

MTA's *Reinventing Educator Evaluation*

Answers to Frequently Asked Questions

The MTA released *Reinventing Educator Evaluation: Connecting Professional Practice with Student Learning* on December 21, 2010, fundamentally changing the terms of the debate about how best to evaluate teachers and administrators in support of improved student learning. The comprehensive proposal advocates revamping the system in dozens of ways. The most controversial section calls for including a review of multiple measures of student achievement, one of which would be three-year trends in student MCAS growth scores, where those are available. Notably, the report says that student achievement data may inform the evaluation, but not determine it. In other words, no high-stakes decisions about educators should be made based on student test scores – or on any other single indicator. If there is a discrepancy between measures of student achievement and what the evaluator found through observation and other means, the evaluator is required to find out why. Ultimately, the final evaluation for educators should rely on professional judgment informed by observation and evidence.

How does the current evaluation system work?

Under current state law and regulations, teachers must be evaluated based on seven Principles of Effective Teaching, and administrators must be evaluated based on the Principles of Effective Administrative Leadership. In their first three years, teachers and administrators must be evaluated at least once a year and teachers who attain Professional Teacher Status must be evaluated at least once every two years.

Districts must determine the specific evaluation protocols, rubrics and timelines for implementation through the collective bargaining process. In many districts, observation is the sole or main instrument of evaluation, though a quarter of MTA members surveyed said that “student outcomes” were part of their local evaluation systems.

Why is the system being changed?

State and federal policymakers and many educators believe that the system is flawed, resulting in inadequate evaluations in some districts. Poor evaluations can impede professional growth and that in turn can adversely affect student learning.

Concerns about the system include that evaluations are often superficial (e.g., based on a single observation), are not completed according to schedule (e.g., a survey of Boston teachers found that more than three-quarters of novice teachers had not been evaluated during the prior school year), or are not effective in helping teachers improve (e.g., only a third of the MTA members surveyed agreed that the supervision and evaluation system identified specific professional development to help them improve their practice).

The Department of Elementary and Secondary Education announced in 2010 that it plans to revamp the evaluation system in 2011 and roll out the changes in stages, with all districts using the new system in the 2013-14 school year.

How does this relate to Race to the Top?

The DESE is planning to change the evaluation system in accordance with the state's RTTT grant. However, the DESE plans to require all districts to adopt the new framework, not just those that are participating in RTTT. Among other provisions, the RTTT grant required the establishment of a task force of stakeholders to research and debate the evaluation issue. MTA has been an active participant in those meetings.

Does MTA agree the system should be changed?

Yes. The MTA Board of Directors has voted to support the changes proposed in a recent report released by MTA's Center for Education Policy and Practice, *Reinventing Educator Evaluation: Connecting Professional Practice with Student Learning*. The recommendations in this report stem from research about effective evaluations and discussions with the MTA Board of Directors, the MTA Professional Development Committee, local presidents and policy experts. The report was also informed by two surveys of thousands of MTA members conducted in 2010. While some MTA members report that the evaluation system works well, others express concerns about the kinds of deficiencies described above.

How would students benefit from a better evaluation system?

Educators who are evaluated properly and provided with meaningful feedback and support become better teachers and administrators. Students are the ultimate beneficiaries of these improvements.

While better evaluations can help, the MTA recognizes that they are no silver bullet. Improving the evaluation system is important, but should never be a substitute for society addressing the most significant cause of poor student performance: poverty.

How would the MTA proposal work?

The MTA plan would reduce the number of principles on which educators are evaluated to five and also reduce the number of indicators linked to each principle.

The proposal then calls for a triangulated system for evaluating educator performance relative to these five standards. At each stage, the evaluation would be based on multiple measures. The belief is that a more accurate assessment of educator strengths and weaknesses can be determined by examining educator practice and outcomes using multiple sources of evidence. As indicated by the term "triangulation," the evaluation would have three distinct parts.

- (1) Observation of Practice and Examination of Artifacts: The entry point is at the top of the triangle. The most important measures would be an evaluation of the educator's practice through observation by a trained evaluator, the educator's self-assessment and the evaluator's examination of educator work products (artifacts), such as lesson plans, school schedules or district budgets. Training is essential. Too many supervisors have little or no training in how to conduct effective evaluations.

The other two components of the triangle would be examined in order to "validate" the results of the first measure.

- (2) Measures of Student Learning and Outcomes. These would have to include multiple measures of student learning. For example, they could include district-based subject tests, teacher developed quizzes and tests, student portfolios, projects and performances. Where available, three-year trends in student MCAS growth scores would also be used. According to the

DESE, currently such growth scores can only be calculated for fewer than 17 percent of all teachers. Additional measures would be needed for all teachers, including those for whom growth scores can be calculated.

- (3) Evidence of Professional Contributions. Educators are members of a learning community and must contribute to it. Contributions could include participation in grade-level teams, developing new curricula or joining an advisory committee – among many others. Any review of these contributions would have to take into consideration the educator’s years of experience and other obligations.

The resulting evaluation would lead to one of four possible ratings: Unsatisfactory, Needs Improvement, Proficient or Exemplary. (Note: Discussions are ongoing about whether educators would receive an overall rating in one of these four categories or a rating for each of the five principles on which they are being evaluated.)

The same categories would apply to both teachers and administrators. However, the standards on which they would be evaluated – as well as the indicators, rubrics and procedures – would be based on their differing job functions and responsibilities.

Would the same evaluation procedures be used for all teachers – new teachers and veterans?

No. As under current law, teachers in their first three years would be subject to evaluations at least once a year while those with PTS would be evaluated at least once every two years. But under the MTA plan, peers could, subject to bargaining, be more active participants in the process. For example, teachers who are rated Proficient or Exemplary could be part of a professional learning community and could be used to assess and support similarly rated colleagues, though the final evaluation would still rest with an administrator.

The MTA report also suggests that some districts may want to negotiate more comprehensive Peer Assistance and Review programs under which panels of peers become more deeply involved in assessing the performance of novice teachers and teachers with PTS whose performance is rated Unsatisfactory.

The concept behind involving peers is twofold. First, administrators contend that they simply have too many people to evaluate. By spending time evaluating teachers who have consistently done well in the past, they have less time to focus on those who are new or struggling. Second, high percentages of MTA members report that they value support and insights from respected peers.

What parts of the system would be subject to bargaining?

As under current law, the Board of Elementary and Secondary Education would establish an evaluation framework, including performance standards and other required components of a district’s system. The specific procedures, rubrics and timelines would be bargained locally.

What are student growth scores?

The term “growth score” refers to the change in an individual student’s performance over time. The Massachusetts growth model is based on one implemented in Colorado. Under this model, each student’s change in performance is compared to the change in performance of other students who have a similar test-score history. Those students are referred to as “academic peers.”

For example, student Will Smith has the following scores in MCAS math:

5th grade: 236

6th grade: 242

7th grade: 238

Will's score in seventh grade will be compared to all of the other students in the state who scored 236 in 5th grade and 242 in 6th grade – students with a similar math profile, regardless of where they were taught or their demographic characteristics. If Will's score of 238 in 7th grade was better than the scores of 60 percent of his academic peers, Will's Student Growth Percentile (SGP) score would be 60.

Have growth scores been calculated for Massachusetts teachers yet?

No. The state has produced SGP scores for schools and districts, which can be found on the DESE website. The DESE does not expect to produce such scores for teachers until the summer of 2011. The score for a classroom would be the median of all the growth scores of the students in the class – that is, the point at which half the scores are above and half below.

What are trends in growth scores?

A trend looks at growth scores over multiple years. The MTA plan calls for a minimum of three years of growth scores before a trend can be established. Even then, a meaningful trend cannot be determined if the scores are high one year, low the next, high again in the third year and so on. The DESE wrote that a trend of low growth scores “typically refers to a teacher whose students are in the 39th growth percentile or lower for multiple years.”

How useful are growth scores at identifying effective teachers?

Most researchers and educators – including the National Education Association – say that growth is a much better measure of effectiveness than simply comparing one year's class to the previous year's class or comparing absolute score levels from one district to another. It is well understood that some classes are stronger than others and that, on average, for a lot of reasons unrelated to teacher quality, schools serving low-income students rarely achieve the same score levels as those serving affluent students. Growth models attempt to adjust for that by looking at how much a student's scores improved over a set period of time – regardless of the starting point or absolute levels reached.

Several recent studies find that there is some value to using student growth scores as part of an educator evaluation system – but they are far from perfect. Douglas Staiger, an economics professor at Dartmouth College cited in *The New York Times* on Dec. 26, said of a similar model in New York City: “This information is useful but has to be used with caution. It's that middle ground. It's not useless, but it's not perfect.” A Rand Corporation analysis summarized findings from multiple studies, concluding that the research base is currently insufficient to support the use of these kinds of measures “for high-stakes decisions about individual teachers.”

Because these measures are not foolproof, the MTA proposal states that student growth scores should be one of multiple measures and should *not* be used to make high-stakes decisions about educators. Instead, these measures should be used to validate – or not – conclusions reached based on the evaluator's observations.

What does it mean to “validate” the supervisor's observations?

Under the MTA plan, measures of student learning and outcomes and evidence of professional contributions may inform the evaluation, but not determine it. This may best be explained with an example. Jane Doe is a middle school math teacher. After she goes through the first part of the process – a self-assessment, observation by her supervisor and a review of her work artifacts – her supervisor

determines that her practice is Proficient. Before developing a final rating or a professional growth plan, the supervisor reviews multiple measures of student growth. In her district that means reviewing assessments in 7th-grade math developed by the district, student projects submitted to her school's math fair and three years of data about her students' growth scores. The supervisor then looks at her contributions to the school or district, noting her involvement in the team of educators who revamped the middle school math curriculum. If the supervisor determines that those measures also reflect Proficient performance, that rating is given, and a professional growth plan is developed.

What happens if the other measures don't validate the observation?

The evaluator would then have to figure out why. Let's say Doe's students showed average growth on district-based measures but her students had below average MCAS growth scores – a level more in line with an educator receiving a rating of Needs Improvement. The supervisor would be tasked with resolving the discrepancy, first by meeting with Doe to go over the results and find out her perspective. The reassessment would likely include more observations and a closer look at Doe's curriculum and instructional practices.

Ultimately, human judgment – the supervisor's – would be used to reconcile any differences and render a final evaluation and rating. Education is too complex an enterprise to base evaluations on any single measure. Observations alone are imperfect, as are measures of student growth, self-assessments or any other single component of an evaluation system. The more measures used, the more likely it is the supervisor will develop an accurate picture of the educator's strengths and weaknesses. With that information, it is easier to provide educators with the guidance and professional development they need to meet the goals in their professional growth plans.

After conducting many evaluations, the supervisor would also develop a better picture of the whole school's strengths and weaknesses. That could lead to changes in what textbooks, curriculum guides and professional development strategies are used, among other possible changes.

What happens once the evaluations are completed?

Under the MTA plan, educators with ratings above Unsatisfactory would develop professional growth plans with their supervisors, while those with an Unsatisfactory rating would have improvement plans. As noted earlier, these plans may involve working collaboratively with peers to build on strengths or improve knowledge and skills in a particular area.

As under current law, teachers without PTS who have poor evaluations may be dismissed; no "just cause" must be established. Teachers with PTS who are rated Unsatisfactory would be given support and close supervision to help them improve. If they did not improve, they could be dismissed or demoted within one year.

Would PTS teachers still have "just cause" dismissal rights?

Yes. As under current law, poor performance – as determined by a valid, fairly administered evaluation process that includes an opportunity to improve – would constitute just cause for dismissing a teacher.

How would the ratings be used for educators seeking career advancement?

Under the MTA plan, teachers who are rated Exemplary would be eligible to apply for new positions requiring greater skill and experience and giving them more responsibility. These positions typically offer higher pay.

What is a Peer Assistance and Review program?

Under the MTA plan, a PAR program could be established only if negotiated locally. It would involve a process under which trained Exemplary teachers would be involved in many aspects of the evaluation and support process, including making recommendations that inform decisions about awarding PTS and validating whether a rating of Unsatisfactory is warranted. These panels would advise, but principals would make the final decisions regarding hiring and firing personnel, as under current law.

What process did MTA use to develop this proposal?

The MTA's Center for Education Policy and Practice took the lead in researching the issue, meeting with other stakeholder groups and writing the plan, with input from MTA leadership. A conceptual framework of the plan was approved by the MTA Board of Directors on October 23, 2010, and the board gave the full plan its final approval on December 4, 2010. In addition, the MTA had presentations and received feedback at two meetings with the large urban district presidents and held several webinars to which presidents of the smaller local associations were invited. The MTA also conducted two statewide web surveys that were sent to 43,000 MTA members in April and December on key elements of the proposal, receiving more than 3,000 responses to the first and more than 5,100 responses to the second. The MTA has been sending regular RTTT updates to local presidents and has sought input from MTA field staff and Boston office staff.

MTA leadership felt it was important to put out the MTA plan before the state's Educator Evaluation Task Force made its proposals in order to help shape the debate. While some states and jurisdictions are giving student test scores a specific weight in the evaluation process, such as requiring it to account for 50 percent of the evaluation, the MTA plan says that these scores are just one of many indicators to be reviewed to determine if the educator's practice is leading to the desired student outcomes.

What happens now?

That task force is expected to issue a report to the Commissioner of Education this winter based on input from stakeholders, including the MTA. The DESE is expected to issue proposed regulations in March and invite public comment. The final regulations are expected to be approved by the BESE in May. These deadlines could be extended if deliberations are not completed.

Throughout this process, the MTA will actively work for a multi-pronged approach based on multiple measures, requiring training for evaluators, mandating more support for struggling educators and using student performance data – including MCAS growth scores – appropriately.

This FAQ was produced by MTA Communications and MTA's Center for Education Policy and Practice. A copy of the full report, a policy brief and a press release on MTA's proposal can be downloaded at www.massteacher.org.