



Educator evaluation data:

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

April 2014



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.

© 2014 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..... 1

Findings: Student growth percentiles 3

Findings: Race/ethnicity 5

Findings: Gender..... 7

Findings: Professional teaching status 8

Conclusion..... 10

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 report expands upon the November report in two ways: by showing how the summative performance rating relates to one measure of impact on student learning, the MCAS median student growth percentile; and by disaggregating the overall performance ratings by race and gender.

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 of 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:

- **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 achieved low student academic growth.**
 - Less than 10 percent of teachers rated as Exemplary had a median student growth percentile (SGP) below 35.5 in English language arts, as compared to 41 percent of teachers rated Unsatisfactory. Conversely, 33 percent of teachers rated as Exemplary had a median SGP above 64.5 in English language arts, versus 5 percent of teachers rated Unsatisfactory. Median student growth percentiles in mathematics showed similar patterns.
 - Teachers rated as Exemplary in the summative performance rating had an average median student growth percentile of 56.7 in English language arts and 58.3 in mathematics, as compared to 42.5 and 43.1 respectively for teachers rated Unsatisfactory.
- **The distribution of ratings for educators of color is more disperse than the distribution for white educators.**
 - Looking at all types of educators, 7.1 percent of white educators received an Exemplary rating, versus 10.7 percent of African Americans and 10.0 percent of

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.

Hispanics and Latinos. Likewise, 6.5 percent of white educators were rated as Needs Improvement and 0.6 percent Unsatisfactory, versus 10.3 and 2.4 percent of African Americans and 9.6 and 1.1 percent of Hispanics and Latinos, respectively.

- **Female educators were more likely than males to receive high summative performance ratings and less likely to receive low ratings.**
 - Statewide, 8.0 percent of all female educators were rated as Exemplary, versus 5.4 percent of males. Similarly, 5.9 percent of female educators were rated as Needs Improvement and 0.6 Unsatisfactory, versus 9.6 and 1.1 percent of male educators, respectively.
- **Teachers without professional teaching status (PTS, or tenure) were more likely to be evaluated than PTS teachers and were more likely to receive low ratings.**
 - 66 percent of PTS teachers eligible to be evaluated in 2012–13 were actually evaluated, as compared with 82 percent of non-PTS teachers.
 - Statewide, 7.7 percent of PTS teachers were rated as Exemplary, as compared to 3.0 percent of non-PTS teachers. PTS teachers were also one-third as likely to receive a rating of Needs Improvement as non-PTS teachers (4.6 percent versus 13.5 percent).
 - Patterns for professional teaching status by race/ethnicity and gender were similar to the statewide results.

These data should be considered in light of several important methodological notes.

- These data are from the 2012–13 school year, 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 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 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 that 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).
- Most of the educators of color in Massachusetts are concentrated in a small number of districts. Thus, in the accompanying district-level report we can only show disaggregated ratings by district for educators of color in the 19 districts with sufficient numbers of those educators.

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 new 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 (i.e., 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 district-determined measures of student growth that include state assessment data where applicable.² In 2012–13, the year to which these results pertain, the Race to the Top districts were required to issue a summative performance rating only. The student impact rating will not be issued until the end of the 2015–2016 school year.

Data and methodology

In November 2013, the Massachusetts Department of Elementary and Secondary Education (ESE) released statewide data on the distribution of educator evaluation ratings among the

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

37,940 educators³ evaluated in 2012–13. These findings showed that 85.2 percent of educators evaluated that year were rated Proficient and 7.4 percent Exemplary, while 6.8 percent were rated Needs Improvement and 0.7 percent Unsatisfactory. This report expands upon the previous analysis in two ways: by showing how the summative performance rating relates to one measure of impact on student learning, the median student growth percentile; and by disaggregating the overall performance ratings by race/ethnicity and gender.

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 the 2012–13 school year, 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 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 that 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, versus 65.8 percent of those with professional teaching status. As additional data become available in future years, we will be able to determine how representative this initial sample is of educators statewide.

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 that 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.

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.

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

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.

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. For instance, 60 percent of all African-American educators in Massachusetts and 33 percent of all Hispanic or Latino educators work in the Boston Public Schools, as compared to 8 percent of white educators. Thus, in the accompanying district report, we can only show disaggregated ratings by district for educators of color in the 19 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 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 1: Percent 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
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 1 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 8.5 percent had median English language arts SGPs in the moderate (58.4 percent) or high (33.1 percent) category. Similarly, all but 10.8 percent had median mathematics SGPs considered moderate or high. At the other end of the spectrum, in both English language arts and mathematics about 40 percent of the teachers rated unsatisfactory had low median SGPs, and relatively few had high median SGPs.

Table 2: Percent 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
Exemplary	12.9%	56.1%	31.1%	132 (8.4%)	15.4%	50.5%	34.1%	91 (6.8%)
Proficient	20.8%	61.7%	17.4%	1,238 (79.7%)	22.7%	56.3%	21.0%	1,075 (80.5%)
Needs improvement	35.5%	57.4%	7.1%	169 (10.9%)	38.1%	50.3%	11.6%	155 (11.6%)
Unsatisfactory	66.7%	33.3%	0.0%	15 (1.0%)	57.1%	35.7%	7.1%	14 (1.0%)

Table 2 shows the same breakdown, but just for teachers in the 24 urban districts⁵. The patterns in these districts are generally similar to the statewide patterns. However, urban teachers rated Exemplary are somewhat more likely to have high SGPs and those rated Unsatisfactory are substantially more likely to have low SGPs. Indeed, not a single teacher 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 3 shows these results.

Table 3: Average median SGP for teachers statewide, by summative performance rating

Summative performance rating	All teachers		Teachers in urban districts	
	Average ELA SGP	Average mathematics SGP	Average ELA SGP	Average mathematics SGP
Exemplary	56.7 (n=317)	58.3 (n=231)	54.7 (n=132)	55.9 (n=91)
Proficient	51.1 (n=3,329)	51.7 (n=3,015)	49.0 (n=1,238)	49.4 (n=1,075)
Needs improvement	44.7 (n=270)	46.8 (n=281)	41.3 (n=169)	42.9 (n=155)
Unsatisfactory	42.5 (n=37)	43.1 (n=28)	34.0 (n=15)	33.2 (n=14)

⁵ 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.

As anticipated, teachers rated Exemplary had the highest average median SGPs, at 56.7 in English language arts and 58.3 in mathematics. The average median SGP decreases for each performance level, with the lowest SGPs among the teachers rated Unsatisfactory. The patterns in urban districts are similar, though the average median SGP for each summative performance rating category is lower than it is for teachers statewide.

Taken together, the findings related to student growth percentiles provide early, suggestive evidence that the system 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: About 10 percent of educators rated as Exemplary have a low impact on student learning as measured by the median student growth percentile, and between 5 and 11 percent of educators rated Unsatisfactory have a high impact on student learning.

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 4: Summative performance ratings by race/ethnicity, all educators

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All educators	61,441	37,940	61.8	7.4	85.2	6.8	0.7
African-American	2,380	1,677	70.5	10.7	76.6	10.3	2.4
American Indian or Alaskan Native	67	47	70.1	6.4	87.2	4.3	2.1
Asian	797	563	70.6	10.1	80.3	8.5	1.1
Hispanic or Latino	1,926	1,339	69.5	10.0	79.2	9.6	1.1
Multi-race	260	173	66.5	5.8	84.4	8.1	1.7
Native Hawaiian or Pacific Islander	37	29	78.4	10.3	69.0	20.7	0.0
White	55,974	34,112	60.9	7.1	85.9	6.5	0.6

Table 4 shows the statewide breakdown of the summative performance ratings by race and ethnicity. Overall 61.8 percent of educators were evaluated in this first year of implementation of the new system, with a higher percentage of educators of color being evaluated as compared to white educators.

The distribution of ratings for educators of color is wider than it is for the state as a whole. For instance, 10.7 percent of African-American educators were rated Exemplary, as compared to 7.4 percent overall, and 12.7 percent were rated below Proficient, as compared to 7.5 percent overall. We see similarly wide distributions for Asian, Hispanic or Latino, and Native Hawaiian or Pacific Islander educators, while American Indian/Alaskan Native, multi-race, and white educators show patterns similar to the state as a whole.

Table 5: 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	25,272	16,200	64.1	9.0	80.6	9.3	1.1
African-American	1,949	1,413	72.5	10.0	77.5	9.9	2.6
American Indian or Alaskan Native	41	32	78.0	9.4	84.4	6.3	0.0
Asian	510	380	74.5	11.3	79.5	8.2	1.1
Hispanic or Latino	1,448	1,027	70.9	10.1	79.9	8.6	1.4
Multi-race	103	66	64.1	7.6	84.8	6.1	1.5
Native Hawaiian or Pacific Islander	19	15	78.9	20.0	60.0	20.0	0.0
White	21,202	13,267	62.6	8.7	81.1	9.3	0.9

Table 5 shows the same breakdown by race/ethnicity, but just for educators in the 24 urban districts. Urban educators show a wider range ratings than the statewide results, not surprising since the urban districts house the majority of the state’s educators of color. For instance, out of the state’s 2,380 African-American educators, 82 percent of them (1,949) work in urban districts; similarly, urban districts employ 75 percent of the state’s Hispanic or Latino educators and 64 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 6 summarizes these results.

Table 6: 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	50,729	32,945	64.9	6.9	85.1	7.3	0.7
African-American	1,826	1,387	76.0	10.2	75.6	11.6	2.7
American Indian or Alaskan Native	55	38	69.1	5.3	86.8	5.3	2.6
Asian	688	519	75.4	9.4	80.5	9.1	1.0
Hispanic or Latino	1,566	1,145	73.1	8.8	79.9	10.0	1.2
Multi-race	218	153	70.2	3.9	85.0	9.2	2.0
Native Hawaiian or Pacific Islander	30	26	86.7	11.5	65.4	23.1	0.0
White	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 (64.9 percent of those evaluated), almost irrespective of racial/ethnicity group. This is unsurprising since many districts prioritized evaluating teachers (versus other staff) as they

began their initial implementation of the new educator evaluation framework. 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 receive lower ratings on average than their female counterparts.

Table 7: Summative performance ratings by gender, all educators

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All educators	61,441	37,940	61.8	7.4	85.2	6.8	0.7
Female	46,804	29,012	62.0	8.0	85.6	5.9	0.6
Male	14,637	8,928	61.0	5.4	83.9	9.6	1.1

Table 7 shows that female and male educators were about equally likely to receive a summative performance rating during the first year of implementation. Female educators were more likely than males to be rated as Exemplary (8.0 percent, versus 5.4 percent) and less likely to be rated as Needs Improvement (5.9 percent, versus 9.6 percent) or Unsatisfactory (0.6 percent, versus 1.1 percent).

Table 8: 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	25,272	16,200	64.1	9.0	80.6	9.3	1.1
Female	19,290	12,435	64.5	9.7	81.2	8.2	0.9
Male	5,982	3,765	62.9	6.6	78.8	13.0	1.7

In Table 8, which looks just at the 24 urban districts, male urban educators were more likely to be rated as Needs Improvement (13.0 percent, versus 8.2 percent of female) and Unsatisfactory (1.7 percent, versus 0.9 percent), and less likely to be rated as Exemplary (6.6 percent, versus 9.7 percent).

Table 9: Summative performance ratings by gender, teachers only

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All teachers	50,729	32,945	64.9	6.9	85.1	7.3	0.7
Female	38,579	25,133	65.1	7.5	85.6	6.4	0.6
Male	12,150	7,812	64.3	5.1	83.5	10.2	1.2

Lastly, Table 9 shows the breakdown of ratings by gender just for teachers, as opposed to all educators. 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.

Findings: Professional teaching 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, districts appeared to focus their evaluation efforts first on the teachers for whom they will need to make future tenure decisions. As Table 8 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).

Table 8: Summative performance ratings by professional teaching status

Demographic group	Total N	N evaluated	% evaluated	Among those evaluated			
				% Exemplary	% Proficient	% Needs Improvement	% Unsatisfactory
All teachers	50,729	32,945	64.9	6.9	85.1	7.3	0.7
PTS teachers	33,902	22,302	65.8	7.7	87.1	4.6	0.6
Non-PTS teachers	10,244	8,446	82.4	3.0	82.5	13.5	1.0

Breaking down the professional teaching status findings by race/ethnicity (Table 9) 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. 9.2 percent Hispanic or Latino PTS educators were rated Exemplary, as compared to 3.4 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, 8.6 percent of African-American PTS educators received this rating, versus 19.1 percent of African-American non-PTS educators. We see similar patterns for Asian (5.8 percent versus 13.8 percent) and Hispanic or Latino (6.8 percent versus 13.8 percent) educators.

Comparing Table 9 to Table 4 (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, 10.7 percent of African-American PTS teachers were rated Exemplary, equal to the 10.7 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.

Table 9: 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	50,729	32,945	64.9	6.9	85.1	7.3	0.7
PTS teachers	33,902	22,302	65.8	7.7	87.1	4.6	0.6
African-American	1,131	945	83.6	10.7	77.5	8.6	3.3
American Indian or Alaskan Native	31	21	67.7	4.8	95.2	0.0	0.0
Asian	425	329	77.4	9.7	84.2	5.8	0.3
Hispanic or Latino	947	731	77.2	9.2	82.2	6.8	1.8
Multi-race	133	93	69.9	5.4	86.0	7.5	1.1
Native Hawaiian or Pacific Islander	12	10	83.3	20.0	60.0	20.0	0.0
White	31,223	20,173	64.6	7.5	87.8	4.2	0.5
Non-PTS teachers	10,244	8,446	82.4	3.0	82.5	13.5	1.0
African-American	401	299	74.6	4.3	74.9	19.1	1.7
American Indian or Alaskan Native	19	14	73.7	7.1	78.6	14.3	0.0
Asian	165	145	87.9	5.5	78.6	13.8	2.1
Hispanic or Latino	354	297	83.9	3.4	82.5	13.8	0.3
Multi-race	38	35	92.1	0.0	88.6	8.6	2.9
Native Hawaiian or Pacific Islander	16	14	87.5	7.1	71.4	21.4	0.0
White	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 10 shows these results.

Table 10: 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	50,729	32,945	64.9	6.9	85.1	7.3	0.7
<i>PTS teachers</i>	<i>33,902</i>	<i>22,302</i>	<i>65.8</i>	<i>7.7</i>	<i>87.1</i>	<i>4.6</i>	<i>0.6</i>
Female	26,095	17,216	66.0	8.3	87.1	4.1	0.5
Male	7,807	5,086	65.1	5.4	87.3	6.2	1.1
<i>Non-PTS teachers</i>	<i>10,244</i>	<i>8,446</i>	<i>82.4</i>	<i>3.0</i>	<i>82.5</i>	<i>13.5</i>	<i>1.0</i>
Female	7,642	6,332	82.9	3.2	84.2	11.7	0.9
Male	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, 3.2 percent of females were rated Exemplary, versus 2.4 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 of implementation of the new system and provide information to help districts improve their continued implementation.

This preliminary evidence from the first year 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 and may be explained by the fact that the educators rated in this first year of implementation were not a random or representative sample of educators statewide.

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. Indeed, we were only able to provide separate district-level breakouts of summative performance ratings for non-white educator subgroups in 19 districts. The remainder had too few educators of color to be able to preserve their confidentiality in district-level disaggregations. The statewide Diversity Initiative Task Force has focused

attention on this important issue in recent months, and ESE will be working 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.