



Educator evaluation data:

Student growth percentiles, race/ethnicity,
gender, and professional teaching status

June 2015



This document was prepared by the
Massachusetts Department of Elementary and Secondary Education
Mitchell D. Chester, Ed.D.
Commissioner

The Massachusetts Department of Elementary and Secondary Education, an affirmative action employer, is committed to ensuring that all of its programs and facilities are accessible to all members of the public. We do not discriminate on the basis of age, color, disability, national origin, race, religion, sex, gender identity, or sexual orientation.

Inquiries regarding the Department's compliance with Title IX and other civil rights laws may be directed to the Human Resources Director, 75 Pleasant St., Malden, MA 02148-4906. Phone: 781-338-6105.

© 2015 Massachusetts Department of Elementary and Secondary Education
Permission is hereby granted to copy any or all parts of this document for non-commercial educational purposes. Please credit the "Massachusetts Department of Elementary and Secondary Education."

This document printed on recycled paper

Massachusetts Department of Elementary and Secondary Education
75 Pleasant Street, Malden, MA 02148-4906
Phone 781-338-3000 TTY: N.E.T. Relay 800-439-2370
www.doe.mass.edu



Table of Contents

Executive summary..... i

Background 1

Data and methodology..... 2

Findings: Student growth percentiles 3

Findings: Race/ethnicity 7

Findings: Gender..... 9

Findings: Professional teaching status 11

Conclusion..... 15

Executive summary

In November 2013, the Massachusetts Department of Elementary and Secondary Education (ESE) released the first set of summative performance ratings under the state’s new educator evaluation system. The ratings included educators in the 234 Race to the Top districts required to implement the new regulations and evaluate at least half of their educators in the 2012–13 school year. Ultimately, 37,940 educators were evaluated in 2012–13 through the Commonwealth’s new system, representing 62 percent of the 61,441 educators in the districts that met the criteria to be evaluated and 43 percent of educators statewide.

This past November, in 2014, the second set of evaluation ratings produced under the state framework was released. The ratings included not only Race to the Top districts, but also districts that did not participate in Race to the Top, which were required to evaluate at least half of their educators for the first time in the 2013-14 school year. In all, 71,675 educators in 372 districts were evaluated using systems aligned to the new state framework.

This year’s report builds upon last year’s report by comparing two years of evaluation data and by revisiting last year’s key findings with an expanded set of data from districts both in their first and second year of reporting. Similar to last year, this report shows how the summative performance rating relates to one measure of impact on student learning, the MCAS median student growth percentile; and disaggregates the overall performance ratings by race and gender.

A primary purpose for conducting this analysis over time is to promote continuous learning and improvement, a goal of the educator evaluation system itself. By examining the state’s early evaluation data, we can better understand the initial implementation of the new system and provide information to help districts improve their continued implementation. This report also helps support two goals of the educator evaluation system: placing student learning at the center and setting a high bar for professional teaching status.

Key findings include:

- **Similar to last year’s findings, teachers¹ rated Exemplary in the summative performance rating were more likely than other teachers to have achieved high student academic growth, and teachers rated Needs Improvement or Unsatisfactory were more likely than other teachers to have produced low student academic growth.**
 - Fewer than 9 percent of teachers rated as “Exemplary” had a median student growth percentile (SGP) below 35.5 in Mathematics, as compared to 51 percent of teachers rated Unsatisfactory. Conversely, 34 percent of teachers rated as Exemplary had a median SGP above 64.5 in Mathematics, versus 3 percent of teachers rated Unsatisfactory.
 - Teachers rated as Exemplary in the summative performance rating had an average median student growth percentile of 56.2 in English language arts and 58.1 in

¹ Throughout this report we use the term “educators” to mean all educators that must be evaluated according to state regulation, including classroom teachers, specialized instructional support personnel, principals, and others. “Teachers” refers just to classroom teachers.

mathematics, as compared to 45.5 and 34.9 respectively for teachers rated Unsatisfactory.

- **As in 2012-13, the distribution of ratings for educators of color is more disperse than the distribution for white educators.**
 - Looking at all types of educators, 8.0 percent of white educators received an Exemplary rating, versus 9.5 percent of African Americans and 10.6 percent of Hispanics and Latinos. Likewise, 4.5 percent of white educators were rated as Needs Improvement and 0.5 percent Unsatisfactory, versus 9.2 and 1.5 percent of African Americans and 8.4 and 0.7 percent of Hispanics and Latinos, respectively.
- **Again this year, female educators were more likely than males to receive high summative performance ratings and less likely to receive low ratings.**
 - Statewide, 8.8 percent of all female educators were rated as Exemplary, versus 5.9 percent of males. Similarly, 4.1 percent of female educators were rated as Needs Improvement and 0.4 Unsatisfactory, versus 7.1 and 0.8 percent of male educators, respectively.
- **Teachers without professional teaching status (PTS, or tenure) were more likely to receive low ratings and continue to be more likely to be evaluated than PTS teachers.**
 - Statewide, 3.5 percent of non-PTS teachers were rated as Exemplary, as compared to 8.3 percent of PTS teachers. Non-PTS teachers were also three times more likely than PTS teachers to receive a rating of Needs Improvement (10.4 percent versus 3.1 percent).
 - 86.9 percent of PTS teachers eligible to be evaluated in 2013–14 were evaluated, as compared with 93.4 percent of non-PTS teachers who were evaluated.

The data in this report should be considered in light of several important methodological notes.

- Data from the 2012–13 school year represent the first year of large-scale implementation of the educator evaluation system. Only Race to the Top districts were required to implement the new system that year; those districts were required to evaluate at least 50 percent of their educators. Thus, the 2012-13 data on the summative performance ratings comes only from the 37,940 educators in Race to the Top districts who were rated in 2012–13. The 2013–14 summative ratings represent a greatly expanded data set. For the first time, it includes both Race to the Top and non-Race to the Top Districts. However, it is important to note that non-Race to the Top districts were only required to evaluate at least half of their educators in the 2013-14 school year.
- The educators evaluated in 2012–13 are not a random or representative sample of all educators, but rather are representative of those educators in Race to the Top districts who districts chose to evaluate in the first year of implementation. The educators evaluated in 2013-14 in Race to the Top Districts should be a representative sample of all

educators. However, in non-Race to the Top districts the data are representative of those educators who districts chose to evaluate in the first year of implementation.

- Data on the distribution of individual ratings within districts is suppressed when the number of educators in a group is fewer than six or publishing the data would compromise the confidentiality of individual educators' ratings (for instance, when all educators or all but one within a district have the same rating).

Background

On June 28, 2011, the Massachusetts Board of Elementary and Secondary Education adopted new regulations to guide the evaluation of all educators serving in positions requiring a license: teachers, principals, superintendents, and other administrators. The new regulations were based in large part on recommendations from a 40-member statewide task force charged by the Board of Elementary and Secondary Education with developing a new framework for educator evaluation in Massachusetts.

The educator evaluation framework described in the new regulations was explicitly developed to support the following goals:

- Promote growth and development of leaders and teachers,
- Place student learning at the center, using multiple measures of student learning, growth and achievement,
- Recognize excellence in teaching and leading,
- Set a high bar for professional teaching status, and
- Shorten timelines for improvement.

The regulations specify several key elements of the evaluation process. All educators engage in a five-step evaluation cycle that includes self-assessment; analysis, goal setting, and plan development; implementation of the plan; a formative assessment/evaluation; and a summative evaluation. Throughout this process, three categories of evidence are collected: multiple measures of student learning, growth, and achievement, including statewide assessment data (e.g., MCAS) where available; judgment based on observations, including unannounced observations; and additional evidence relating to performance.

Ultimately, educators receive two ratings: a summative performance rating related to their performance on the statewide standards of effective practice, and a rating of their impact on student learning. The summative performance rating is categorized into four levels of performance (Exemplary, Proficient, Needs Improvement, and Unsatisfactory) and is composed of ratings on the four standards of effective teaching or administrative leadership defined in state regulation. The impact on student learning is categorized as high, moderate, or low and is based on trends and patterns in student learning, growth and achievement that include state assessment data where applicable and data from local common measures.² In 2012–13 and 2013–14, the years to which these results pertain, districts were required to issue a summative performance rating only. The student impact rating will not begin to be issued until the end of the 2015–2016 school year.

² More information about the educator evaluation framework is available at <http://www.doe.mass.edu/eval/>.

Data and methodology

In November 2013, the Massachusetts Department of Elementary and Secondary Education (ESE) released initial statewide data on the distribution of educator evaluation ratings among the 37,940 educators³ evaluated in 2012–13. In November 2014, ESE released the second year of statewide data, with a total of 71,675 educators evaluated in 2013–14. This year’s report builds upon last year’s report by comparing two years of evaluation data and by revisiting last year’s key findings with an expanded set of data from districts both in their first and second years of reporting. Similar to last year, this report will also show how the summative performance rating relates to one measure of impact on student learning, the MCAS median student growth percentile; and disaggregates the summative performance ratings by race and gender.

Table 1: Percent of teachers statewide in each SGP growth category, by summative performance rating

Demographic group	2012–2013	2013–2014
% exemplary	7.4	8.1
% proficient	85.2	86.5
% needs improvement	6.8	4.8
% unsatisfactory	0.7	0.5
# of educators evaluated	37,940	71,765

To conduct these analyses, we relied upon evaluation ratings data reported to the state through the Education Personnel Information Management System (EPIMS), the statewide system for collecting demographic and work assignment data on educators. We also used the Student Course Schedule (SCS) data, a separate state data collection, to determine which teachers were assigned to which students. This allowed us to calculate how much improvement each teacher’s students made on statewide assessments.

The data presented in this report are from both the 2012–13 school and 2013-14 school years. The 2012-13 data represents the first year of large-scale implementation of the educator evaluation system. Only the 234 Race to the Top districts were required to implement the new system that year; those districts were required to evaluate at least 50 percent of their teachers. Thus, the data on the summative performance ratings comes from the 37,940 educators in Race to the Top districts rated in 2012–13. This represents 62 percent of the 61,441 educators in those districts and 43 percent of educators statewide in that year. The 2013-14 summative ratings represent a greatly expanded data set. It includes not only Race to the Top districts, but also for the first time non-Race to the Top Districts. However, it is important to note that non-Race to the Top districts were only required to evaluate at least half of their educators in the 2013-14 school year. In all, 71,675 educators were rated, representing 81.5 percent of educators statewide in that year.

³ Throughout this report we use the term “educators” to mean all educators that must be evaluated according to state regulation, including classroom teachers, specialized instructional support personnel, principals, and others. “Teachers” refers just to classroom teachers.

The educators evaluated in 2012–13 are not a random or representative sample of all educators, but rather are representative of those educators in Race to the Top districts who districts chose to evaluate in the first year of implementation. For instance, many districts selected to focus first on evaluating their non-professional teaching status (non-tenured) educators. Indeed, 82 percent of non-PTS teachers were evaluated in 2012-2013, versus 65.8 percent of those with professional teaching status.

The 2013-14 data represents a larger sample, especially in Race to the Top districts, where 89.6 percent of educators were evaluated. However, in non-Race to the Top districts, the educators evaluated are not a random or representative sample of educators. Similar to last year, districts in their first year of implementation targeted supports to newer teachers: 91.2 percent of teachers identified as not having Professional Teacher Status were evaluated, as compared to 68.6 percent of teachers identified as having Professional Teacher Status.

To examine how the summative performance rating relates to student improvement, we examined the data on student growth percentiles (SGPs), which measure a student's improvement from one year to the next on state assessments relative to other students with similar test score histories. We calculate a student growth percentile for each student and then find the median SGP for the students taught by each teacher.⁴ Only teachers who had at least 20 students with available student growth percentile data are included in this analysis. We also only attribute student assessment data to teachers for whom they are directly relevant. For instance, for middle school mathematics teachers, we include their students' SGP in mathematics but not English language arts. As a result, data on student growth percentiles are only available for approximately 10 percent of the educators that received a summative performance rating in 2012–13.

Educators in Massachusetts are accustomed to thinking of the definition of moderate growth for schools or districts as a median student growth percentile between 40 and 60. However, teachers typically have smaller numbers of students contributing to their SGP than schools or districts do. Thus in this analysis we expanded the definition of moderate to include median SGPs between 35.5 and 64.5 in order to account for the greater variability of the measure at the teacher level.⁵

As part of this report, we are also publishing district-level disaggregations of the summative performance ratings by race/ethnicity and gender. In order to protect educators' confidentiality, data are suppressed for groups of fewer than six educators and when all educators or all but one within a group received the same rating. Further, most of Massachusetts' educators of color are concentrated in a small number of districts. In the accompanying district report, we only show disaggregated ratings by district for educators of color in districts with sufficient numbers of those educators.

Findings: Student growth percentiles

Our first analysis compares the summative performance ratings, which are based on professional judgment and a robust evidentiary base, against the student growth measure, which is based on improvement on statewide assessments. If the two generate similar results, this is an indication

⁴ More information on student growth percentiles is available at <http://www.doe.mass.edu/mcas/growth/>.

⁵ More information on the use of student growth percentiles in educator evaluation is available at <http://www.doe.mass.edu/eval/ddm/GrowthPercentiles.pdf>.

that the summative performance rating is related to improved student outcomes. If the two are different, this could signal to the state and districts that additional support or training for evaluators is needed to ensure that ratings are appropriately calibrated.

Table 2: Percentage of teachers statewide in each SGP growth category, by summative performance rating

Summative Performance Rating	English Language Arts					Mathematics				
	Low 0-35 SGP	Moderate 35.5-64.5 SGP	High 65-99 SGP	N	% of total	Low 0-35 SGP	Moderate 35.5-64.5 SGP	High 65-99 SGP	N	% of total
2013-14										
Exemplary	9.4%	58.6%	32.0%	694	9.4%	8.6%	57.0%	34.5%	537	8.1%
Proficient	16.1%	66.4%	17.5%	6,321	86.0%	17.6%	61.1%	21.3%	5,721	86.5%
Needs Improvement	20.5%	71.0%	8.5%	307	4.2%	29.8%	56.7%	13.5%	319	4.8%
Unsatisfactory	23.3%	66.7%	10.0%	30	0.4%	51.3%	46.2%	2.6%	39	0.6%
2012-13										
Exemplary	8.5%	58.4%	33.1%	317	8.0%	10.8%	50.2%	39.0%	231	6.5%
Proficient	15.5%	64.8%	19.7%	3,329	84.2%	16.7%	60.3%	23.0%	3,015	84.8%
Needs Improvement	28.9%	59.3%	11.9%	270	6.8%	29.2%	56.6%	14.2%	281	7.9%
Unsatisfactory	40.5%	54.1%	5.4%	37	0.9%	39.3%	50.0%	10.7%	28	0.8%

Table 2 breaks down teachers’ median student growth percentile data into three categories: low growth (median SGP of 0 to 35), moderate (median SGP between 35.5 and 64.5), and high (median SGP of 65 to 99). It then shows, for a given summative performance rating, what percentage of teachers at that rating exhibited a low, moderate, or high impact on student learning as measured by the student growth percentile.

For instance, among teachers rated Exemplary, all but 9.4% percent had median English language arts SGPs in the moderate (58.6 percent) or high (32 percent) category in 2013-14. Similarly, all but 8.6 percent had median mathematics SGPs considered moderate or high. This pattern is consistent with the 2012-13 results. At the other end of the spectrum, in English language arts, 23.3 percent of teachers rated unsatisfactory had low median SGPs in 2013-14, which is down from 40.5 percent in 2012-13. In both years, relatively few had high median SGPs. It should be noted that the number of teachers rated “Unsatisfactory” that can be linked to a median SGP is relatively small. As a result, relatively large percentage changes may reflect small shifts in numbers. In mathematics, 51.3 percent of the teachers rated unsatisfactory had low median SGPs in 2013-14, up 12 percentage points from last year; and less than 3 percent had high median SGPs.

Table 3: Percentage of teachers in each SGP growth category, by summative performance rating, urban districts only

Summative Performance Rating	English Language Arts					Mathematics				
	Low 0-35 SGP	Moderate 35.5-64.5 SGP	High 65-99 SGP	N	% of total	Low 0-35 SGP	Moderate 35.5-64.5 SGP	High 65-99 SGP	N	% of total
2013-14										
Exemplary	10.2%	51.1%	38.7%	235	10.3%	12.7%	51.0%	36.3%	157	8.6%
Proficient	19.2%	63.9%	16.9%	1868	81.7%	23.90%	55.90%	20.20%	1,501	81.8%
Needs Improvement	25.50%	67.30%	7.30%	165	7.2%	37.40%	52.30%	10.30%	155	8.4%
Unsatisfactory	27.80%	61.10%	11.10%	18	0.8%	65.20%	30.40%	4.30%	23	1.3%
2012-13										
Exemplary	12.90%	56.10%	31.10%	132	8.4%	15.40%	50.50%	34.10%	91	8.4%
Proficient	20.80%	61.70%	17.40%	1,238	79.7%	22.70%	56.30%	21.00%	1,075	80.5%
Needs Improvement	35.50%	57.40%	7.10%	169	10.9%	38.10%	50.30%	11.60%	155	8.4%
Unsatisfactory	66.70%	33.30%	0.00%	37	1.0%	57.10%	35.70%	7.10%	14	1.0%

Table 3 shows the same breakdown, but just for teachers in the 24 urban districts in 2013-14 and 2012-13⁶. The patterns in these districts are generally similar to the statewide patterns. However, again this year, urban teachers rated Exemplary are somewhat more likely to have high SGPs and those rated Unsatisfactory are more likely to have low SGPs. Though the N size in the Unsatisfactory rating category is small, just two teachers rated Unsatisfactory had a high median SGP in English language arts and only one did in mathematics.

A different way to look at these same data is to calculate the average median student growth percentile for educators in each summative performance rating category. For example, we calculate the median SGP for each educator rated as Exemplary, then average those SGPs across all Exemplary educators to find the average median SGP for those educators. If the system is working well, the average median SGP should increase as the summative performance rating improves. Table 4 shows these results.

Table 4: Average median SGP for teachers statewide, by summative performance rating, all educators, urban districts only

⁶ The urban districts are: Boston, Brockton, Cambridge, Chelsea, Chicopee, Everett, Fall River, Fitchburg, Framingham, Haverhill, Holyoke, Lawrence, Leominster, Lowell, Lynn, Malden, New Bedford, Pittsfield, Revere, Salem, Somerville, Springfield, Taunton, and Worcester.

Summative Performance Rating	All Teachers				Teachers in Urban Districts			
	Average ELA SGP	N	Average Math SGP	N	Low 0-35 SGP	Moderate 35.5-64.5 SGP	High 65-99 SGP	N
2013-14								
Exemplary	56.2%	694	58.1	537	57.1	235	56.7	157
Proficient	50.3%	6,321	51.1	5,721	48.9	1868	49.1	1,501
Needs Improvement	45.3%	307	44.4	319	43.5	165	40.9	155
Unsatisfactory	45.6%	30	34.9	39	43.6	18	31.9	23
2012-13								
Exemplary	56.7%	317	58.3	231	54.7	132	55.9	91
Proficient	51.1%	3,329	51.7	3,015	49	1,238	49.4	1,075
Needs Improvement	44.7%	270	46.8	281	41.3	169	42.9	155
Unsatisfactory	42.5%	37	43.1	28	34	37	33.2	14

As anticipated, in both years, teachers rated Exemplary had the highest average median SGPs, at 56.2 in English language arts and 58.1 in mathematics in 2013-14. In mathematics, the average median SGP decreases for each performance level in both years, with the lowest SGPs among the teachers rated Unsatisfactory. The pattern is a little different for English language arts. In 2012-13, the average median SGP decreased with each performance level, but in 2013-14 the average median SGP for the Needs Improvement and Unsatisfactory rating were about the same, at 45.3 and 45.6 respectively. Given the small N sizes in the Unsatisfactory rating category, it is difficult to draw conclusions explaining the change from 2012-13 to 2013-14. More years of data will confirm whether this is the beginning of a trend or simply an anomaly. The patterns in urban districts are generally similar, including the similar SGP median ratings for the Needs Improvement and Unsatisfactory rating for English language arts in 2013-14.

Taken together, the findings related to student growth percentiles provide early, suggestive evidence that the system, in most respects, is working as it should. The educators who have been rated the strongest on the basis of professional judgment are also, on average, those who have the strongest impact on student learning. Nonetheless the relationship is not perfect. In 2013-14, between 10 and 13 percent of educators rated as Exemplary had a low impact on student learning as measured by the median student growth percentile, and between 4 and just over 11 percent of educators rated Unsatisfactory had a high impact on student learning, and the average median SGPs at the Needs Improvement and Unsatisfactory ratings for English language arts were nearly the same.

Findings: Race/ethnicity

Our second analysis disaggregates the summative performance ratings by race/ethnicity to examine whether the patterns of ratings are similar across demographic groups. We present findings for all educators and just for teachers, both for all evaluated educators and just for those in the 24 urban districts.

Table 5: Summative performance ratings by race/ethnicity, all educators

Demographic group	Total N	N evaluated	% evaluated	% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All educators							
2013-14	87,923	71,675	81.5	8.1	86.5	4.8	0.5
2012-13	61,441	37,940	61.8	7.4	85.2	6.8	0.7
African-American							
2013-14	2,517	2,008	79.8	9.5	79.8	9.2	1.5
2012-13	2,380	1,677	70.5	10.7	76.6	10.3	2.4
American Indian or Alaskan Native							
2013-14	91	77	84.6	5.2	87.0	6.5	1.3
2012-13	67	47	70.1	6.4	87.2	4.3	2.1
Asian							
2013-14	1,063	869	81.7	10.4	81.5	7.8	0.3
2012-13	797	563	70.6	10.1	80.3	8.5	1.1
Hispanic or Latino							
2013-14	2,235	1,887	84.4	10.6	80.2	8.4	0.7
2012-13	1,926	1,339	69.5	10.0	79.2	9.6	1.1
Multi-race							
2013-14	358	292	81.6	7.5	84.6	7.2	0.7
2012-13	260	173	66.5	5.8	84.4	8.1	1.7
Native Hawaiian or Pacific Islander							
2013-14	65	56	86.2	3.6	87.5	7.1	1.8
2012-13	37	29	78.4	10.3	69.0	20.7	0.0
White							
2013-14	81,594	66,486	81.5	8.0	87.0	4.5	0.5
2012-13	55,974	34,112	60.9	7.1	85.9	6.5	0.6

Table 5 shows the statewide breakdown of the summative performance ratings by race and ethnicity. Overall, 81.5 percent of educators were evaluated in the second year of implementation, an increase of 19.7 percentage points from the first year. Unlike in 2012-13, where a higher percentage of educators of color were evaluated as compared to white educators, the figures in 2013-15 show a more uniform distribution, ranging from 79.8 for African-American educators, to 86.2 for Native Hawaiian or Pacific Islanders.

The distribution of ratings for educators of color is again wider than it is for the state as a whole. For instance, 9.5 percent of African-American educators were rated Exemplary, as compared to 8.1 percent overall, and 10.7 percent were rated below Proficient, as compared to 5.3 percent overall. We see similarly wide distributions for Asian, and Hispanic or Latino. For Native Hawaiian or Pacific Islander, American Indian/Alaskan Native, and multi-race educators the

percents of educators rated Exemplary were lower in comparison to the state, while the percentagerated below Proficient were higher. As in 2012-13, white educators show patterns similar to the state as a whole.

Table 6: Summative performance ratings by race/ethnicity, all educators, urban districts only

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All urban educators							
2013-14	26,790	22,658	84.6	9.3	82.4	7.4	0.9
2012-13	25,272	16,200	64.1	9.0	80.6	9.3	1.1
African-American							
2013-14	1,869	1,489	79.7	9.9	80.1	8.5	1.6
2012-13	1,949	1,413	72.5	10.0	77.5	9.9	2.6
American Indian or Alaskan Native							
2013-14	43	35	81.4	5.7	88.6	5.7	0.0
2012-13	41	32	78.0	9.4	84.4	6.3	0.0
Asian							
2013-14	552	459	83.2	14.2	78.0	7.4	0.4
2012-13	510	380	74.5	11.3	79.5	8.2	1.1
Hispanic or Latino							
2013-14	1,464	1,247	85.2	12.0	78.2	9.1	0.6
2012-13	1,448	1,027	70.9	10.1	79.9	8.6	1.4
Multi-race							
2013-14	124	100	80.6	6.0	84.0	9.0	1.0
2012-13	103	66	64.1	7.6	84.8	6.1	1.5
Native Hawaiian or Pacific Islander							
2013-14	28	24	85.7	8.3	91.7	0.0	0.0
2012-13	19	15	78.9	20.0	60.0	20.0	0.0
White							
2013-14	22,710	19,304	85.0	9.0	83.0	7.2	0.8
2012-13	21,202	13,267	62.6	8.7	81.1	9.3	0.9

Table 6 shows the same breakdown by race/ethnicity, but just for educators in the 24 urban districts. Urban educators show a wider range of ratings than the statewide results, which is not surprising since the majority of the state’s educators of color work in the urban districts. For instance, out of the state’s 2,517 African-American educators in 2013-14, 74 percent of them (1,869) work in urban districts; similarly, urban districts employ 66 percent of the state’s Hispanic or Latino educators and 52 percent of the Asian educators.

Looking next at the breakdowns just for teachers, as opposed to all educators, we see similar patterns once again. Table 7 summarizes these results.

Table 7: Summative performance ratings by race/ethnicity, teachers only

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All teachers							
2013-14	72,518	60,720	83.7	7.6	86.8	5.1	0.6
2012-13	50,729	32,945	64.9	6.9	85.1	7.3	0.7
African-American							
2013-14	1,891	1,629	86.1	8.4	79.4	10.4	1.8
2012-13	1,826	1,387	76.0	10.2	75.6	11.6	2.7
American Indian or Alaskan Native							
2013-14	72	62	86.1	4.8	87.1	6.5	1.6
2012-13	55	38	69.1	5.3	86.8	5.3	2.6
Asian							
2013-14	904	762	84.3	10.0	81.5	8.1	0.4
2012-13	688	519	75.4	9.4	80.5	9.1	1.0
Hispanic or Latino							
2013-14	1,807	1,592	88.1	10.4	80.3	8.5	0.9
2012-13	1,566	1,145	73.1	8.8	79.9	10.0	1.2
Multi-race							
2013-14	292	246	84.2	6.9	85.0	7.3	0.8
2012-13	218	153	70.2	3.9	85.0	9.2	2.0
Native Hawaiian or Pacific Islander							
2013-14	51	45	88.2	4.4	84.4	8.9	2.2
2012-13	30	26	86.7	11.5	65.4	23.1	0.0
White							
2013-14	67,501	56,384	83.5	7.4	87.2	4.8	0.5
2013-13	46,346	29,677	64.0	6.7	85.8	6.9	0.6

Here we see that a larger share of the state’s teachers have been evaluated than educators overall (83.7 percent of those evaluated), almost irrespective of racial/ethnicity group. This is unsurprising since the districts that began their initial implementation in 2013-14 of the new educator evaluation framework prioritized evaluating teachers (versus other staff). The spread across summative performance ratings categories again shows a wider distribution of ratings at both ends of the spectrum among teachers of color as compared to white teachers.

Findings: Gender

Next, we examined the distribution of summative performance ratings by gender, for educators overall and for teachers. In general we find that male educators received lower ratings on average than their female counterparts.

Table 8: Summative performance ratings by gender, all educators

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All educators							
2013-14	87,923	71,675	81.5	8.1	86.5	4.8	0.5
2012-13	61,441	37,940	61.8	7.4	85.2	6.8	0.7
Female							
2013-14	67,099	54,887	81.8	8.8	86.6	4.1	0.4
2012-13	46,804	29,012	62.0	8.0	85.6	5.9	0.6
Male							
2013-14	20,824	16,788	80.6	5.9	86.2	7.1	0.8
2012-13	14,637	8,928	61.0	5.4	83.9	9.6	1.1

Table 9 shows that, as in 2012-13, in 2013-14 female and male educators were about equally likely to receive a summative performance rating. Female educators were more likely than males to be rated as Exemplary (8.8 percent, versus 5.9 percent) and less likely to be rated as Needs Improvement (4.1 percent, versus 7.1 percent) or Unsatisfactory (0.4 percent, versus 0.8 percent), consistent with the data from 2012-13.

Table 9: Summative performance ratings by gender, all educators, urban districts only

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All urban educators							
2013-14	26,790	22,658	84.6	9.3	82.4	7.4	0.9
2012-13	25,272	16,200	64.1	9.0	80.6	9.3	1.1
Female							
2013-14	20,440	17,396	85.1	10.1	82.6	6.5	0.7
2012-13	19,290	12,435	64.5	9.7	81.2	8.2	0.9
Male							
2013-14	6,350	5,262	82.9	6.8	81.8	10.1	1.3
2012-13	5,982	3,765	62.9	6.6	78.8	13.0	1.7

In Table 9, which looks just at the 24 urban districts, male urban educators in 2013-14 were more likely to be rated as Needs Improvement (10.1 percent, versus 6.5 percent of female) and Unsatisfactory (1.3 percent, versus 0.7 percent), and less likely to be rated as Exemplary (6.8 percent, versus 10.1 percent). Again, this replicates patterns established from the 2012-13 data.

Lastly, Table 10 shows the breakdown of ratings by gender just for teachers, as opposed to all educators in both 2012-13 and 2013-14. We again see a similar pattern with male teachers more likely to receive ratings below Proficient and less likely to receive a rating of Exemplary.

Table 10: Summative performance ratings by gender, teachers only

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All teachers							
2013-14	72,518	60,720	83.7	7.6	86.8	5.1	0.6
2012-13	50,729	32,945	64.9	6.9	85.1	7.3	0.7
Female							
2013-14	55,342	46,474	84.0	8.3	87.0	4.3	0.5
2012-13	38,579	25,133	65.1	7.5	85.6	6.4	0.6
Male							
2013-14	17,176	14,246	82.9	5.3	86.2	7.6	0.9
2012-13	12,150	7,812	64.3	5.1	83.5	10.2	1.2

Findings: Professional teacher status

Finally, we examined the distribution of summative performance ratings by professional teaching status (PTS, or tenure). In Massachusetts, teachers, including school librarians, school adjustment counselors, social workers, school nurses, and school psychologists, are typically awarded professional teaching status after three consecutive years of satisfactory service. Without PTS, a teacher is considered probationary and is employed on an annual basis, allowing a district to not renew the teacher’s contract without stating a specific reason. With PTS, the teacher is considered continuously employed unless the district terminates the employment for cause, and dismissal decisions can be appealed. As such, PTS teachers are more experienced than their non-PTS counterparts.

In this first year of statewide implementation, in 2012-13, districts appeared to focus their evaluation efforts first on the teachers for whom they would need to make future tenure decisions. As Table 11 shows, 66 percent of the 33,902 PTS teachers eligible to be evaluated in 2012–13 were actually evaluated that year, as compared with 82 percent of the non-PTS teachers. Further, as compared to non-PTS teachers, PTS teachers were more than twice as likely to be rated Exemplary (7.7 percent versus 3.0 percent) and one-third as likely to be rated as Needs Improvement (4.6 percent versus 13.5 percent). In the second year of implementation, 2013-14, we see the gap between the percent of PTS and non-PTS evaluated move closer together, no doubt due to the requirement that Race to the Top districts fully implement and evaluate all staff in their second year of implementation. Still, there is an gap between the percent of PTS and non-PTS educators in the second year as well. In 2013-14, 86.9 percent of the 51,206 PTS teachers eligible to be evaluated in 2012–13 were actually evaluated that year, as compared with 93.4 percent of the non-PTS teachers. Evidence shows that much like last year, districts implementing in their first year in 2013-14 chose to focus their efforts on non-PTS teachers rather than PTS teachers.

Table 11: Summative performance ratings by professional teaching status

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All teachers	72,518	60,720	83.7	7.6	86.8	5.1	0.6
2013-14	50,729	32,945	64.9	6.9	85.1	7.3	0.7
2012-13							
PTS teachers							
2013-14	51,206	44,516	86.9	8.3	88.2	3.1	0.4
2012-13	33,902	22,302	65.8	7.7	87.1	4.6	0.6
Non-PTS teachers							
2013-14	15,281	14,268	93.4	3.5	85.3	10.4	0.9
2012-13	10,244	8,446	82.4	3.0	82.5	13.5	1.0

Analyzing the professional teaching status findings by race/ethnicity (Table 12) shows similar patterns to the statewide results. Educators of color have more disperse summative performance ratings than white educators do, whether or not they have professional teaching status. Further, within most racial/ethnic groups, PTS educators were more likely to receive Exemplary ratings than their non-PTS counterparts. In 2013-14, 12.2 percent Hispanic or Latino PTS educators were rated Exemplary, as compared to 4.4 percent of non-PTS Hispanic or Latino teachers.

Non-PTS educators of color, however, were more likely to receive Needs Improvement or Unsatisfactory ratings than were PTS educators of color. Looking just at the Needs Improvement category, 7.8 percent of African-American PTS educators received this rating, versus 12.7 percent of African-American non-PTS educators. We see similar patterns for Asian (3.7 percent versus 13.3 percent) and Hispanic or Latino (6.5 percent versus 12 percent) educators. Both of these patterns are consistent with last year's findings.

Comparing Table 12 to Table 5 (which shows the overall statewide breakdown of summative performance ratings by race) demonstrates that PTS teachers within a given racial/ethnic subgroup are similarly likely to receive an Exemplary rating as educators in that subgroup overall. For example, 9.3 percent of African-American PTS teachers were rated Exemplary, nearly equal to the 9.5 of African-American educators overall that received that rating. PTS teachers of color are somewhat more likely to receive Needs Improvement or Unsatisfactory ratings than educators of color statewide, however, in both years.

Table 12: Summative performance ratings by professional teaching status and race/ethnicity

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All teachers							
2013-14	72,518	60,720	83.7	7.6	86.8	5.1	0.6
2012-13	50,729	32,945	64.9	6.9	85.1	7.3	0.7
PTS teachers							
2013-14	51,206	44,516	86.9	8.3	88.2	3.1	0.4
2012-13	33,902	22,302	65.8	7.7	87.1	4.6	0.6
African-American							
2013-14	1,176	1,058	90.0	9.3	81.3	7.8	1.7
2012-13	1,131	945	83.6	10.7	77.5	8.6	3.3
American Indian or Alaskan Native							
2013-14	47	41	87.2	4.9	85.4	7.3	2.4
2012-13	31	21	67.7	4.8	95.2	0.0	0.0
Asian							
2013-14	537	464	86.4	12.5	83.6	3.7	0.2
2012-13	425	329	77.4	9.7	84.2	5.8	0.3
Hispanic or Latino							
2013-14	1153	1033	89.6	12.2	80.5	6.5	0.8
2012-13	947	731	77.2	9.2	82.2	6.8	1.8
Multi-race							
2013-14	189	162	85.7	7.4	90.1	1.9	0.6
2012-13	133	93	69.9	5.4	86.0	7.5	1.1
Native Hawaiian or Pacific Islander							
2013-14	22	19	86.4	10.5	78.9	5.3	5.3
2012-13	12	10	83.3	20.0	60.0	20.0	0.0
White							
2013-14	48,082	41,739	86.8	8.1	88.6	2.8	0.4
2012-13	31,223	20,173	64.6	7.5	87.8	4.2	0.5
Non-PTS teachers							
2013-14	15,281	14,268	93.4	3.5	85.3	10.4	0.9
2012-13	10,244	8,446	82.4	3.0	82.5	13.5	1.0
African-American							
2013-14	460	408	88.7	3.9	81.1	12.7	2.2
2012-13	401	299	74.6	4.3	74.9	19.1	1.7
American Indian or Alaskan Native							
2013-14	20	19	95.0	5.3	89.5	5.3	0.0
2012-13	19	14	73.7	7.1	78.6	14.3	0.0
Asian							
2013-14	258	241	93.4	3.7	82.6	13.3	0.4
2012-13	165	145	87.9	5.5	78.6	13.8	2.1
Hispanic or Latino							
2013-14	467	435	93.1	4.4	82.3	12.0	1.4
2012-13	354	297	83.9	3.4	82.5	13.8	0.3
Multi-race							

2013-14	54	52	96.3	3.8	82.7	13.5	0.0
2012-13	38	35	92.1	0.0	88.6	8.6	2.9
Native Hawaiian or Pacific Islander							
2013-14	26	25	96.2	0.0	88.0	12.0	0.0
2012-13	16	14	87.5	7.1	71.4	21.4	0.0
White							
2013-14	13,996	13,088	93.5	3.5	85.5	10.2	0.8
2012-13	9,251	7,642	82.6	2.9	82.8	13.3	1.0

In a final analysis, we disaggregated the findings by gender for PTS and non-PTS teachers. Table 13 shows these results.

Table 13: Summative performance ratings by professional teaching status and gender

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All teachers							
2013-14	72,518	60,720	83.7	7.6	86.8	5.1	0.6
2012-13	50,729	32,945	64.9	6.9	85.1	7.3	0.7
PTS teachers							
2013-14	51,206	44,516	86.9	8.3	88.2	3.1	0.4
2012-13	33,902	22,302	65.8	7.7	87.1	4.6	0.6
Female							
2013-14	39,370	34,324	87.2	9.1	88.0	2.6	0.3
2012-13	26,095	17,216	66.0	8.3	87.1	4.1	0.5
Male							
2013-14	11,836	10,192	86.1	5.7	89.0	4.5	0.8
2012-13	7,807	5,086	65.1	5.4	87.3	6.2	1.1
Non-PTS teachers							
2013-14	15,281	14,268	93.4	3.5	85.3	10.4	0.9
2012-13	10,244	8,446	82.4	3.0	82.5	13.5	1.0
Female							
2013-14	11,437	10,706	93.6	3.8	86.7	8.8	0.7
2012-13	7,642	6,332	82.9	3.2	84.2	11.7	0.9
Male							
2013-14	3,844	3,562	92.7	2.6	80.9	15.1	1.3
2012-13	2,602	2,114	81.2	2.4	77.2	18.9	1.5

Similar to the findings for gender overall, we see that both for PTS and non-PTS teachers, females are more likely than males to receive Exemplary ratings. Among non-PTS teachers, for instance, 3.8 percent of females were rated Exemplary, versus 2.6 percent of males. At the other end of the spectrum, males were more likely to receive Needs Improvement or Unsatisfactory ratings, whether or not they had professional teaching status.

Conclusion

A primary purpose for conducting this analysis was to promote continuous learning and improvement, a goal of the educator evaluation system itself. By examining the state's early evaluation data, we can better understand the first year and second years of implementation of the new system and provide information to help districts improve their continued implementation.

This preliminary evidence from the first two years of implementation of the new Massachusetts educator evaluation system suggests that implementation is off to a strong start. Most educators who are rated as Proficient or Exemplary on the summative performance rating also exhibit moderate or high growth among their students, as measured by the median student growth percentile. The distribution of summative performance ratings is wider for educators of color than for white educators, and male educators receive lower ratings on average, but the differences are not stark.

These data provide a point of comparison for districts, so they can understand whether the patterns they see in their own evaluation results are typical of those statewide. Where results are unexpected, districts should dig deeper to understand why these results have occurred. For instance, districts that see a larger than average number of discordant ratings (Exemplary educators with low student growth or vice versa) should closely examine their evaluation processes to ensure that the summative performance ratings are appropriately calibrated across evaluators and relative to available student impact data.

These data also underscore the small number of educators of color across the state and their heavy concentration in the urban districts. The statewide Diversity Initiative Task Force has focused attention on this important issue, and ESE is continuing to [implement their recommendations](#) to help increase the diversity of the state's educator workforce.

As more and better data becomes available about the effectiveness of the state's educator workforce, reports such as this will help to shed light on the distribution of effective educators without resorting to proxies such as educational attainment or length of service. Ultimately this will help districts and the state to develop policies and programs to ensure that the most effective teachers serve the students who most need their support.