

# **A Proposal to Eliminate the Unworkable, Harmful and Unnecessary “DDM Mandate” in Light of Changes to Federal Education Law**

## **Summary**

This paper summarizes the widespread problems associated with the mandate from the Massachusetts Department of Elementary and Secondary Education (DESE) that districts issue “impact on student learning” ratings to educators based on MCAS/PARCC growth scores and “district-determined measures” (DDMs). In light of these problems, and in consideration of the new federal education law (the Every Student Succeeds Act, or ESSA) eliminating the federal requirement to tie educator evaluations to student outcomes, changes to the Massachusetts system are warranted. This paper recommends eliminating DESE’s requirement that districts issue student impact ratings while preserving evidence of student learning in districts’ locally negotiated educator evaluation systems.

## **Background**

In June 2011, the Massachusetts Board of Elementary and Secondary Education (BESE) approved final regulations on educator evaluation.<sup>1</sup> The regulations outline the process by which all educators in the state receive one of four overall performance ratings: exemplary, proficient, needs improvement, or unsatisfactory. These regulations were informed by the report of a 40-member Massachusetts Task Force on the Evaluation of Teachers and Administrators, which included teacher and administrator representatives.<sup>2</sup>

The final evaluation regulations included a section —35.09— that was never seriously considered or approved by the task force. This section requires that every educator receive a separate “impact on student learning” rating of low, moderate or high. This rating is to be based on “student growth percentiles” (SGPs) derived from student scores on the MCAS assessments; student scores on the Massachusetts English Proficiency Assessment; and/or student outcomes on so-called “district-determined measures,” or DDMs.<sup>3</sup> DDMs are locally created or selected measures of student learning and growth that are comparable across grade or subject district-wide.<sup>4</sup>

When these regulations were passed, many questions were raised about the separate student impact rating, including its feasibility, fairness and possible negative effects on teaching and learning. State education officials insisted, however, that the rating was sound; furthermore, they claimed that the rating was necessary to comply with federal mandates under the now-defunct Race to the Top (RTTT) grant program. RTTT required that student outcomes be a “significant factor” in educator evaluation. This federal requirement was further extended through the No Child Left Behind waivers issued by the U.S. Department of Education from 2012 to 2015.

---

<sup>1</sup> <http://www.doe.mass.edu/lawsregs/603cmr35.html>

<sup>2</sup> <http://www.doe.mass.edu/eval/breakthroughframework.pdf>

<sup>3</sup> <http://www.doe.mass.edu/lawsregs/603cmr35.html?section=09>

<sup>4</sup> <http://www.doe.mass.edu/lawsregs/603cmr35.html?section=02>

There was major bipartisan opposition by parents and educators to many aspects of RTTT and the NCLB waivers across the country, much of it against the inappropriate way test scores were to be used in evaluating educators. As public opposition grew, Congress finally passed, and President Obama signed, ESSA in December 2015. ESSA eliminates the federal requirement that educator evaluations be tied to student test outcomes. This is a significant policy shift, tantamount to an admission that the federal government's policy in this regard had been misguided. As a result of ESSA, NCLB waivers end entirely in July 2016 and the federal mandates linking evaluation to student outcomes have already ceased.

To attempt to address districts' concerns about the DDM/impact rating mandate, DESE has delayed implementation timelines several times and issued hundreds of pages of guidance.<sup>5</sup> Yet even though nearly five years have passed since the establishment of the 2011 evaluation regulations, it appears that no district in Massachusetts has a fully operationalized DDM program in place that uses test scores as prescribed to yield impact ratings. This is probably because educators — superintendents, administrators, and teachers alike — have found the DDM/impact rating mandate to be invalid, unreliable and educationally undesirable for the reasons that are outlined in this paper.

The passage of ESSA creates an important opportunity for Massachusetts to reconsider its DDM/impact rating mandate by modifying its educator evaluation regulations. Policymakers should now make the relatively simple fixes to regulations that would relieve districts of this educationally unproductive and bureaucratically onerous requirement. Such changes would allow districts that want to use common assessments to continue to do so for sound educational purposes without corrupting their use through the impact rating process.

### **Why the DDM/Impact Rating Mandate Is Problematic**

Based on research by testing experts and feedback from district/school-level practitioners, there are three major problems associated with the DDM/impact rating mandate:

#### *1. It Is Methodologically Unsound*

The DDM/impact rating program rests on the premise that it is possible to separate an individual educator's contribution to student learning from the many other complex factors outside of the educator's control, all of which shape student learning, and that this can be done in an accurate, objective and nearly scientific manner.

Yet common sense and extensive research show that this premise is flawed, whether an evaluator is looking at student growth percentiles (SGPs) derived from state standardized tests or student growth on DDMs:

- Leading education researchers do not condone the practice of using student test scores to evaluate educators. For example, a recent statement<sup>6</sup> by the esteemed American

---

<sup>5</sup> <http://www.doe.mass.edu/eval/sir/guidance.html>

<sup>6</sup> <http://edr.sagepub.com/content/early/2015/11/10/0013189X15618385.full.pdf+html>

Educational Research Association (AERA) argues that so-called “value-added models,” which are designed to tease out an individual educator’s contribution to student test scores, have numerous technical and psychometric problems and are too flawed to use for educators’ evaluations. The statement notes that SGPs are simply a type of value-added model.

ESE Response: There are no scores, formulas, or algorithms in the MA Framework. The AERA statement, like others of its kind, cautions against “VAM being used to have a high-stakes, dispositive weight in evaluations,” a position ESE fully supports as reflected by the MA Educator Evaluation Framework. Our framework does not include VAM scores. We do believe that median student growth percentiles from state assessments are an important piece of information for educators and evaluators to study when drawing conclusions about an educator’s impact on student learning. The reality is that since median SGPs are only available for educators who teach ELA or math in grades 4-8 (about 12% of educators), they are used as evidence in the evaluation of a very small percentage of the Commonwealth’s educators. This is in stark contrast to the models used in many other states that rely on formulas in which a teacher’s VAM score represents a specific percentage of their overall evaluation “score.”

- Respected education professor Bruce Baker of Rutgers University argues that SGPs are especially flawed as measures of teacher effectiveness. He writes, “SGPs weren’t ever meant to nor are they designed to actually measure the causal effect of teachers or schools on student achievement growth. They are merely descriptive measures of relative growth and include no attempt to control for the plethora of factors one would need to control for when inferring causal effects.”<sup>7</sup>

ESE Response: The MA Framework employs an appropriate amount of caution in its use of SGPs. First off, median SGP is only part of the evidence used in evaluation for 4<sup>th</sup> through 8<sup>th</sup> grade ELA and math teachers who can be matched with at least 20 students with SGPs. Further, the Student Impact Rating is informed by at least two different measures in each of at least two years. The evaluator uses his/her professional judgment to determine a rating, allowing him or her to consider the instructional context and a host of other information about the educator’s impact determined at the local level. The Framework doesn’t place a minimum or expected weight on median SGP; it just requires it to be part of the conversation, where available. Lastly, the MA Framework does not consider SGPs to be measures of teacher effectiveness. SGPs are measures of student growth that, in the MA Framework, are used as one piece of evidence among many to inform a judgment of educator impact.

- Using DDMs to derive educator ratings is equally, if not more, flawed than the SGP measure. Most of the DDMs being proposed are district-, school-, or teacher-created tests that are administered unevenly, under a variety of testing and scoring protocols, and in classrooms that vary dramatically in terms of student composition and students’ prior knowledge and skill levels. To our knowledge, no district in Massachusetts has figured

---

<sup>7</sup> <https://schoolfinance101.wordpress.com/2011/09/02/take-your-sgp-and-vamit-damn-it>

out how to fairly and accurately attribute student performance on these tests to an individual educator.

ESE response: It is true that most districts have brought teams of teachers together to develop DDMs, and we wouldn't want it any other way. It is inaccurate to assume that a teacher-created measure will yield less meaningful information about student learning than a commercial assessment. A related oft-cited concern about DDMs is that teachers lack the skill to develop measures sophisticated enough to use as evidence in the evaluation process. ESE rejects that presumption and notes that educators regularly create assessments that are used in high-stakes decision-making about students (e.g., the final exam that determines whether a student passes a course). The MA Framework is intentionally flexible to encourage the use of a variety of assessment types authentic to the learning process. All MA districts currently have an ESE-approved plan in place to implement the Student Impact Rating, such that all educators in MA will receive these ratings no later than the end of the 2017-18 school year. In fact, over 40 districts have reported to us that they will be ready to report these ratings for at least some educators by the end of the current school year.

## *2. It Is Harmful to Teaching and Learning*

Deriving educator ratings from student test outcomes, as envisioned in DESE's regulations and guidance, impedes effective teaching and learning for the following reasons:

- It requires districts to use student assessments for purposes for which they were never designed or intended. Educators routinely rely on student assessments to inform key decisions about curriculum, instruction and student supports and interventions. This is the proper use of these kinds of assessments: to facilitate good teaching practice and continuous improvement. However, when those same tests are used improperly for the purpose of deriving impact ratings, the integrity of those tests becomes corrupted, generating data of little or no value and hindering teaching and learning.

ESE Response: The purpose of common assessments, or DDMs, is to support strong teaching and learning, not to derive ratings. We hear from teachers all the time how important it is for them to have tools, such as common assessments, to know how much their students are learning. We are not clear how using student outcomes on assessments to inform a Student Impact Rating "corrupts" the assessment, but if the implication is that MA educators would somehow manipulate the process to earn a better rating, we simply have more respect for our educators than that.

- It is biased against educators who serve the highest-need students, such as special education students, who typically show lower achievement and/or growth on standardized tests.

ESE Response: This is inaccurate. By emphasizing the use of growth measures of varied formats (e.g., portfolio assessments, performance assessments, projects), the entire initiative is grounded in the notion that all student assessments must be designed to allow every student an equal opportunity to demonstrate growth. One of the most important outcomes of the work districts have done to date to develop DDMs has been taking stock of the accessibility of existing assessments and refining them as needed to address the needs of all learners. Improvements to local measures that allow educators to accurately assess the growth of all students are essential to ensuring that all students have access to an excellent education.

- It discourages collaboration among educators because it incentivizes educators to compete for the best students and the best test scores.

ESE Response: The claim here seems to be that the Student Impact Rating is somehow a zero-sum game among teachers. There is no reason to suspect that educators will compete for students and scores when it is possible for all students to demonstrate high growth, and thereby all educators to demonstrate high impact. DDMs involve educators coming together to decide not only what content is most important, but also to establish shared expectations for student learning. When a teacher team can be clear about what they want to measure and how much growth they expect from students, it can have much more focused conversations about instruction.

- It narrows instruction to the tests that “count,” often at the expense of creative and engaging lessons and projects.

ESE Response: All assignments, learning tasks, and assessments measure student learning of just a sample of the content of the standards. The task at hand for educator teams is to design measures that assess the most important content and allow students to engage in authentic learning experiences. ESE frequently works with educators who are refining performance tasks and project-based assessments for use as DDMs.

- It demoralizes and drives away good educators by inaccurately judging educator performance through the use of invalid and unreliable measures.

ESE Response: The expectation is that districts are leveraging existing high quality assessments that are well-aligned to content and provide meaningful information to educators about student learning.

### 3. *It Is an Unnecessary and Time-Consuming Distraction*

- With the passage of ESSA, separate impact ratings are no longer necessary to receive federal education dollars; this was a primary rationale for DDMs/impact ratings before the federal law was changed.

ESE Response: The Board did not establish the Student Impact Rating because of federal policy. The Framework was born out of a belief that all educators are striving to answer an essential question: Does my actual impact meet or exceed the impact I aspire to have on my students? The MA Framework, with its two independent but linked ratings, is designed to use a broad evidence base to home in on the intersection of educator practice and impact on student learning to provide all educators with opportunities for growth and development tailored to their individual needs. That said, while ESSA reserves decisions about evaluation systems to the states, it does require districts, and in turn states, to report on the rates at which minority and economically disadvantaged students are taught by ineffective educators. We believe the MA Framework is essential to a statewide definition of effectiveness. In that regard, Student Impact Ratings are a matter of educational equity, ensuring that all districts are paying close attention to how much and in what ways students are demonstrating progress, especially our most at risk students.

- The DDM/impact rating mandate adds no value to the evaluation process. The state regulations — without impact ratings — *already* set forth a comprehensive educator evaluation system that relies on multiple sources of evidence, including evidence of student learning and demonstrated progress toward student learning goals, to spur educator growth and development. Likewise, Massachusetts educators *already* develop and use a wide range of formative and summative assessments to monitor student learning and inform their instructional practices.

ESE Response: Without the Student Impact Rating, the MA Framework is incomplete, measuring teacher inputs without considering impact on student learning. The Student Impact Rating supports focused conversations about whether student learning is meeting shared expectations. The MA Framework sensibly keeps the consequences for educators low; the Student Impact Rating determines the length of the educator's next growth plan.

- The DDM/impact rating mandate imposes an unproductive bureaucratic burden, according to many district leaders and practitioners. Their objection is not directed at locally developed measures of student performance that educators, schools and districts find useful; it is to the unproductive use of these measures to meet the impact rating requirement. Educators are already drowning in red tape and top-down mandates that take precious time and energy away from teaching and learning; the impact rating requirement simply adds to this problem.

ESE Response: Districts that have undertaken this work in earnest are finding value in it. For example, our team has worked closely with the leadership of a number of public schools this year. In one district, for example, as teams of teacher leaders reviewed their common assessments, they have uncovered gaps in student learning they had never perceived before and identified opportunities to refine their assessments to ensure that their higher and lower performing students had the same opportunities to demonstrate growth. This work is too important to the efficacy of our public schools systems to abandon before it fully gets off the ground in all districts. The actual regulatory language affords districts tremendous flexibility in terms of how to set up structures to assess educator impact. In fact, there is very little in terms of “top-down mandates.”

### **A Relatively Simple Solution Will Fix the Problem**

This problem has a relatively simple fix: Eliminate the impact rating mandate while keeping indicators of student learning as a source of evidence in the educator development and evaluation process. This would enable administrators and educators who value the common assessments they have created to continue their work, while relieving them of the unproductive obligation to create new assessments and complex scoring and educator-evaluation metrics that they do not believe improve teaching and learning in their schools and classrooms.