory; rating declined from proficient in 2013 to needs improvement in 2014, and the like.
	1	Improvement from previous year	Rating improved from unsatisfactory in 2013 to needs improvement, proficient, or exemplary in 2014; rating improved from needs improvement to proficient or exemplary; rating improved from proficient to exemplary.
Improvement from 2013/14 to 2014/15	0	No improvement or a decline in rating from previous year	Rating stayed proficient, needs improvement, or unsatisfactory; rating declined from proficient in 2014 to needs improvement in 2015, and the like.
	1	Improvement from previous year	Rating improved from unsatisfactory in 2014 to needs improvement, proficient, or exemplary in 2015; rating improved from needs improvement to proficient or exemplary; rating improved from proficient to exemplary.
Improvement from 2012/13 to 2014/15	0	No improvement or a decline in rating between 2012/13 and 2014/15	Rating stayed proficient, needs improvement, or unsatisfactory; rating declined from proficient in 2013 to needs improvement in 2015; and the like.
	1	Improvement in rating between 2012/13 and 2014/15	Rating improved from unsatisfactory to needs improvement, proficient, or exemplary; rating improved from needs improvement to proficient or exemplary; rating improved from proficient to exemplary.

Source: Authors' creation.

Appendix B. Supplementary tables

This appendix includes additional tables referenced throughout the report. The first set of tables provides the summative performance ratings for all teachers combined. The second set of tables provides the demographic composition of the teacher population, with summative performance ratings in the district, by year, and the demographic composition of teachers with below proficient summative performance ratings, by year. Those tables are followed by tables for the summative performance ratings, disaggregated by characteristic.

Summative performance ratings for all teachers combined

Table B1. Summative performance ratings, 2012/13–2014/15

Rating	2012/13 (N = 3,287)		2013/14 (N = 2,930)		2014/15 (N = 2,615)	
	Number	Percent	Number	Percent	Number	Percent
Below proficient	249	7.6	216	7.4	161	6.2
At least proficient	3,038	92.4	2,714	92.6	2,454	93.8

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

Table B2. Summative performance ratings using all four rating categories, 2012/13–2014/15

Rating	2012/13 (N = 3,287)		2013/14 (N = 2,930)		2014/15 (N = 2,615)	
	Number	Percent	Number	Percent	Number	Percent
Unsatisfactory	49	1.5	43	1.5	37	1.4
Needs improvement	200	6.1	173	5.9	124	4.7
Proficient	2,661	81.0	2,288	78.1	2,019	77.2
Exemplary	377	11.5	426	14.5	435	16.6

Note: Percentages may not sum to 100 because of rounding.

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

Teacher demographics

Table B3. Demographic characteristics of all teachers with a summative performance rating, 2012/13–2014/15

Characteristic	2012/13 (N = 3,287)		2013/14 (N = 2,930)		2014/15 (N = 2,615)	
	Number	Percent	Number	Percent	Number	Percent
Race/ethnicity ^a						
Black	730	22.2	685	23.4	566	21.6
White	2,018	61.4	1,758	60.0	1,563	59.8
Other	539	16.4	487	16.6	483	18.5
Age						
Younger than 30	582	17.7	573	19.6	583	22.3
30–49	1,808	55.0	1,602	54.7	1,422	54.4
50 and older	897	27.3	755	25.8	610	23.3
Gender						
Male	822	25.0	791	27.0	708	27.1
Female	2,465	75.0	2,139	73.0	1,906	72.9

Note: Data on race/ethnicity in 2014/15 were missing for three teachers, and data on gender in 2014/15 were missing for one teacher. Percentages may not sum to 100 because of rounding.

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.

Table B4. Demographic characteristics of teachers with a below proficient summative performance rating, 2012/13–2014/15

Characteristic	2012/13 (N = 249)		2013/14 (N = 216)		2014/15 (N = 161)	
	Number	Percent	Number	Percent	Number	Percent
Race/ethnicity ^a						
Black	106	42.6	75	34.7	59	36.6
White	101	40.6	101	46.8	67	41.6
Other	42	16.9	40	18.5	35	21.7
Age						
Younger than 30	31	12.4	29	13.4	23	14.3
30–49	95	38.2	89	41.2	70	43.5
50 and older	123	49.4	98	45.4	68	42.2
Gender						
Male	93	37.3	79	36.6	58	36.0
Female	156	62.7	137	63.4	103	64.0

Note: Percentages may not sum to 100 because of rounding.

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.

Summative performance ratings by teacher characteristics

Table B5. Summative performance ratings, by teacher characteristics, 2012/13

Characteristic	At least proficient		Below proficient	
	Number	Percent	Number	Percent
Race/ethnicity ^a				
Black (N = 730)	624	85.5	106	14.5
White (N = 2,018)	1,917	95.0	101	5.0
Other (N = 539)	497	92.2	42	7.8
Age				
Younger than 30 (N = 582)	551	94.7	31	5.3
30–49 (N = 1,808)	1,713	94.7	95	5.3
50 and older (N = 897)	774	86.3	123	13.7
Gender				
Male (N = 822)	729	88.7	93	11.3
Female (N = 2,465)	2,309	93.7	156	6.3

Note: N = 3,287.

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.

Table B6. Summative performance ratings, by teacher characteristics, 2013/14

Characteristic	At least proficient		Below proficient	
	Number	Percent	Number	Percent
Race/ethnicity ^a				
Black (N = 685)	610	89.1	75	10.9
White (N = 1,758)	1,657	94.3	101	5.7
Other (N = 487)	447	91.8	40	8.2
Age				
Younger than 30 (N = 573)	544	94.9	29	5.1
30–49 (N = 1,602)	1,513	94.4	89	5.6
50 and older (N = 755)	657	87.0	98	13.0
Gender				
Male (N = 791)	712	90.0	79	10.0
Female (N = 2,139)	2,002	93.6	137	6.4

Note: N = 2,930.

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 2013/14.

Table B7. Summative performance ratings, by teacher characteristics, 2014/15

Characteristic	At least proficient		Below proficient	
	Number	Percent	Number	Percent
Race/ethnicity^a				
Black (N = 566)	507	89.6	59	10.4
White (N = 1,563)	1,496	95.7	67	4.3
Other (N = 483)	448	92.8	35	7.2
Age				
Younger than 30 (N = 583)	560	96.1	23	3.9
30–49 (N = 1,422)	1,352	95.1	70	4.9
50 and older (N = 610)	542	88.9	68	11.1
Gender				
Male (N = 708)	650	91.8	58	8.2
Female (N = 1,906)	1,803	94.6	103	5.4

Note: N = 2,615. Data on race/ethnicity were missing for three teachers, and data on gender were missing for one teacher.

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 2014/15.

Table B8. Percentage of teachers with a below proficient summative performance rating, by race/ethnicity and age, 2012/13–2014/15

Characteristic	2012/13 (N = 249)		2013/14 (N = 216)		2014/15 (N = 161)	
	Number	Percent	Number	Percent	Number	Percent
Black						
Younger than age 30	9	10.6	5	5.3	12	10.8
Age 30–49	41	10.9	27	7.7	24	8.5
Age 50 and older	56	20.8	43	18.0	23	13.4
Total	106	14.5	75	10.9	59	10.4
White						
Younger than age 30	15	3.9	18	5.1	6	1.8
Age 30–49	31	2.7	46	4.6	28	3.1
Age 50 and older	55	11.4	37	9.3	33	10.0
Total	101	5.0	101	5.7	67	4.3
Other						
Younger than age 30	7	6.1	6	4.8	5	3.8
Age 30–49	23	8.3	16	6.6	18	7.4
Age 50 and older	12	8.2	18	15.0	12	11.1
Total	42	7.8	40	8.2	35	7.2

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.

Table B9. Number and percentage of teachers with a summative performance evaluation rating who left the district, by age and race/ethnicity, 2012/13–2014/15

Characteristic	Departures from 2012/13 to 2014/15	
	Number	Percent
Age 50 and older		
Black (N = 269)	93	35
White (N = 482)	110	23
Age 50 and older and below proficient		
Black (N = 56)	35	63
White (N = 55)	28	51

Note: Black includes African American.

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

Table B10. Percentage of teachers receiving a below proficient summative performance rating, by age and gender, 2012/13–2014/15

Characteristic	2012/13 (N = 249)		2013/14 (N = 216)		2014/15 (N = 161)	
	Number	Percent	Number	Percent	Number	Percent
Younger than age 30						
Male	10	8.3	6	4.7	7	5.5
Female	21	4.5	23	5.2	16	3.5
Total	31	12.8	29	9.9	23	9.0
Age 30–49						
Male	34	7.3	35	7.9	32	7.7
Female	61	4.5	54	4.7	38	3.8
Total	95	11.8	89	12.6	70	11.5
Age 50 and older						
Male	49	20.7	38	17.2	19	11.6
Female	74	11.2	60	11.2	49	11.0
Total	123	31.9	98	28.4	68	22.6

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

Table B11. Percentage of teachers receiving a below proficient summative performance rating, by gender and race/ethnicity, 2012/13–2014/15

Characteristic	2012/13 (N = 249)		2013/14 (N = 216)		2014/15 (N = 161)	
	Number	Percent	Number	Percent	Number	Percent
Male						
Black	33	17.7	25	13.7	18	12.2
White	42	8.3	45	9.1	23	5.3
Other	18	13.5	9	8.0	17	13.6
Total	93	39.5	79	30.8	58	31.1
Female						
Black	73	13.4	50	9.9	41	9.8
White	59	3.9	56	4.4	44	3.9
Other	24	15.4	31	8.3	18	5.0
Total	156	32.7	137	22.6	103	18.7

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.

Improvement in summative performance ratings, by teacher characteristics

Table B12. Percentage of teachers whose summative performance ratings improved, by teacher characteristics, 2012/13–2014/15

Characteristic	2012/13 to 2013/14		2013/14 to 2014/15		2012/13 to 2014/15	
	Number	Percent	Number	Percent	Number	Percent
Total	211	13.5	134	11.7	280	18.0
Race/ethnicity^a						
Black	48	12.7	33	11.2	64	19.2
White	118	12.9	78	12.0	165	17.4
Other	45	17.0	23	11.3	51	19.0
Age						
Younger than 30	63	17.8	36	10.4	62	20.7
30–49	98	11.9	61	11.0	158	18.2
50 and older	50	13.1	37	15.0	60	15.6
Gender						
Male	54	13.3	40	12.3	72	17.3
Female	157	13.6	94	11.4	208	18.3

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.

Improvement in summative performance ratings, by combinations of teacher characteristic

Table B13. Percentage of teachers whose summative performance ratings improved, by race/ethnicity and age, 2012/13–2014/15

Characteristic	2012/13 to 2013/14		2013/14 to 2014/15		2012/13 to 2014/15	
	Number	Percent	Number	Percent	Number	Percent
Black						
Younger than age 30	9	17.3	2	3.6	10	21.3
Age 30–49	25	12.6	12	7.9	35	18.8
Age 50 and older	14	11.0	19	21.8	19	19.0
White						
Younger than age 30	37	15.5	25	11.4	35	18.3
Age 30–49	54	11.0	40	12.9	100	18.3
Age 50 and older	27	14.3	13	10.8	30	14.2
Other						
Younger than age 30	17	26.6	9	12.5	17	27.4
Age 30–49	19	14.1	9	9.8	23	17.2
Age 50 and older	9	13.8	5	12.8	11	15.1

Note: Total number that improved is 211 out of 1,559 for 2012/13 to 2013/14 analysis; 134 out of 1,147 for 2013/14 to 2014/15 analysis; 80 out of 1,552 for 2012/13 to 2014/15 analysis. Data do not include teachers who remained at exemplary. 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.

Table B14. Percentage of teachers whose summative performance ratings improved, by age and gender, 2012/13–2014/15

Characteristic	2012/13 to 2013/14		2013/14 to 2014/15		2012/13 to 2014/15	
	Number	Percent	Number	Percent	Number	Percent
Younger than age 30						
Male	15	17.4	8	9.5	13	20.3
Female	48	17.9	28	10.6	49	20.8
Age 30–49						
Male	27	12.3	17	10.0	45	18.1
Female	71	11.7	44	11.5	113	18.3
Age 50 and older						
Male	12	12.0	15	20.8	14	13.6
Female	38	13.5	22	12.6	46	16.3

Note: Total number that improved is 211 out of 1,559 for 2012/13 to 2013/14 analysis; 134 out of 1,147 for 2013/14 to 2014/15 analysis; 80 out of 1,552 for 2012/13 to 2014/15 analysis. Data do not include teachers who remained at exemplary.

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

Table B15. Percentage of teachers whose summative performance ratings improved, by gender and race/ethnicity, 2012/13–2014/15

Characteristic	2012/13 to 2013/14		2013/14 to 2014/15		2012/13 to 2014/15	
	Number	Percent	Number	Percent	Number	Percent
Male						
Black	11	11.3	10	13.7	16	18.6
White	30	12.3	25	12.4	44	16.7
Other	13	20.0	5	9.6	12	18.2
Female						
Black	37	13.2	23	10.4	48	19.4
White	88	13.0	53	11.8	121	17.6
Other	32	16.1	18	11.9	39	14.8

Note: Total number that improved is 211 of 1,559 for 2012/13 to 2013/14 analysis; 134 of 1,147 for 2013/14 to 2014/15 analysis, 80 of 1,552 for 2012/13 to 2014/15 analysis. Data do not include teachers who remained at exemplary. 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.

Table B16. Characteristics of teachers on one-year and two-year evaluation plans, 2012/13

Characteristic	One-year evaluation plans		Two year evaluation plans	
	Number	Percent	Number	Percent
Total	3,336	84.8	597	15.2
Race/ethnicity^a				
Black	739	87.0	110	13.0
White	2,054	83.9	394	16.1
Other	543	85.4	93	14.6
Age				
Younger than 30	585	91.5	54	8.5
30–49	1,842	83.5	364	16.5
50 and older	909	83.5	179	16.5
Gender				
Male	2,503	85.2	436	14.8
Female	833	83.8	161	16.2

Note: The chi-squared test statistic was 5.00 ($p = .08$) for race/ethnicity, 26.83 ($p < .01$) for age, and 1.07 ($p = .05$) for gender.

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.

Table B17. Characteristics of teachers on one-year and two-year evaluation plans, 2013/14

Characteristic	One-year evaluation plans		Two year evaluation plans	
	Number	Percent	Number	Percent
Total	3,004	74.3	1,041	25.7
Race/ethnicity ^a				
Black	689	79.7	175	20.3
White	1,822	72.0	708	28.0
Other	493	75.7	158	24.3
Age				
Younger than 30	577	87.2	85	12.8
30–49	1,636	71.7	646	28.3
50 and older	791	71.8	310	28.2
Gender				
Male	795	76.9	239	23.1
Female	2,209	73.4	802	26.6

Note: The chi-squared test statistic was 21.01 for race/ethnicity ($p < .01$), 68.88 ($p < .01$) for age, and 4.99 ($p < .05$) for gender.

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 2013/14.

Table B18. Characteristics of teachers on one-year and two-year evaluation plans, 2014/15

Characteristic	One-year evaluation plans		Two year evaluation plans	
	Number	Percent	Number	Percent
Total	2,680	71.3	1,079	28.7
Race/ethnicity ^a				
Black	571	73.8	203	26.2
White	1,617	69.7	704	30.3
Other	489	74.0	172	26.0
Age				
Younger than 30	586	91.4	55	8.6
30–49	1,452	67.7	692	32.3
50 and older	642	65.9	332	34.1
Gender				
Male	711	72.9	264	27.1
Female	1,968	70.7	815	29.3

Note: The chi-squared test statistic was 7.65 for race/ethnicity ($p < .05$), 154.00 ($p < .01$) for age, and 1.72 ($p = .19$) for gender. Data on race/ethnicity were missing for three teachers on a one-year evaluation plan, and data on gender were missing for one teacher on a one-year evaluation plan.

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 2014/15.

Notes

1. The analysis conducted for research question 1 did not follow the same teachers over the three years, whereas the analysis for research question 2 did. See box 2 in the main text and appendix A for more information.
2. As of 2015/16, teachers also receive a separate rating for their influence on student learning (for example, student growth scores on the state assessment tests). That rating is also part of the plan determination.
3. These percentages were calculated using the number of teachers with summative performance ratings within each racial/ethnic group as the denominator, which allows for comparisons across racial/ethnic groups relative to each group's size.
4. Although this is considered a population, statistical tests were performed to highlight the important differences.
5. Teachers with a rating of exemplary were included in the analysis if their rating declined from one year to the next; however, if their rating remained at exemplary, they were not included as having improved because exemplary is the highest rating.

References

- 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/?id=EJ889247>
- Curtis, R. (2012). *Building it together: The design and implementation of Hillsborough County public schools' teacher evaluation system*. Washington, DC: The Aspen Institute. <http://eric.ed.gov/?id=ED532207>
- Every Student Succeeds Act, Public Law 114–95, 114th Cong., 1st sess. (December 2015). Retrieved January 25, 2016, from <https://www.congress.gov/bill/114th-congress/senate-bill/1177?q=%7B%22search%22%3A%5B%22every+student+succeeds+act%22%5D%7D&resultIndex=1>.
- 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. <http://epa.sagepub.com/content/22/2/129.abstract>.
- Hamilton, L. C. (1990). *Modern data analysis: A first course in applied statistics*. Belmont, CA: Brooks/Cole Publishing Company.
- 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/?id=EJ1018910>
- Harris, D., & Sass, T. (2009). *What makes for a good teacher and who can tell?* National Center for Analysis of Longitudinal Data in Education Research. Washington, DC: CALDER Urban Institute. 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>.
- Ho, A. D., & Kane, T. J. (2013). *The reliability of classroom observations by school personnel*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved August 28, 2015, from http://www.metproject.org/downloads/MET_Reliability%20of%20Classroom%20Observations_Research%20Paper.pdf.
- 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/?id=EJ1024230>
- Riordan, J., Lacireno-Paquet, N., Shakman, K., Bocala, C., & Chang, Q. (2015). *Redesigning teacher evaluation: Lessons from a pilot implementation* (REL 2015–030). 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://eric.ed.gov/?id=ED552484>

Sartain, L., Stoelinga, S. R., & Brown, E. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. Chicago, IL: Consortium on Chicago School Research, University of Chicago Urban Education Institute. Retrieved January 23, 2012, from <http://ccsr.uchicago.edu/publications/Teacher%20Eval%20Report%20FINAL.pdf>.

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