

The Massachusetts Association of School Superintendents submits the following proposal in order to provide stability and restore some sense of equity with respect to our state accountability system.

1. Over the last three years students in the Commonwealth have taken three different state exams - MCAS, PARCC [paper based], and PARCC [computer based]. Now, two new forms of the state exam (Next-Generation MCAS [paper based] and Next-Generation MCAS [computer based]) will be given in grades 3 through 8 in the spring of 2017. M.A.S.S. understands that the department has done the best it can to crosswalk the scores for these assessments in order to place schools in levels and assign percentile rankings. However, with many different assessment scores comprising districts' four-year accountability calculation the validity of those calculations are very much in question at least by many of us in the field. Furthermore, regardless of the efforts taken to be able to compare scores/growth/achievement levels on the various exams, not every variable has been taken into account.

For example, student performance across the state has demonstrated that students who take the PARCC exam on paper score higher than those who take the PARCC exam on computers. This phenomenon was not unique to Massachusetts as was noted in the February 3, 2016 online version of *Education Week*. In his article "PARCC Scores Lower for Students Who Took Exams on Computers," Benjamin Harold points out the discrepancies that exist with the PARCC assessment. Harold referenced in that piece that the Illinois State Board of Education recently found that 43 percent of students who took the PARCC English/Language Arts exam on paper scored proficient or above compared with 36 percent of students who took the exam online. This finding is consistent with the evidence derived from an analysis of the testing data here in Massachusetts as well.

Although analysis of the testing data has proven this to be true, this discrepancy is not taken into account by the state for purposes of determining accountability and those scores are compared as if they were of equal weight. Furthermore, in the state's own reporting of Grade 3-8 testing scores from Spring 2016, DESE did not release statewide data for ELA or math because they recognized that there was a demographic selection bias when comparing the districts that chose to administer PARCC and those than administered MCAS. If this concern is great enough to prohibit the release of statewide performance data, why should it be used as a baseline to judge schools and districts?

In addition to this discrepancy between paper and computer-based assessment results, at least 40 schools across the state saw their accountability levels negatively impacted due to opt-out students lowering participation rates. Not only do opt-out students impact participation rates, but this has the added effect of lowering performance scores as we have seen that the vast majority of the students that opt-out are higher achieving students. Thus, achievement in those schools is negatively impacted along with participation.

"Opt-out" rates are only part of the problem with creating an equitable means to calculate accountability that incorporates both participation rates and performance in districts across the state. The instability inherent in the lives of many of our students due to influences such as living in a group home, homelessness, challenge of being in DCF custody, transience, serious mental health issues, and trauma in the home all have a detrimental impact upon participation rates and performance.

Furthermore, many of our districts, particularly in urban areas, have seen an influx of refugees and non-English speaking ethnic groups. These students often struggle to perform on our state assessments due to language barriers. The department must consider waiving the state testing requirement for those ELL students for a three-year period as performance and growth can be assessed through the ACCESS test which is more appropriate for that student demographic to assess acquisition of English proficiency. It is imperative that as we overhaul our accountability system to align with the new

federal requirements and our new exam, we also look to recognize and incorporate these elements that dramatically impact performance in our districts.

Our reason for pointing out the above concerns is to encourage that, while we finalize both the exam and its impact on our accountability system, the department should temporarily suspend the calculation of accountability ratings and then reset accountability when all districts are administering the Next-Generation MCAS.

A temporary pause of accountability determinations only makes sense and follows the precedent set when the original MCAS exam was developed and piloted. The development of any new state assessment is a process that comes with difficulties and no assessment is perfect from the start. The State itself has recognized this fact and pushed the administration of the Next-Generation MCAS exam at the high school level back until 2019. It is our assertion that the same process should be followed with the Next-Generation MCAS exam in grades 3 through 8. If the state doesn't believe that the test will be ready as a state assessment for high school performance until 2019, why should schools and districts be held accountable by that assessment prior to that date? This temporary suspension of the calculation of accountability determinations would allow us to work out any problems in the new exam before accountability attaches for our schools and districts.

Once that process is complete the department should then reset accountability determinations for all schools so that any variables, uncertainties, or problems with having multiple state exams figure into the accountability determination is eliminated, once again providing equity between districts. At that point we would have meaningful assessment data, consistent among all schools, and a system that had been vetted and improved through a thoughtful deliberate process.

We recognize that some challenges with respect to identifying and providing support to underperforming districts as well as identifying the districts performing in the lowest 5% to meet the federal ESSA mandate will need to be addressed. We would assert that the state already has the data to identify those districts and that is unlikely to change dramatically over the next couple of years. Furthermore, Massachusetts is the highest performing state in the country. It is time for us to take the lead once again. Even should the federal government not be receptive to a waiver of accountability while we work implement a new assessment, it is time for our state to take a stand and take the time to get it right rather than just work feverishly to get it done.

Should the Board of Education be unwilling to pause the determination of accountability ratings while we work to implement our new state assessment, then we must at least reset accountability after the spring 2017 administration of Next-Generation MCAS exam. Although resetting the accountability system would preclude using growth as a factor, it is still possible to meet those challenges using the one year of achievement data which would be inherently more reliable than basing determinations on data from the various assessments taken by students over the past three years using multiple testing modes. Growth can once again be factored into the equation in subsequent years.

2. Although it is our strong professional opinion as educators that the calculation of accountability ratings should be temporarily suspended, should the Board of Education be unwilling to take that course of action, M.A.S.S. recommends that the department investigate the development of a calculation to weight the paper vs. computer-based exam scoring for Next-Generation MCAS during this transition from one mode to the other.

This has not been a smooth transition so far. Some districts have made the decision to move immediately to full computer-based testing because, although they know in the short term it will negatively impact their scores compared to those districts that remain with paper-based exams, they feel it will be a benefit in the long run as their students will gain familiarity with that platform. These districts are making a strategic calculation to take a big hit now rather than smaller ones over time so that in upcoming years their scores would be more competitive. This would also have the additional

consequence of higher percentile rankings for that district over time as the year with a lower score would be further back in time; thus it would be given less weight under our current system.

Other schools systems, even if they have the capability to take the MCAS 2.0 on computers, are reluctant to make that move before they have to because of the likely negative impact on scores and they want to delay that drop for as long a possible. Still other districts, that have already taken state assessments on computers, have now elected to go back to paper based exams because their scores dropped. These choices being made are not educational decisions and they do not essentially affect the delivery of educational services to our students one way or the other. These are strategic gamesmanship decisions and calculations which all superintendents must consider due to the nature of this assessment system being in flux for years. This should not be the case.

3. M.A.S.S. feels compelled to point out that, in looking to mitigate detrimental impact to districts of this system in transition the State's notion of "hold harmless" is not a viable solution to this problem. To say that districts will be "held harmless" really has no meaning for those of us in the field for the following reasons:

a. We are not truly held harmless as, if we continue with the current method of calculating accountability, those scores are still factored into our four-year accountability determination. Consequently, those scores continue to follow (harm) us for 4 years and thus it is simply "harm deferred."

b. Even in the current year we are not "held harmless," The harm is in public perception, not our actual accountability rating. That perception is dramatically shaped by the percentile ranking of a school even more so than the accountability "level". Consequently, since the drop in percentile ranking is still shown on the district profile, even though the district is "held harmless," public perception of the district is harmed

In summation, M.A.S.S. recommends temporarily suspending the calculation of accountability ratings for schools and districts until 2019 as we transition to the computer based Next-Generation MCAS state assessment. Should the Board of Education elect not to do that then, at a minimum, we need to reset accountability levels after the administration of the Next-Generation MCAS in the spring of 2017 so that all districts have a level playing field with the same assessment. Taking this action would mean that districts truly were held harmless during this transition.

Furthermore, should the Board be unwilling to pause accountability determinations, we ask that the department develop a method to weight computer versus paper based testing as this will add validity to the system and help spur the transition to a fully computer based system by removing districts' incentives to delay. In the end, all that educators are looking for is an accountability system that gives a fair and accurate picture of the health of a district. M.A.S.S. believes that the steps outlined above would certainly help in moving us in a direction that provides greater clarity and equity in our current accountability system.