



MEMORANDUM

To: Commissioner Mitchell Chester
From: Tom Gosnell, President, AFT Massachusetts
Date: April 22, 2011
Subject: Proposed Educator Evaluation Regulations

My staff and I have assembled a number of questions relating to your proposed educator evaluation regulations, which we received on Saturday, April 16. If you or your staff could respond in writing to these questions (ideally before the April 27 Board meeting), that would be greatly appreciated. We would also be open to a face-to-face meeting. Thank you very much for your time and consideration.

Questions relating to the grid and three colored zones

We understand that your proposal combines an educator's "summary" or "overall" rating in one of four categories (exemplary, proficient, needs improvement, or unsatisfactory) with a rating of the teacher's impact on student performance in one of three categories (low, moderate, or high), resulting in the educator's placement in one of 12 boxes on a three-by-four grid (per p. 9 of your memo to the Board). Each of the 12 boxes falls into one of three colored zones, with the green zone triggering a self-directed growth plan, the yellow zone triggering a directed growth plan, and the red zone triggering an improvement plan. We have several questions relating to this proposal.

1. There is an apparent discrepancy between the memo and regulations.

The memo breaks down the 12 boxes as follows:

- Four green zone boxes: Exemplary or proficient with moderate or high.
- Six yellow zone boxes: Exemplary or proficient with low; needs improvement or unsatisfactory with moderate or high
- Two red zone boxes: Needs improvement or unsatisfactory with low.

Yet, the regulations (on p. 12) break down the boxes as follows:

- Four green boxes: Same as above
- Four yellow zone boxes (rather than six, as above): Exemplary or proficient with low; needs improvement with moderate or high
- Four red zone boxes (rather than two, as above): ALL unsatisfactory ratings (regardless of impact on student performance), plus needs improvement with low.

Which breakdown is correct?

2. The meaning of the “summary” rating.

Both the memo and regulations refer to the rating of either exemplary, proficient, needs improvement, or unsatisfactory (henceforth, the EPNU rating) as the “summary” or “overall” rating. Further, the regulations require that multiple measures of student learning and growth factor into this rating. Yet, it appears that this rating isn’t really “summary” in nature since it must then be combined with an “impact on student performance” rating of low, moderate, or high (henceforth, the LMH rating). Can you please provide some elaboration on how educators in the field should interpret the EPNU rating? Is it indeed a summary rating, or is it a “measure of educator practice,” as the memo describes it on p. 8? If it is indeed a summary rating that must incorporate multiple measures of student learning and growth, why is the LMH rating needed at all? What’s the rationale for having evaluators come up with two separate sets of ratings?

3. Under the proposed system, an educator rated “exemplary” or “proficient” who has “low impact” on student performance must be placed on a directed growth plan and “may be dismissed at any time” (p. 14 of regulations). Please explain why educators who are rated highly against the performance standards would be placed on a track that is quite punitive in nature. If the evaluator uses his/her professional judgment to arrive at a summary rating of exemplary or proficient—and if this rating is confirmed by the evaluator’s supervisor, as the regulations require—then what is the rationale for *automatically* putting the educator on a punitive track?
4. What happens to teachers in the yellow zone if their combination of ratings changes from one year to the next? For example, take the case of an educator with an EPNU rating of “needs improvement” and a LMH rating of “moderate.” The memo says that yellow zone educators must move to either the green zone or red zone at the end of the directed growth plan (p. 11). And the regulations (on p. 16) say: “An educator who receives a Needs Improvement rating shall receive this rating for one year only. The educator either must earn a higher rating in the following summative review cycle, which shall be for no more than one year, or shall be rated Unsatisfactory.”

But suppose after one year that the educator raises his EPNU rating to “proficient” but his LMH rating falls to “low.” This would leave the educator in the yellow zone, but for different reasons. Would the educator face discipline/possible dismissal, or would the “proficient” EPNU rating trump the “low” LMH rating and allow the educator to remain in the yellow zone for an additional year? Similarly, what happens to a “proficient” or “exemplary” educator who receives a “low” LMH rating for two or more consecutive years?

5. Bottom line: Do consequential decisions such as dismissal flow primarily from the EPNU summary rating, or does the combination of ratings (i.e., the zone you’re in) matter, too? If yes, how?

Questions relating to the “impact on student performance” rating

6. Determining an individual educator’s impact on student performance is a new and untested concept in education policy. Assuming that multiple measures of student learning are used in arriving at this judgment (as the regulations require to a certain degree), will DESE give guidance to districts and evaluators on how to distinguish among high, moderate, and low ratings? How will DESE help ensure reliability and consistency in ratings across districts, schools, and classrooms?
7. The memo (p. 7) says: “For each year of instruction: moderate impact is represented by student learning gains of a year’s growth; growth of less than one year represents low impact; and high impact is represented by growth of more than one year.” How does the DESE define and measure “one year’s growth?” What tools/measures exist to do this?
8. The regulations say that “at least two state or district-wide measures of student learning gains shall be employed at each grade and subject in determining impact on student learning” (p. 17). Since MCAS growth scores are available for less than 20 percent of educators, most educators will be judged based on these “district-wide measures.” Does the DESE have examples of what these measures might be? For example, what measures will be used for foreign language, physical education, and art teachers? If districts don’t have such measures in place, what happens until they do?
9. The Educator Evaluation Task Force made a point to elevate school- and classroom-based measures “set between the educator and evaluator” as important measures of student performance that would be used in a true “multiple measures” system. While these measures are mentioned on p. 15 of the regulations, they get no attention elsewhere. For example, an educator’s goals for student learning must be measured against “either statewide assessments or district-wide assessments” (p. 12 of regulations), while judgments of impact on student learning require “at least two state or district-wide measures of student learning gains” (p. 17 of regulations). How does the DESE reconcile this language favoring state/district measures with the Task Force’s strong emphasis on the need for school- and classroom-based measures? What is the role for school- and

classroom-based measures, including teacher-selected measures, under the Department's proposal?

Questions relating to the use of MCAS growth scores

The regulations place a heavy emphasis on a teacher's median student growth percentile (median SGP) as an indicator of teacher effectiveness. Median SGP is an entirely new concept for teacher evaluation. Several questions arise from this:

10. Both the commissioner and Task Force members have stressed the importance of having at least three years of data before using median SGPs for evaluative purposes, yet the regulations are silent on this issue. Is it DESE's intent that districts must have at least three years of data before considering median SGPs?
11. DESE staff members have indicated that there should be a minimum number of students in a cohort (i.e., a minimum N size) before a median SGP is calculated from that cohort. Will the Department be providing guidance to districts on this matter? If yes, what is that guidance?
12. The research is clear that student assignment has a profound impact on a teacher's "value-added" or "growth" scores. As evaluators consider median SGPs when assigning impact ratings of low, moderate, or high, will districts be required to randomly assign students or at least consider student assignment patterns in their judgments? What will DESE do to ensure that teachers assigned a disproportionate number of poor test-takers or students with academic challenges are not unfairly penalized under this system?
13. The research also is clear that classroom conditions and resources have a sizable impact on student achievement/growth. The Task Force acknowledged this reality in its report, recommending that "conditions and resources needed by the educator to meet state standards should be considered by the evaluator, where appropriate" (p. 18). In determining LMH ratings for educators, will districts be required to consider variations in conditions and resources, such as class size, school climate, the quality of the curriculum and professional development, the availability of textbooks and technology, and the availability of classroom aides and other professional supports?
14. How will transfer students or students with high rates of absenteeism factor into an educator's median SGP score? For example, if a student transfers to a new school in January and takes the MCAS in March, will his score "count" in calculating the educator's median SGP? What about a student who misses several weeks of instruction due to an extended illness or absence?
15. What is DESE's guidance on how a district should determine the "educator of record" for a particular cohort of students who take the MCAS? For example, if teachers work collaboratively in grade-level teams, or if SPED/ESL teachers, paraprofessionals, and

other specialists provide assistance to a classroom teacher, how does the district determine which individual gets the “credit” or “blame” for the student scores?

16. The memo and regulations say that MCAS growth scores must be used where they exist. Given the well-documented challenges of attributing gains to an individual educator, if a district or school wants to consider median SGPs for a grade-level or department-level team *instead of* for individual educators, would that be acceptable under the regulations?

Miscellaneous questions

17. The memo (p. 10) says that an educator with professional teacher status (PTS) must spend “at least three consecutive years in the green zone to be eligible for additional roles, responsibilities, and compensation,” yet the regulations are silent on this matter. Please clarify the Department’s position.
18. The definitions (pp. 2-4) in the regulations relating to self-directed growth, directed growth, and improvement plans neglect to mention the “impact on student performance” component that affects the plan upon which the educator is placed. Is this an oversight?