## DESE logo

Draft White Paper

Replacing Medians with Means When Summarizing MCAS Student Growth Percentiles

Beginning in 2018, the Massachusetts Department of Elementary and Secondary Education (DESE) will switch to using means (averages) instead of medians for official summaries of student growth percentiles. This white paper discusses the reasons for and notable effects of this change.

# Background

DESE annually aggregates a wide variety of student performance variables at the state, district, school, and subgroup levels for use in the [School and District Profiles](http://profiles.doe.mass.edu/). (DESE also aggregates reports at the classroom level on its secure Edwin Analytics website.) These aggregations support state and federal reporting requirements, provide public feedback and support school improvement planning for all students. As part of these efforts, the Department summarizes changes in MCAS scores at the student level with student growth percentiles (SGPs) rolled up to the group level.

A *student growth percentile* measures a student’s progress on MCAS compared to the progress of other students with similar MCAS performance histories. SGPs range from 1 to 99, where higher numbers represent higher relative growth and lower numbers represent lower growth. A student who grew at the 90th percentile had a higher MCAS score than 90% of the students who performed like him or her in the past.

Individual SGPs are aggregated to summarize the growth of subgroups, classes, schools, and districts. Since the introduction of student growth as an assessment metric in 2008, median SGPs have been the main statistic used for these summaries representing the middle value of each group. This paper presents a rationale for changing the primary summary tool for SGPs to means.

Critical findings and recommendations of this report include the following:

* Means are more sensitive and representative than medians in describing group performance, according to recent research
* Means align better with the Department’s guiding philosophy that all students contribute to accountability results.
* The opportune time for the change is now, as the Department transitions to the next-generation MCAS assessments.

**How SGPs Are Calculated for MCAS**

DESE calculates growth percentiles in ELA and mathematics for students in grades 4 through 8 and grade 10. An SGP is provided for each student with a valid score from the prior grade-level test. When available, the two most recent years of prior test data are used. Because there are no MCAS tests at grade 9, SGPs for grade 10 students compare 10th grade test scores with 8th grade test scores in each subject. No growth percentiles are calculated for grade 3 (the first grade of MCAS testing) or for science (because science is tested only in grades 5 and 8 and in high school).

SGPs are calculated using a specialized form of statistical regression called quantile regression with B-spline adjustments. This method ensures that, no matter what past scores students have, they each have an equal opportunity to grow at a high or low percentile.[[1]](#footnote-1)

# Reasons for the Change

The Department is making this change, beginning in 2018, for the following reasons:

1. Studies conducted by Katherine Castellano and Andrew Ho reveal that means are significantly more sensitive and accurate than medians.[[2]](#footnote-2) Castellano and Ho (who was influential in guiding Massachusetts to the student growth percentile metric in 2007 when the state was developing its value-added measure) have written a series of papers comparing the measurement qualities of various growth models. The authors used simulated data as well as operational data from two states to compare the aggregations of several value-added models, including SGPs. Castellano and Ho evaluated each model’s ability to recover estimates of growth in a large population and concluded that, in all instances, means are substantially more efficient and accurate than medians and that the difference between means and medians was greater than the differences between any of the different models commonly used by states. Importantly, means were more accurate than medians when there were changes in the underlying test scale or test program, which is a factor in the MCAS program. Ho and Betebenner - who devised the student growth percentile model - have both recommended that the Department make the switch to improve the accuracy and interpretability of the results.

2. Means align better with the Department’s philosophy that the performance of all students should contribute to accountability results. When aggregating growth with medians, the scores in the middle of the distribution have the greatest influence on the result. Consider Table 1, which shows

two schools with test results for 10 students apiece, created to illustrate the differences that can arise when calculating means and medians.

**Table 1. A comparison of means and medians**

|  |  |
| --- | --- |
| **School 1** | **School 2** |
| **Student** | Score (1-9) | **Student**  | Score (1-9) |
| **A** | 1 | **K** | 5 |
| **B** | 1 | **L** | 5 |
| **C** | 1 | **M** | 5 |
| **D** | 1 | **N** | 5 |
| **E** | 5 | **O** | 5 |
| **F** | 5 | **P** | 5 |
| **G** | 5 | **Q** | 9 |
| **H** | 5 | **R** | 9 |
| **I** | 5 | **S** | 9 |
| **J** | 5 | **T** | 9 |
| **Mean** | **3.4** |  | **6.6** |
| **Median** | **5** |  | **5** |

The students in School 2 collectively scored almost twice as many points on the test as the students in School 1, but because the students at the middle of the distribution in both schools had the same score, both schools had the same median score of 5.

In this example, the scores for students A, B, C, and D do not have an effect on the median score for their school; only the scores of students E, F, G, H, I, and J are contributing to the median. Every student score influences the mean, while only the middle values are represented in a median.

SGPs, like the test scores in this example, are restricted. No matter how much a student learns and how many more questions he or she correctly answers on the MCAS test, the highest obtainable SGP is 99 and the lowest is 1. In a school with the minimum number of 20 students, that means that while no student can influence the mean by more than 5 points, every student does have an influence on the mean.

3. This year is an opportune time to introduce the change as the Department calculates its first growth percentiles based on two years of next-generation results and introduces its new school and district accountability system under ESSA.

In 2018, students in grades 4 through 8 will receive the first SGPs based on both baseline and comparison scores from the next-generation tests introduced in 2017. Changing the method for aggregating growth in 2018 will underscore the fact that growth from the legacy and transition years should not be directly compared to next-generation results without employing statistical transformations and controls.

In 2018, the Department also plans to revise its school and district accountability system. The old Progress and Performance Index, based in part upon median growth scores, will be replaced with a new system with new weights for growth and achievement. Mean growth results can be incorporated into the new system in 2018 without a complicated transition.

# Implications of the change

*Implication #1: Means will be less volatile than medians from year-to-year; groups will be less likely to fluctuate in and out of the moderate growth range*

School-level mean SGPs vary less from year to year than the medians summarizing the same groups. Comparing 2013 school-level math growth to 2014, there were 1,583 schools that received 20 or more SGPs in both years. The standard deviation of school means was 7.6 points and the standard deviation of school medians was 10.9 points using the same student results. Standard deviations measure the dispersion around an average, so it is reasonable to interpret the smaller standard deviations as evidence of less volatility in means than medians.

In another analysis, the Department compared means to medians using 2017 4th grade math SGPs.

One of the notable effects of aggregating with means instead of medians in this and other analyses is that school (or group) averages tend to be closer to 50, whereas medians are more widely distributed. In recent years, roughly 20% of schools have mean SGPs above 60 or below 40, compared to 25% with median SGPs above 60 or below 40.

Figure 1 compares school SGP distributions in two ways: on the left using medians and on the right using means. The standard deviation for school means is almost four points smaller than the standard deviation for school medians.

**Figure 1: 2017 Grade 4 School median and mean SGP distributions with standard deviations**



One way to think about this difference is to recall that 68% of the scores will be within a standard deviation of the mean. So, for SGPs, 68% of the school means will be between 37.9 and 62.5 (50.18 +/-12.33 points) whereas 68% of the medians will be between 34.3 and 66.8 (50.54 +/- 16.28 points).

*Implication #2: Schools with higher medians will have higher means and schools with lower medians will have lower means*

Overall, for MCAS SGPs, school means and medians are very highly correlated, with an R-squared value of 97. This means that, if you graph them together, 97% of the variation in school medians is directly proportional to the variation in school means.

**Figure 2: School means graphed against school medians**



The most extreme differences between medians and means occur when medians are either exceptionally high or exceptionally low. (More information is provided above about the tendency for means to be closer to 50.) Of the 903 schools serving 20 or more 4th grade students in 2017, 21 (2.3%) had a point difference of 10 or more between school means and medians. Figure 3 shows that 31% of school means and school medians were within two points of each other, and 57% were within four points. These differences mean that some schools would be identified as having high growth using medians but moderate growth using means, or vice versa. There are no cases in 2017 where a school would have been identified as high growth using one aggregation and low growth using the other. (Moderate growth is defined by the Department as ranging from 40 to 59.5.)

**Figure 3: Differences between school medians and school means among grade 4 schools in 2017**



*Implication #3: Schools and classrooms with very high or very low median growth will have more moderate mean growth if they have students whose growth is at the other extreme*

Historically, schools and classrooms with extremely high or low medians have also had significantly more students at the other extreme than one would expect under a normal distribution. In other words, teachers or schools with very high classroom growth tend to have an unusually large number of students with very low growth, and vice versa.

There are two scenarios where this phenomenon is most likely to be observed. Being aware of these scenarios may be helpful for those trying to understand and intervene with programs where very high or very low *median* growth has been moderated by the switch to means.

1. *Accelerated programs where a small but significant number of students can’t keep pace.* When SGPs are aggregated using the median, a program’s score is defined by the middle student. As long as half of the class keeps up, growth scores are not influenced by a small handful of students who not only fall behind the rapidly progressing students in the class, but also fall behind students across the state who started the year with similar knowledge, skills and ability. In rough terms, a program fits into this category if 5% of students have SGPs that are 50 points or more below the class median. In programs that formerly had very high medians, but now have moderate means, educators should consider an approach that benefits the students who are struggling to keep pace with the students who are growing at extremely high rates.
2. *Programs that are extremely effective for a small proportion of students, but do not engage the majority*. This scenario (the opposite of the one described above) is also evident in the data, though less common. An analysis of 4th grade students taking the MCAS in 2010 showed 2,012 students attended schools with median growth below the 25th percentile and 152 (7%) had individual growth percentiles higher than 75. The students with the high growth in low growth schools had lower average scaled scores the year before than the other students in the class (3.2 points; p=0.04). This pattern suggests that some of these programs were effective for the small minority of students needing remediation, but did not keep pace with the average needs of the majority of their students.

Figure 4 illustrates the situation described in scenario 1 using 2017 4th grade math results. The graphic displays the SGPs of all students in schools where median SGPs were 75 or higher. (Such schools are in the top 1% in the state.) The average SGP of the students in these schools was 72.02, which is lower than the median in any of their schools. The average is lower is because there is a larger than normal number of students in these schools with very low growth (on the left side of the chart.)

By the definition of a normal distribution, we would expect 2.5% of the students to be more than two standard deviations (~50 SGP points) below the mean or median, but statewide 7% of the students in these schools were more than two standard deviations below their school medians.

###### Figure 4: State-level student growth percentiles in schools with median SGPs above 75



The reference line shows a normal distribution centered around the mean of all students in these schools. The distribution of students is not normal, however, as many more students than expected (~7%) in these high growth schools are showing very low growth while hundreds of other students showed growth in the 80s and 90s in these high growth schools.

1. Betebenner, D. W. 2009. “Norm- and criterion-referenced student growth.” *Educational Measurement: Issues and Practice* 28 (4), 42–51. [↑](#footnote-ref-1)
2. Castellano, K. E., and A. D. Ho. 2015. “Practical differences among aggregate-level conditional status metrics: From median student growth percentiles to value-added models.” *Journal of Educational and Behavioral Statistics* 40 (1): 35-68. [↑](#footnote-ref-2)