



Massachusetts Model System for Educator Evaluation

Part VII: Rating Educator Impact on Student Learning Using District-Determined Measures of Student Learning, Growth and Achievement

August 2012

Massachusetts Department of Elementary and Secondary Education
75 Pleasant Street, Malden, Massachusetts 02148-4906
Phone 781-338-3000 TTY: N.E.T. Relay 800-439-2370
www.doe.mass.edu



This document was prepared by the
Massachusetts Department of Elementary and Secondary Education
Mitchell D. Chester, Ed.D.
Commissioner

Board of Elementary and Secondary Education Members

Ms. Maura Banta, Chair, Melrose
Ms. Beverly Holmes, Vice Chair, Springfield
Dr. Vanessa Calderón-Rosado, Milton
Ms. Harneen Chernow, Jamaica Plain
Mr. Gerald Chertavian, Cambridge
Mr. Matthew Gifford, Chair, Student Advisory Council, Brookline
Dr. Jeff Howard, Reading
Ms. Ruth Kaplan, Brookline
Dr. Dana Mohler-Faria, Bridgewater
Mr. Paul Reville, Secretary of Education, Worcester
Mr. David Roach, Sutton

Mitchell D. Chester, Ed.D., Commissioner and Secretary to the Board

The Massachusetts Department of Elementary and Secondary Education, an affirmative action employer, is committed to ensuring that all of its programs and facilities are accessible to all members of the public. We do not discriminate on the basis of age, color, disability, national origin, race, religion, sex or sexual orientation. Inquiries regarding the Department's compliance with Title IX and other civil rights laws may be directed to the Human Resources Director, 75 Pleasant St., Malden, Massachusetts 02148-4906. Phone: 781-338-6105.

© 2012 Massachusetts Department of Elementary and Secondary Education
*Permission is hereby granted to copy any or all parts of this document for non-commercial educational purposes.
Please credit the "Massachusetts Department of Elementary and Secondary Education."*

This document printed on recycled paper

Massachusetts Department of Elementary and Secondary Education
75 Pleasant Street, Malden, Massachusetts 02148-4906
Phone 781-338-3000 TTY: N.E.T. Relay 800-439-2370
www.doe.mass.edu



Contents

A Letter From the Commissioner	ii
The Massachusetts Model System for Educator Evaluation.....	iii
About This Guide	1
Introduction and Purpose	2
Implementation Timetable.....	5
Getting Started	7
Identifying and Selecting District-Determined Measures	8
Key Criteria	8
Selecting Appropriate Measures of Student Learning	10
Resources	11
Matching Educators with Appropriate Measures	19
Matching Students to Their Educators.....	22
Implementation Timelines and Reporting Requirements for District-Determined Measures	22
Rating Educator Impact on Student Learning.....	24
Defining Student Growth as High, Moderate, or Low	24
Identifying Trends and Patterns in Student Growth	27
Using the Impact Rating.....	29
Looking Ahead: Reporting the Rating of Impact to ESE.....	33
Identifying and Beginning to Address District Capacity and Infrastructure Needs.....	34
Identifying and Addressing District and School Needs	34
Accessing More Resources From ESE.....	35
Planning Collective Bargaining	37
Immediate Next Steps.....	39
Appendix A. What the Regulations Say	A-1
Appendix B. Technical Guide A (District-Determined Measures).....	B-1
Appendix C. Technical Guide B (Rating Educator Impact on Student Learning)	C-1
Appendix D. Looking Ahead: Attribution and Roster Verification	D-1
Appendix E. Educator Evaluation and Collective Bargaining	E-1



Mitchell D. Chester, Ed.D.
Commissioner

A Letter From the Commissioner

Massachusetts Department of Elementary and Secondary Education

75 Pleasant Street, Malden, Massachusetts 02148-4906

Telephone: (781) 338-3000
TTY: N.E.T. Relay 1-800-439-2370

August 10, 2012

Dear Educators and other interested Stakeholders,

I am pleased to present Part VII of the Massachusetts Model System for Educator Evaluation. Since late June, when the Board of Elementary and Secondary Education adopted regulations to improve student learning by overhauling educator evaluation in the Commonwealth, staff here at the Department has been working closely with stakeholders to develop the Model System called for in the regulations. With the help of thoughtful suggestions and candid feedback from a wide range of stakeholders, we have now developed the first six components of the Model System:

- District-Level Planning and Implementation Guide
- School-Level Planning and Implementation Guide
- Guide to Rubrics and Model Rubrics for Superintendent, Administrator, and Teacher
- Model Collective Bargaining Contract Language
- Implementation Guide for Principal Evaluation
- Implementation Guide for Superintendent Evaluation
- Rating Educator Impact of Student Learning Using District—Determined Measures of Student Learning, Growth, and Achievement

I am excited by the promise of Massachusetts' new regulations. Thoughtfully and strategically implemented, they will improve student learning by supporting analytical conversation about teaching and leading that will strengthen professional practice. At the same time, the new regulations provide the opportunity for educators to take charge of their own growth and development by setting individual and group goals related to student learning.

The Members of the State Board and I know that improvement in the quality and effectiveness of educator evaluation will happen only if the Department does the hard work ahead "with the field," not "to the field." To that end, we at the Department need to learn with the field. We will continue to revise and improve the Model System including the Implementation Guides based on what we learn with the field over the next few years. To help us do that, please do not hesitate to send your comments, questions and suggestions to us at EducatorEvaluation@doe.mass.edu. Please also visit the Educator Evaluation webpage at www.doe.mass.edu/edeval/. We will be updating the page regularly.

Please know that you can count on the Department to be an active, engaged partner in the challenging, but critical work ahead.

Sincerely,

Mitchell D. Chester, Ed.D.
Commissioner of Elementary and Secondary Education

The Massachusetts Model System for Educator Evaluation

The Model System is a comprehensive educator evaluation system designed by the Department of Elementary and Secondary Education (ESE), pursuant to the new educator evaluation regulations, 603 CMR 35.00. The following eight-part series was developed to support effective implementation of the regulations by districts and schools across the Commonwealth.

Part I: District-Level Planning and Implementation Guide

This Guide takes district leaders—school committees, superintendents and union leaders—through factors to consider as they decide whether to adopt or adapt the Model System or revise their own evaluation systems to meet the new educator evaluation regulation. The Guide describes the rubrics, tools, resources and model contract language ESE has developed, and describes the system of support ESE is offering. It outlines reporting requirements, as well as the process ESE will use to review district evaluation systems for superintendents, principals, teachers and other licensed staff. Finally, the Guide identifies ways in which district leaders can support effective educator evaluation implementation in the schools.

Part II: School-Level Planning and Implementation Guide

This Guide is designed to support administrators and teachers as they implement teacher evaluations at the school level. The Guide introduces and explains the requirements of the regulation and the principles and priorities that underlie them. It offers guidance, strategies, templates and examples that will support effective implementation of each of the five components of the evaluation cycle: self-assessment; goal setting and Educator Plan development; plan implementation and evidence collection; formative assessment/evaluation; and summative evaluation.

Part III: Guide to Rubrics and Model Rubrics for Superintendent, Administrator, and Teacher

The Guide presents the Model Rubrics and explains their use. The Guide also outlines the process for adapting them.

Part IV: Model Collective Bargaining Contract Language

This section contains the Model Contract that is consistent with the regulation, with model language for teacher evaluation. The Guide will contain model language for administrators represented through collective bargaining by March 15, 2012.

Part V: Implementation Guide for Principal Evaluation

This section details the model process for principal evaluation and includes relevant documents and forms for recording goals, evidence and ratings. The Guide includes resources that principals and superintendents may find helpful, including a school visit protocol.

Part VI: Implementation Guide for Superintendent Evaluation

This section details the model process for superintendent evaluation and includes relevant documents and a form for recording goals, evidence and ratings. The Guide includes resources that school committees and superintendents may find helpful, including a model for effective goal setting.

Part VII: Rating Educator Impact on Student Learning Using District-Determined Measures of Student Learning, Growth and Achievement

This document contain guidance for districts on identifying and using district determined measures of student learning, growth and achievement, and determining ratings of high, moderate, or low for educator impact on student learning.

Part VIII: Using Staff and Student Feedback in the Evaluation Process (May 2013)

Part VIII is scheduled for publication in May 2013. It will contain direction for districts on incorporating student and staff feedback into the educator evaluation process.

About This Guide

Advancing the academic growth of students is the core work of schools. The state's new educator evaluation framework also puts it at the center of the evaluation and development of teachers and administrators by asking districts to identify at least two district-determined measures of student growth for all educators and to use those measures to assess each educator's impact on student learning. This will help support a primary function of the new system, which is to provide timely, useful feedback to teachers to improve their practice and better support student learning.

The use of common measures of student performance for all educators also provides a groundbreaking opportunity to better understand student knowledge and learning patterns throughout the Commonwealth. What does it look like to achieve excellence in musical expression, or to excel in reading comprehension in a foreign language? Are there discrepancies in Grade 1 growth rates within a school or across a district, and what types of instruction support higher growth rates? Not only can multiple measures of learning better inform the instructional practice of individual educators, they have the potential to inform our overall understanding of *how* students learn and excel throughout the educational continuum.

Identifying credible, instructionally useful measures of student growth will require districts to think deeply about their instructional priorities and how those are reflected in their assessment strategies. Selecting district-determined measures gives districts a long-sought opportunity to broaden the range of *what* knowledge and skills they assess and *how* they assess learning.

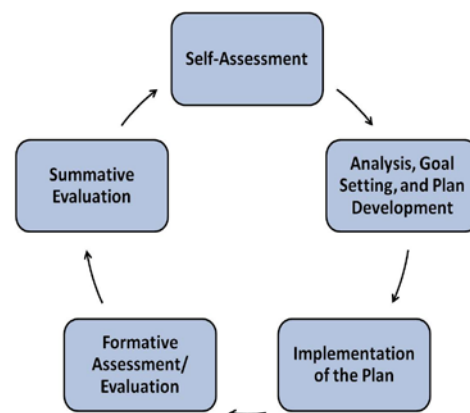
Yet measuring how much students have learned is challenging, particularly in subjects and grades not typically covered by traditional assessments. This guide is meant to help districts as they begin this process. Questions this guide will answer include:

- How does student growth fit into the 5-Step Cycle?
- What criteria should I use when selecting district-determined measures of growth?
- Which types of measures work best for which types of educators?
- For which educators must I use state assessment data?
- What is meant by high, moderate, and low impact on student learning?
- By when must I establish my district's measures?
- What supports will I need to establish in my district to implement this system?
- What do I need to report to ESE, and when?
- When will additional information be available from ESE?

While this guide will help districts start these important conversations, it will not address all the questions educators are likely to have or issues they are likely to confront as they undertake this work. Responses to many critical questions will evolve over the next months and years as districts in Massachusetts undertake this important work and as research and results from other states engaged in similar work accumulates. ESE will continue to provide updated guidance as new information becomes available.

Introduction and Purpose

On June 28, 2011, the Massachusetts Board of Elementary & Secondary Education adopted new regulations to guide evaluation of all licensed educators: superintendents, principals, other administrators, teachers and specialized instructional support personnel. Under these regulations, all educators will participate in a 5-Step Cycle of continuous improvement, at the end of which they receive a summative rating based on both their performance against the Standards and Indicators contained in the regulations, as well as attainment of goals established in their Educator Plans.¹ Educators in Level 4 schools and early adopter districts began implementing the 5-Step Cycle in 2011–12. All Race to the Top (RTTT) districts will be launching the 5-Step Cycle in fall 2012.



The regulations include a second dimension to the educator evaluation system: Every educator will receive a rating of high, moderate, or low for their impact on student learning. This impact rating is based on trends and patterns in student learning, growth, and achievement using at least two years of data and at least two measures of student learning, each of which is comparable across grade or subject district-wide.² These measures are referred to as **district-determined measures** and are defined in the regulations as:

...measures of student learning, growth, and achievement related to the Massachusetts Curriculum Frameworks, Massachusetts Vocational Technical Education Frameworks, or other relevant frameworks, that are comparable across grade or subject level district-wide. These measures may include, but shall not be limited to: portfolios, approved commercial assessments and district-developed pre and post unit and course assessments, and capstone projects.” (CMR 35.02)

At the heart of the work for districts in this phase of educator evaluation work is selecting and using credible district-determined measures of student learning, growth and achievement for a broad range of subjects and grade levels. Selecting district-determined measures gives districts a long-sought opportunity to broaden the range of *what* knowledge and skills they assess and also *how* they assess learning. Districts will be identifying or developing at least two measures for assessing student learning for educators in all grade spans and subject areas, including English language arts, family and consumer science and industrial arts, fine and performing arts, foreign languages, history and social studies, mathematics, physical education and health, science and technology, vocational and business education, and others.

¹ The process involved in arriving at an educator’s summative rating has been described in detail in Parts I-VI of the Model System (<http://www.doe.mass.edu/edeval/model/http://www.doe.mass.edu/edeval/model/>).

² See Appendix A for excerpts from the state regulations related to district-determined measures and rating educator impact on student learning. See Section VI, Part C of this Guide for details on how the Impact Rating is used.

Districts will be able to consider measures of both academic and social/emotional learning. The Department encourages districts to assess the application of skills and concepts embedded in the new English language arts and mathematics curriculum frameworks that cut across subjects, such as expository writing, non-fiction reading, reasoning and analysis, and public presentation. Districts are encouraged by the regulations and the Department to look beyond traditional standardized, year-end assessments to performance assessments and capstone projects scored against district rubrics and scoring guides, as well as interim and unit assessments with pre- and post- measures of learning. A district might consider, for example, using:

- Annual science fair projects in grades 6, 7, and 8 as the basis for assessing student growth in critical scientific skills and knowledge, as well as public presentation skills;
- Periodic student surveys of nutritional and exercise knowledge and practices, combined with annual fitness assessments, to assess student growth in critical life skills; and,
- The essay portion of common quarterly exams in U.S. History to assess growth in persuasive writing skills.

In short, the requirement to identify district-determined measures can be the impetus for broadening the range and strengthening the relevance of assessments and assessment practices used in schools and districts in the Commonwealth.

Selecting and using credible district-determined measures will also support successful implementation of the 5-Step Cycle. Strong student learning goals need to be built upon strong assessments. Sound district-determined measures will provide educators with useful data for self-assessment and a basis for relevant targets to consider for the “specific, actionable, and measurable” goal(s) for improving student learning required in the 5-Step Cycle. Since district-determined measures of student learning are “comparable across grade or subject level district-wide,” their existence will be especially helpful for teams of educators as they analyze data about student learning together and consider shared goals. Finally, and most importantly, credible district-determined measures focused on student growth will improve educator practice and student learning.

Districts will not be left alone to find and develop district-determined measures on their own. ESE has ambitious plans, described in more detail later in this Guide, to work with knowledgeable practitioners—administrators, teachers and others—to identify, share, and, in some cases, develop exemplar assessments for a wide range of content and grade levels that districts can consider.

This guide is designed to assist district leaders and other educators as they begin to explore appropriate district-determined measures and begin to envision a process for using results from district-determined measures to rate each educator’s impact on student learning. It is intended to lay the groundwork and provide tools for districts to begin planning for the challenging and important work of establishing multiple, instructionally useful, and credible measures of student learning, and using results from these measures to arrive at ratings of educator impact on student learning.

This guide will not address all the questions educators are likely to have or issues they are likely to confront as they undertake this work. Responses to many critical questions will evolve over the next months and years as districts in Massachusetts undertake this important work and as research and results from other states engaged in similar work accumulates. The Department is working on several supplements to this guide, based on continuing consultation with stakeholders and lessons learned locally and nationally.

- By winter 2012-13, educators can expect Technical Guide A on District-Determined Measures, which will also include details on reporting requirements.³
- By spring 2013, educators can expect Technical Guide B on Rating Educator Impact.⁴
- By spring 2013, the Department will disseminate details of how districts will be expected to report the district-determined measures it intends to use.
- By summer 2013, educators can expect to see the first exemplar assessments identified and/or developed by ESE in partnership with stakeholders in the Commonwealth.
- By summer 2013, the Department expects to have completed work with stakeholders to produce model contract language districts can use as a starting point for the collective bargaining required to implement the process for rating educator impact on student learning based on trends and patterns.
- By spring 2014, the Department will provide guidance on how to use results of ACCESS, the new statewide assessment of English language proficiency replacing MEPA, to help determine trends in student learning and growth.

³ See Appendix B for a preliminary table of contents for Technical Guide A (District-Determined Measures).

⁴ See Appendix C for the preliminary table of contents for Technical Guide B (Rating Educator Impact on Student Learning)

Implementation Timetable

For many districts, the change in culture and practice required by the new educator evaluation framework is dramatic. Implementation in phases is critical. It is not realistic to expect most districts to implement the 5-Step Cycle of evaluation for the first time and simultaneously identify and, where necessary, pilot all of the district-determined measures they will begin using the next year. Therefore the Department has developed a timetable for implementation that enables districts to do this work in phases, within the requirement established in the regulations that all districts (RTTT and non-RTTT) “identify and report to the Department a district-wide set of student performance measures for each grade and subject that permit a comparison of student learning gains” by September 2013.

- **All districts *without* Level 4 schools as of January 2011** may use the 2013–14 school year to begin to pilot district-determined measures they identify by September 2013. If their pilots lead them to revise their thinking about certain district-determined measures, they can submit a revised plan in September 2014. These districts are expected to begin administering these no later than the 2014–15 school year.⁵ Districts may also choose not to pilot a particular district-determined measure in 2013–14 and instead actually administer the district-determined measure during the year, using the results as the first year of trend data for impact ratings. This may be particularly appropriate for some of the early adopter districts.
- **The 10 districts with Level 4 and/or school improvement grant schools as of January 2011**⁶ are expected to begin administering (not piloting) the district-determined measures they identify by September 2013 in the 2013–14 school years for at least 50 percent of their educators. These districts have had a head start because they began implementation of the 5-Step Cycle in some of their schools in 2011–12. Therefore, these districts will be able to use results from assessing student growth in 2013–14 as the first year of trend data for at least 50 percent of their educators district-wide.⁷ They, too, may submit a revised plan for district-determined measures in September 2014, based on their experience in 2013–14.

The next page includes a chart outlining expectations for implementation of both the 5-Step Cycle (summative performance rating) and district-determined measures (impact on student learning rating) components of the new educator evaluation system.

⁵The regulations permit a district to phase in implementation over a two-year period, with at least half of its educators being evaluated under the new system in the first year.(CMR 35.11(1)(d). For example, a district could begin with all of its administrators and all high school staff the first year (assuming that together they account for 50% of its licensed educators). In the next year, the district could add its elementary and middle school teaching staff.

⁶ Boston, Chelsea, Fall River, Holyoke, Lawrence, Lowell, Lynn, New Bedford, Springfield and Worcester.

⁷ Boston, Chelsea, Fall River, Holyoke, Lawrence, Lowell, Lynn, New Bedford, Springfield and Worcester.

	Implementation Requirements	Districts with Level 4 and/or SIG schools⁸ (District-wide)	RTTT Districts	Non-RTTT Districts
1	Submit collective bargaining agreements for 5-Step Cycle to ESE	September 2012	September 2012	September 2013
2	Implement 5-Step Cycle with at least 50 percent of educators	2012–13	2012–13	2013–14
3	Implement 5-Step Cycle with all remaining educators	2013–14	2013–14	2014-15
4	Submit collective bargaining agreements for rating educator impact to ESE	September 2013	September 2014 ⁹	September 2014 ⁹
5	Submit initial plans for district-determined measures to ESE	September 2013	September 2013	September 2013
6	Begin piloting district-determined measures	2012–13	2013–14 ¹⁰	2013–14 ¹⁰
7	Submit revised plans for district-determined measures to ESE	September 2014	September 2014	September 2014
8	Begin administering district-determined measures to establish first year of trend data for at least 50 percent of educators	2013–14	2014–15	2014–15
9	Administer district-determined measures to establish first year of trend data for remaining educators	2014–15	2015–16	2015–16
10	If additional changes to district-determined measures are warranted, submit revised plans for them to ESE	September 2015	September 2015	September 2015
11	Report ratings of impact on student learning for at least 50 percent of educators ¹¹	October 2015	October 2016	October 2016
12	Report ratings of impact on student learning for all educators	October 2016	October 2017	October 2017

⁸ At Level 4 and School Improvement Grant schools, the timetable for implementing some of these steps may differ. The Department’s Center for Targeted Assistance and Office for Educator Preparation, Policy and Leadership will continue to work with these school and district leaders to communicate expectations and support effective implementation.

⁹ This deadline assumes the district intends to use 2013–14 for piloting district-determined measures.

¹⁰ Districts have the option of piloting or implementing some or all of their district-determined measures. For example, some measures may already be well established and no pilot is necessary.

¹¹ The regulations require that the trends be based on at least two years of data; if it is decided at the district level that three years of data are required instead of two, then the first ratings of educator impact will be one year later.

Getting Started

Many educators follow this sequence of questions when they plan their work on behalf of students:

- 1) What do I most want my students to know and be able to do by the end of this course/year?
- 2) What assessments are available to assess my students' learning and growth?
- 3) Where are my students starting?
- 4) What do I expect them to achieve by key milestones throughout the year (e.g., by first quarter, second quarter, etc.)?
- 5) How will I chart each student's progress along the way so I can know how to re-group for the purposes of re-teaching, enrichment, and/or acceleration?
- 6) How did my students do? What should I do differently next time?
- 7) How did the results of my students compare to others'? What do I need to do differently next time?

As the Department has worked with early adopters, other stakeholders and national experts to develop initial guidance on this component of the new educator evaluation framework, we have concluded that districts will need to pose similar questions in a similar sequence as they undertake the work on this component of the new educator evaluation framework.

It is not until Question 7 that the focus of the questions turns to comparing results among students and educators. While it is important to know that assessments will need ultimately to permit comparison, there is much critical work to be done on the preceding questions—and that work holds great promise for improving teaching and learning in and of itself.

This guide focuses more on the questions leading up to 7 rather than Question 7 itself. There is much to be learned from answering Questions 1 through 6 at the district level that will inform definitive guidance for Question 7. We will draw on that learning to provide final guidance on Question 7 in one or more supplements to this guide.

Identifying and Selecting District-Determined Measures

Key Criteria

Districts will need to identify and use district-determined measures that are appropriate for assessing student gains. To ensure that they reflect educator impact, district-determined measures should:

1) **Measure student learning growth, not just achievement.**

Many assessments measure the level of a student's performance in a given subject at a particular point in time. For instance, the MCAS measures whether a student has performed at the warning/failing, needs improvement, proficient, or advanced level in, say, grade 4 English language arts or grade 8 science. Measures of growth, by comparison, measure a student's change over time. They answer the question, "How much has this student improved relative to where he or she started from?"

Why focus on growth? Students come to school each year with a wide range of prior academic achievement and therefore begin their next year of instruction with varying levels of readiness to access the curriculum, a situation that is beyond the control of the educator assigned to teach them. Measuring educators' effectiveness solely by the achievement level of their students cannot account for these prior conditions. By comparison, measuring growth can help level the playing field. Improvement in student performance is a more meaningful and fair basis for determining the trends and patterns that will yield the educator's rating of impact on student learning, growth and achievement.

For these reasons, growth measures are best way to rate impact on student learning. However, measures of achievement can serve as acceptable district-determined measures in limited circumstances when they are judged to be the most appropriate and/or most feasible measure for certain educators.

2) **Assess learning as directly as possible.**

Direct measures of student learning are strongly preferred for evaluating educator impact because they measure the most immediately relevant outcomes of the education process. Direct measures assess student growth in a specific area over time using baseline assessment data and end-of-year (or -unit, -term, or -course) assessment data to measure growth. Examples of direct measures include formative, interim and unit pre- and post-assessments in specific subjects, assessments of growth based on performances and/or portfolios of student work judged against a common scoring rubric, mid-year and end-of-course examinations, and progress monitoring. Other options may include measures of social, emotional, or behavioral learning when teaching such skills is an explicit and central component of the curriculum for which an educator bears instructional responsibility. ESE will further explore the potential role of such measures in future guidance.

Indirect measures of student learning do not measure student growth in a specific content area or domain of social-emotional learning but do measure the consequences of that learning. These measures include, among others, changes in promotion and graduation rates, attendance and tardiness rates, rigorous course-taking pattern rates, college course matriculation and course remediation rates, discipline referral and other behavior rates, and other measures of student engagement and progress toward school academic, social-emotional, and other goals for students. Just as for direct measures, baseline data is necessary for indirect measures in order to

Identifying and Selecting District-Determined Measures

assess growth. For some educators, including school leaders, district administrators, and guidance counselors, it may be appropriate to use an indirect measure of student learning along with other direct measures. ESE recommends that at least one measure be direct.

3) **Be administered in the same subject and/or grade across all schools in the district.**

To qualify as a district-determined measure, an assessment must be administered across all schools and classes in the district where the same subject is taught, e.g., Algebra I, grade 2 science, or grade 8 music. Assessing skills that cut across subject areas such as critical analysis, writing, or non-fiction text reading will enable a district to use the same or similar district-determined measures with more than a single teacher in cases where that teacher is the lone teacher of a subject.

4) **Differentiate high, moderate, and low growth.**

State regulations define high, moderate, and low growth:

“(a) A rating of high indicates significantly higher than one year's growth relative to academic peers in the grade or subject.

(b) A rating of moderate indicates one year's growth relative to academic peers in the grade or subject.

(c) A rating of low indicates significantly lower than one year's student learning growth relative to academic peers in the grade or subject.”

The “[Rating Educator Impact on Student Learning](#)” section of this document explains this in more detail.

Districts should also be aware that in Massachusetts, an educator's impact on student learning is determined neither on a single year of data nor on a single measure of student learning. It must be based on a *trend* over time of at least two years, and it should reflect a *pattern* in the results on at least two different assessments. Patterns refer to consistent results from multiple measures, while trends require consistent results over at least two years. Thus, creating a rating of impact on student learning requires at least two measures in each year and at least two years of data. The “[Rating Educator Impact on Student Learning](#)” describes this further.

Key Takeaways

- Wherever possible, district-determined measures of student learning should measure growth, not just achievement, and should measure the direct outcome of student learning rather than indirect outcomes such as promotion or graduation rates.
- To qualify as a district-determined measure, an assessment must be administered across all schools in the district where the same subject is taught and must be able to differentiate high, moderate, and low growth.
- The rating of an educator's impact on student learning must be based on trends and patterns: at least two measures in each year and at least two years of data.

Identifying and Selecting District-Determined Measures

Selecting Appropriate Measures of Student Learning

Identifying and Exploring Options

Every educator will need data from at least two measures of student learning in order for trends and patterns to be identified. State regulations require that the MCAS student growth percentile (SGP) scores must be used as one measure “where available.” Similarly, regulations require that gain scores on the soon-to-be-replaced Massachusetts English Proficiency Assessment (MEPA) be used where available. The chart below shows educators for whom MCAS student growth percentiles are available and should be used as one measure for determining trends and patterns in their students’ learning.

Educator Category	Grade Levels	Subject Area Taught	MCAS SGP Required
Teachers	4–8 ¹²	Mathematics	Mathematics SGP
		English/Reading	English Language Arts SGP
Administrators	Schools enrolling grades 4, 5, 6, 7, 8 and/or 10	Mathematics and/or English language arts	ELA and mathematics
Instructional Specialists (Educators who support specific teachers or subjects)	Grades 4, 5, 6, 7, 8 and/or 10	Mathematics and/or English language arts	ELA and/or mathematics depending on the subject(s) being taught by the educators being supported

The Massachusetts English Proficiency Assessment (MEPA) will be replaced in the 2012–13 school year by ACCESS, a new assessment of English language proficiency developed by a multi-state consortium. ESE has not yet determined how progress of English language learners will be measured on ACCESS. The Department will issue guidance on ACCESS in the educator evaluation system by spring 2014.

Districts may determine whether to use MCAS growth percentiles or the eventual ACCESS-based growth measures as a district-determined measure for educators who do not have a direct role in teaching or overseeing instruction in those subject areas. For instance, in a district with a high priority on literacy across the curriculum, it may make sense to include an MCAS English language arts growth measure as a district-determined measure for many additional teachers or other staff.

But even for educators with a measure of growth available from a state assessment, districts may still need or want to identify at least one other district-determined measure of growth as well. And for many educators, a state assessment will not be relevant. Thus districts will need to select district-determined measures for the majority of subjects and grades taught.

To identify those measures, districts can start by assessing measures of student learning already in use in their schools: standardized tests, commercial and textbook assessments, departmental exams, performance assessments, and the like. *A scan of a district’s currently used assessments is the most critical first step.* The “Suggested Criteria for Reviewing District-Determined Measures of Student Learning” on pages 15-16 of this document are designed to support districts in analyzing the strength of their existing measures for evaluation purposes.

¹² Any educator responsible for teaching both subjects must use either the MCAS SGP for mathematics or ELA, but the district is not required to use both for such an educator. If a district chooses to utilize both the MCAS SGP for mathematics and ELA to determine an educator’s impact on student learning, it would meet the requirements of multiple measures. If the district does not use both, another district-determined measure would be needed .

Identifying and Selecting District-Determined Measures

Resources

Districts can also examine what other districts are using. As mentioned earlier, the regulations allow a wide range of assessments to be used as district-determined measures.¹³ Identifying, selecting, and/or developing district-determined measures gives districts an opportunity to expand what content and skills are being assessed and how they are being assessed.

Among the assessments currently in use in a number of Massachusetts districts are these that assess growth and/or progress:

- Assessment Technology, Incorporated: Galileo
- Behavioral and Emotional Rating Scales (BERS-2)
- Developmental Reading Assessment (DRA)
- Dynamic Indicators of Basic Early Literacy Skills (DIBELS)
- MCAS-Alternative Assessment (MCAS Alt)¹⁴
- Measures of Academic Progress (MAP)

Massachusetts districts interested in taking a broader approach to district-determined measures of student learning can also use practices in other states as a starting point. The following list includes samples of non-traditional assessments currently used in states and districts around the country.

- 1) **Connecticut Student Performance Task Database**
<http://www.ctcurriculum.org/search.asp>

Connecticut has developed a tool, located at www.CTcurriculum.org, that allows users to search a database of student performance tasks by grade level and content area. Teachers create and upload student performance tasks that are aligned to state standards along with accompanying scoring rubrics and exemplars. The website is still under development but once complete, there will be a statewide committee in each content area that provides quality control for all uploaded tasks. This tool is a good resource for Massachusetts districts interested in borrowing or developing performance-based assessments for traditionally non-tested grades and subjects.

¹³ "...including, but not limited to: commercial tests, district-developed pre and post unit and course assessments, capstone projects, and performance assessments including portfolios." (35.02)

¹⁴ For more information, visit <http://www.doe.mass.edu/mcas/alt/>. The Department is currently exploring options and potential best practices for the use of the MCAS Alt in teacher evaluation, to be released in later guidance.

Identifying and Selecting District-Determined Measures

2) **Kansas Performance Teaching Student Portfolio: Biology Grade 9**

<http://www.ksde.org/LinkClick.aspx?fileticket=5jJEdT7ihcg%3d&tabid=3769&mid=11995>

NOTE: pre- and post-assessment information can be found on pp. 9, 12, 15, 18 and 19

This teaching portfolio framework is built as a template for practitioners to create and justify particular assessments for traditionally non-tested subjects. The template has space for unit objectives, levels of Bloom's taxonomy that are covered, and a pre- and post-assessment. Within the space given to the assessments, teachers describe the following steps/components: specific types and numbers of questions, description of the assessment, rationale for choosing said assessment, explanation of any adaptations made, objectives covered, scoring, and how student results will affect unit plans. This is a very structured way for teachers to create their own assessments, while simultaneously being reflective and always aligning assessments to objectives. If applied in the same way across a district and monitored for fidelity, this approach could result in comparable measures across a district.

3) **Minneapolis (MN): Literacy and numeracy assessments Grades K-1**

http://rea.mpls.k12.mn.us/uploads/kind_summary_fall_2010_with_2009_bench_2.pdf

This set of pre- and post-assessments are performance assessments for all kindergarteners in Minneapolis in the subjects of literacy and numeracy. Many of the assessments involve the student sitting with a proctor and answering various questions. While the specific questions vary over time, all of the assessments must include the same variety of questions, e.g. questions about which pictures have words that rhyme and which number comes after a particular target number. These simple benchmarking assessments could serve as a guideline for Massachusetts districts designing measures for the early elementary grades.

4) **Illinois: Student growth portion of principal evaluation all grades**

http://www.ilprincipals.org/resources/resource-documents/principal-evaluation/peac_prin_eval_model.pdf (pp. 5, 26)

Developed as a response to the Performance Evaluation Reform Act, this state model for principal evaluation includes, among other things, a section on using student growth information to inform principal evaluation. The framework delineates a definition of student growth, the process for collecting student growth data, definitions of student growth performance levels, and a summative rating matrix. In addition, there is a list of recommended non-test measures for use in principal evaluation, including student attendance, postsecondary matriculation and persistence, graduation rate, discipline information (e.g. referrals), and dual-credit earning rates. These measures mirror many of the recommendations elsewhere in this guidance document, and may offer a useful starting place for identifying district-determined measures for Massachusetts administrators.

Identifying and Selecting District-Determined Measures

Finally, districts may want to look at the performance assessment work underway in several Massachusetts districts through the [Quality Performance Assessment Initiative](#). The project defines quality performance assessments as “multi-step assignments with clear criteria, expectations and processes that measure how well a student transfers knowledge and applies complex skills to create or refine an original product.” These involve portfolios, performances, products and/or projects. Examples described in a recent publication¹⁵ include:

- Fenway Park High School’s “Junior Review, Senior Institute and Senior Position Paper”
- Pentucket Regional’s district-wide assessment of “Habits of Learning” using common rubrics
- Cape Cod Lighthouse Charter School’s “Assessment Validation Process”

Choosing Credible Measures

When reviewing and eventually selecting assessments as district-determined measures, districts should focus on the credibility of those measures. Credibility starts with a measure’s validity: the extent to which the assessment measures what it is intended to measure. A credible measure must also be reliable: if an assessment is reliable, a student who takes it multiple times should get a similar score each time. Finally, a credible assessment must be fair and free of bias: bias occurs in assessments when different groups of test-takers, such as different demographic groups, are advantaged or disadvantaged by the type and/or content of the assessment. Ensuring fairness requires that items and tasks are appropriate for as many of the students tested as possible and are free of barriers to their demonstrating their knowledge, skills and abilities.

These key concepts may also help districts as they begin to consider options for district-determined measures:

- Student growth should be measured as change in student learning over time; that is, it should measure the difference between an assessment that establishes a baseline or starting point and a post-assessment covering similar concepts and content.
- The measure of student growth should offer a way to determine whether each student is attaining one year’s growth in one year’s time, significantly more than one year’s growth in one year’s time, or significantly less than one year’s growth in one year’s time.
- That being said, the measures do not necessarily need to measure growth over the whole school year. A measure that shows that a student progressed at expectations for a particular curriculum unit, for instance, could be considered evidence that the student is on track to achieving one year’s growth in one year’s time. Thus the formative assessments that educators use to inform instruction during the school year can also potentially be used to measure student growth. Recent research confirms the power of interim assessments to improve both instructional practice and student learning,¹⁶ bolstering the case for considering common unit, interim, and quarterly assessments as strong candidates for district-determined measures as long as pre-test or baseline data is available.

¹⁵ Brown, C. and Mevs, P. (2012) Quality Performance Assessment: Harnessing the Power of Teacher and Student Learning. Center for Collaborative Education and Nellie Mae Foundation. Available at www.qualityperformanceassessment.org.

¹⁶ <http://www.cpre.org/testing-teaching-use-interim-assessments-classroom-instruction-0>

Identifying and Selecting District-Determined Measures

- When feasible, assessment results should be compared to regional, state, or national norms or benchmarks. Having the perspective of regional, state, or national norms and benchmarks will help ensure that districts will have comparable expectations for high, moderate, and low growth in student learning. It is particularly important in situations when an educator is a “singleton”—the only person in the district with a particular role, such as a middle school principal in a district with a single middle school. In these circumstances, districts will have no opportunity to compare to other middle school principals within the district to know whether the growth attained by that principal’s students is high, moderate, or low. Collaboratives and other regional consortia may have a role to play in fostering regional efforts in this area. For example, Chelsea, Everett, Malden, Revere and Winthrop are collaborating to develop literacy and mathematics units with pre- and post-assessments that they intend to share.

Four core questions can help frame a district’s process for reviewing options for district-determined measures:

- Is the content assessed by the measure aligned with state curriculum frameworks or district curriculum frameworks?
- Can the measure be used to measure student growth from one time period to the next?
- Is the measure reliable and valid? If this has not been tested, are there sufficient data available from previous administrations to assess reliability and validity?
- Can consistency of administration be assured across all schools serving the same grade spans and content areas within the district?

Districts may find the rubric on the next pages helpful as they weigh the relative strengths and weaknesses of the different assessments they are considering and/or developing. Adapted from the work of Dr. Margaret Heritage from the [Assessment and Accountability Comprehensive Center](#), the rubric focuses attention on eleven key criteria. The first three criteria focus on elements important for measures that intend to assess student growth. The remaining eight focus on criteria that can be applied to measures of growth as well as measures of achievement. Few but the most rigorous commercial or state assessments will earn ratings of 3 on all criteria, and some criteria are more applicable to some kinds of assessments than others. Technical Guide A will provide more guidance on using this rubric as well as other resources designed to help districts assess a range of types of assessments for their use as district-determined measures.

Identifying and Selecting District-Determined Measures

Suggested Criteria for Reviewing District-Determined Measures of Student Learning

Criterion	0	1	2	3
Growth: Measures change over time, not just achievement	Insufficient evidence	The measure does not assess student baseline knowledge or measures growth indirectly (i.e., change in graduation or dropout rates).	The measure assesses student knowledge on related content/standards before and after the learning period.	The measure assesses student knowledge on the same or vertically aligned content/standards before and after the learning period.
Growth: Identifies how much growth is sufficient for the period covered by the assessment	Insufficient evidence	The measure does not allow a definition of "sufficient growth."	The measure defines "sufficient growth" with a relative comparison such as a comparison to other students in the district.	The measure defines "sufficient growth" relative to an external standard such as a national norm or a criterion.
Growth: Measures change relative to an academic peer group	Insufficient evidence	The measure does not quantify change.	The measure quantifies change overall but does not differentiate by academic peer group.	The measure quantifies the change in a student relative to his or her academic peer group: others with similar academic histories.
Consistency of administration	Insufficient evidence	There are no procedures for either (a) when the test is administered and (b) the time allocated for the test, or procedures are not consistently applied.	There are consistently applied procedures for <u>either</u> (a) when the test is administered or (b) the time allocated for the test.	There are consistently applied procedures for <u>both</u> (a) when the test is administered and (b) the time allocated for the test.
Alignment to Standards	Insufficient evidence	The measures are not aligned to targeted grade-level standards.	The measures partially reflect the depth and breadth of targeted grade-level standards.	The measures reflect the full depth and breadth of targeted grade-level standards.
Content Validity: instructional sensitivity	Insufficient evidence	The measure only peripherally addresses content that can be taught in classes, while mostly addressing developmental issues or skills acquired outside of school.	The measure addresses content taught primarily in classes, though some students may have learned the content outside of school.	The measure addresses content explicitly taught in class; rarely if ever is this material learned outside of school.

Identifying and Selecting District-Determined Measures

Criterion	0	1	2	3
Reliability: items	Insufficient evidence	The number of aligned, content-related items is clearly insufficient for reliability.	There are multiple but insufficient items aligned to content for reliable measurement of outcomes.	There are sufficient items aligned to content to enable reliable measurement of outcomes.
Reliability: scoring of open-ended responses	Insufficient evidence	There are no scoring criteria related to the performance expectations.	There are general scoring criteria that are not specifically related to the performance expectations.	There are precise scoring criteria related to the performance expectations.
Reliability: rater training	Insufficient evidence	There are no procedures for training raters of open-ended responses.	There are limited procedures for training raters of open-ended responses.	There are clear procedures for training raters of open-ended responses.
Reliability of scores	Insufficient evidence	There is no evidence of score or rater reliability across similar students in varied contexts.	There is evidence that the scores or raters have low reliability across similar students in varied contexts.	There is evidence that the scores or raters are reasonably reliable across students in varied contexts.
Fairness and freedom from bias	Insufficient evidence	There are many items that contain elements that would prevent some subgroups of students from showing their capabilities.	There are some items that contain elements that would prevent some subgroups of students from showing their capabilities.	The items are free of elements that would prevent some subgroups of students from showing their capabilities.

Adapted from the work of Dr. Margaret Heritage, Assessment and Accountability Comprehensive Center, CRESST, April, 2012 (May 22, 2012)

Identifying and Selecting District-Determined Measures

Considering State-Identified Exemplar Assessments

Starting in fall 2012, using Race to the Top funding, ESE will work with state content and professional associations, volunteer pilot districts, and experts in assessment to identify, adapt, and potentially develop exemplars of district-determined measures for most major subjects at multiple grade levels. ESE will be working with these groups through one or more contractors selected through a rigorous Request for Response process. The Department hopes to make the first exemplars available by summer 2013 to be reviewed, piloted and revised by interested districts in 2013-14.

Below is an initial list of subject areas for which ESE anticipates identifying, adapting and potentially developing one or more measures of student growth for most grade levels:

- Arts
- English language arts
- Foreign languages (grades 7–12)
- History and social studies
- Mathematics
- Physical education and health
- Sciences
- Vocational and business education (grades 9–12)¹⁷

Building Local District-Determined Measures

Districts may choose to build their own district-determined measure for one or more grades and/or subject areas. This process allows districts to create assessments closely aligned to their learning goals and practices and is an opportunity for rich conversations among educators about student learning goals and effective educator practice. That said, identifying or developing credible district-wide assessments in areas beyond reading, English language arts, and mathematics will be new work for at least some districts (and for ESE). Few districts have strong district-wide assessments of growth in many subjects or many grade levels. Few districts use a wide range of formats. Few employ systematic ways of assessing growth in social and emotional learning. Over time, as districts and ESE continue to collaborate to identify, refine, and/or develop measures of student learning and growth for use as district-determined measures, strong exemplars from the field can be identified and shared. The quality of assessments will improve as well so that they will gradually reflect more and more of the characteristics on the right side of the rubric on pages 15 and 16.

Technical Guide A will provide more detailed guidance and resources for districts on selecting and developing district-determined measures, including non-traditional measures and examples of credible locally developed measures already in use in the Commonwealth. In addition, starting in fall 2012, districts will be invited to submit one locally developed measure for review and feedback. These locally developed district-determined measures will be considered for inclusion among the exemplar state measures planned for initial release in summer 2013. ESE will keep districts updated on how to submit local assessments and the process for review.

¹⁷ESE will work with state associations to identify two or three subjects; the tentative plan is to focus on two heavily enrolled courses: culinary arts and automotive repair.

Identifying and Selecting District-Determined Measures

Piloting District-Determined Measures

The Department recommends that districts develop plans for piloting the new assessments they are anticipating using as district-determined measures. Both small- and large-scale piloting offer an opportunity to assess the reliability and validity of measures and work out the logistical details of administering the assessments. Additionally, pilots will enable educators to identify exemplars of student work at different levels of proficiency that can help them calibrate scoring for future administrations.

See the section on "[Implementation Timelines and Reporting Requirements for District-Determined Measures](#)" on page 22 for information about how piloting fits into the overall timeline for implementation.

Key Takeaways

- Districts must use measures of growth from state assessments where they are available. But many educators will not have a relevant measure of growth from a state assessment, and even those that do must have at least one locally determined measure as well. Thus, districts will need to identify credible measures of growth for most subjects and grades.
- Credible measures of growth are valid (measure what they intend to measure), reliable (obtain similar results across repeated assessments of the same student), and free of bias (do not advantage or disadvantage particular groups of test-takers).
- Measures of student growth should identify whether the student has attained one year's growth, significantly more than one year's growth, or significantly less than one year's growth in one year's time. However, measures do not necessarily need to measure growth over the entire year in order to answer this question.
- The state will be working with the field and assessment experts to build exemplar district-determined measures in the core content areas plus vocational and business education. The first exemplars should be available for district piloting as early as summer 2013.
- Districts may choose to build their own district-determined measures for one or more grades and/or subject areas. ESE recommends that these measures be pilot-tested before deployment.

Identifying and Selecting District-Determined Measures

Matching Educators with Appropriate Measures

The types of measures that are most appropriate for measuring an educator's impact on student learning vary by role. Most educators fall into one of four broad categories: teacher, administrator, instructional specialist, and specialized instructional support personnel. Positions in the last category are sometimes called "caseload educators."¹⁸

For some roles, such as teachers who teach a single subject to a consistent group of students, it may be comparatively straightforward to narrow the range of appropriate measures to use to assess the trends and patterns that will serve as an indicator of the educator's impact on student learning in that subject. Yet many teachers teach more than one subject. It is not necessary (nor practical in the case of many elementary teachers) to have a district-determined measure for each subject or course a teacher teaches or an administrator oversees. Districts will need to determine which aspects of an educator's practice are most important to measure, given district priorities. Most importantly, districts will need a process for making choices and ensuring that every educator knows in advance what measures are going to be used to establish the trends and patterns in student learning that will lead to a rating of his or her impact.

Other educators, such as library media specialists, support student learning across multiple classes or schools. Deciding what areas of student learning the district-determined measures applied to them will call for more decisions. For example, should a school-wide writing assessment be one of the measures? In addition, for some educators, such as guidance counselors, indirect measures such as promotion or graduation rates or indicators of college readiness may be more appropriate.

The chart on the next page offers an initial framework for matching educators with appropriate measures. It details for each educator category: 1) educator roles included, 2) the students who are likely to be considered in assessing impact, and 3) appropriate kinds of measures of student learning

¹⁸ Representatives of state associations representing a number of the specialists in this broad category sought a name other than "caseload educator" because the connotation of that term does not take into account the broadened role many of them have in most schools now. "Specialized Instructional Support Personnel" is the term recommended by a national consortium of organizations serving nurses, psychologists, guidance counselors, social workers and others. Neither "Specialized Instructional Support Personnel" nor "caseload educators" includes paraprofessionals or aides, positions which are not covered by the evaluation regulations.

Identifying and Selecting District-Determined Measures

Category	Roles included	Appropriate types of measures*
<p>Teachers</p> <p><i>Assess students in subject area(s) being measured and taught by that teacher</i></p>	<ul style="list-style-type: none"> ▪ Grades prekindergarten through high school <ul style="list-style-type: none"> – English as a Second Language – English language arts – Family and consumer science and industrial arts – Fine and performing arts – Foreign languages – History and social studies – Mathematics – Physical education and health – Science and technology – Special education – Vocational and business education ▪ Others 	<ul style="list-style-type: none"> ▪ Direct measures of learning specific to subjects and grades ▪ Direct measures of learning specific to social, emotional, behavioral, or skill development ▪ Interim assessments, unit tests, end-of-course tests with pre-tests or other sources of baseline data ▪ Performance assessments ▪ Student portfolios, projects and performances scored with common scoring guide
<p>Administrators</p> <p><i>Assess students in the district, school, or department overseen by that educator, depending on the educator's specific role</i></p>	<ul style="list-style-type: none"> ▪ Superintendents ▪ Other district administrators ▪ Principals ▪ Other school-based administrators, including assistant principals ▪ Department chairpersons, including teachers who serve in this role ▪ Others 	<ul style="list-style-type: none"> ▪ Direct measures of learning specific to subjects and grades ▪ Direct measures of learning specific to social, emotional, behavioral, or skill development ▪ Indirect measures of student learning such as promotion and graduation rates <p><i>Impact may be calculated at the district, school, or department level depending on the educator's role</i></p>
<p>Instructional Specialists</p> <p><i>Assess students in the classes of all teachers supported by this educator</i></p>	<ul style="list-style-type: none"> ▪ Instructional coaches ▪ Mentors ▪ Reading specialists ▪ Team leaders ▪ Others 	<ul style="list-style-type: none"> ▪ Direct measures of student learning of the students of the teachers with whom they work, measuring <ul style="list-style-type: none"> ▪ learning specific to subjects and grades ▪ learning specific to social, emotional, behavioral, or skill development <p><i>Impact may be calculated at the district, school, or department level depending on the educator's role</i></p>

Identifying and Selecting District-Determined Measures

Category	Roles included	Appropriate types of measures*
<p>Specialized Instructional Support Personnel**</p> <p><i>Assess students in the school, department, or other group based on whether the educator supports the entire school, a department, a grade, or a specific group of students</i></p>	<ul style="list-style-type: none"> ▪ School nurses ▪ School social workers and adjustment counselors ▪ Guidance counselors ▪ School psychologists ▪ Library/media and technology specialists ▪ Case managers ▪ Others 	<ul style="list-style-type: none"> ▪ Direct measures of learning specific to subjects and grades ▪ Direct measures of learning specific to social, emotional, behavioral, or skill development ▪ Indirect measures of student learning such as promotion and graduation rates <p><i>Impact may be calculated at the district, school, department, or other group levels depending on whether they serve multiple schools, the entire school, a department, a grade, or a specific group of students</i></p>

*These are examples appropriate for one or more of the four categories. Each can be used to measure growth (progress).

** These positions are sometimes called “caseload educators.” See Footnote 14.

Technical Guide A (District-Determined Measures) will supplement this guidance on district-determined measures and will detail ESE’s recommendations for kinds of measures especially appropriate for each educator group.

Key Takeaways

- The appropriate measures of growth for an educator will vary depending on whether he or she is a classroom teacher, administrator, instructional specialist, or specialized instructional support personnel.
- It is not necessary to have a measure of growth for each subject or course an educator teaches or an administrator oversees.
- Districts should identify measures based on district priorities and individual and team professional development goals.

Matching Students to Their Educators

Most students are served by many different educators during a given school year. Attributing an individual educator's impact on student learning must reflect the realities of schools. Team teaching, students who are pulled out for extra support, students who change classes or schools mid-year, and a host of other issues affect how specific students' assessment results are able to be fairly attributed to specific teachers. For administrators, instructional specialists, and specialized support personnel, comparable issues can affect how which students' results should factor into ratings of their impact on student learning.

For evaluations to be fair and accurate, these relationships must be adequately accounted for. Educators must have an opportunity to review and confirm the list, or roster, of students whose learning gains will be taken into account in determining their impact rating. (This process is often called "roster verification.") Districts will have to develop clear policies for attribution, the process of designating responsibility among educators for their impact on students' learning, growth, and achievement, and for roster verification, the process of confirming the accuracy of student-educator links. See Appendix D for more detail on these topics. Technical Guide B, expected to be available in spring 2013, will provide additional guidance for districts.

Implementation Timelines

The regulations require districts to submit their preliminary plans for district-determined measures to ESE for review by September 2013. ESE intends to use district reports to understand the approaches districts are taking in developing and selecting district-determined measures and share promising approaches with districts. Districts will be invited to update their plans and submit changes to ESE in September 2014 and 2015 once pilots are complete. Districts should anticipate reporting the measures they will use to assess educator impact in each content area/grade for teachers and for each other category of educator: administrator, instructional specialist, and professional instructional support personnel. The details of specific reporting requirements will be developed over the next twelve months in consultation with stakeholders and will be detailed in a supplement to this guidance.

Most RTTT districts can use the 2013–14 school year to pilot the district-determined measures they have identified by September 2013, including any of the state exemplars. Additionally, districts have the option of phasing in district-determined measures and educator impact ratings over a two-year period. They can begin with 50 percent of their staff in the first year. That means that they can use 2014–15 to administer district-determined measures for 50 percent of their staff and pilot district-determined measures with the other 50 percent. By 2015–16, they would be administering at least two district-determined measures for each educator.

The table below, excerpted from the table on page 6, summarizes the timeline.

Implementation Timelines

District-Determined Measures Implementation Requirements	Districts with Level 4 and/or SIG schools ¹⁹ (District-wide)	RTTT Districts	Non-RTTT Districts
Submit initial plans for district-determined measures to ESE	September 2013	September 2013	September 2013
Begin piloting district-determined measures	2012–13	2013–14 ²⁰	2013–14 ¹⁹
Submit revised plans for district-determined measures to ESE	September 2014	September 2014	September 2014
Begin administering district-determined measures to establish first year of trend data for at least 50 percent of educators	2013–14	2014–15	2014–15
Administer district-determined measures to establish first year of trend data for remaining educators	2014–15	2015–16	2015–16
If additional changes to district-determined measures are warranted, submit revised plans for them to ESE	September 2015	September 2015	September 2015
Report ratings of impact on student learning for at least 50 percent of educators ²¹	October 2015	October 2016	October 2016
Report ratings of impact on student learning for all educators	October 2016	October 2017	October 2017

Key Takeaways

- Districts must submit their preliminary plans for district-determined measures to ESE for review by September 2013. They may update their plans and submit changes in September 2014 and 2015.
- Most districts can pilot-test measures and then phase in implementation over two years.

¹⁹ At Level 4 and School Improvement Grant schools, the timetable for implementing some of these steps may differ. The Department’s Center for Targeted Assistance and Office for Educator Preparation, Policy and Leadership will continue to work with these school and district leaders to communicate expectations and support effective implementation.

²⁰ Districts have the option of piloting or implementing some or all of their district-determined measures. For example, some measures may already be well established and no pilot is necessary.

²¹ The regulations require that the trends be based on at least two years of data; if it is decided at the district level that three years of data are required instead of two, then the first ratings of educator impact will be one year later.

Rating Educator Impact on Student Learning

Measuring student learning gains is a first step. The second step is identifying trends and patterns in student growth over at least two years. Only when both of these steps are complete can an educator's impact on student learning be rated as high, moderate, or low, as required by regulation. ESE recommendations for step two are detailed in this section.

Considerably more detail on rating student growth as high, moderate, or low, analyzing trends and patterns, and determining impact ratings for educators will appear in Technical Guide B (Rating Educator Impact on Student Learning), expected to be released in spring 2013. See Appendix C for a preliminary table of contents.

Defining Student Growth as High, Moderate, or Low

State regulations define high, moderate, and low growth:

- “(a) A rating of high indicates significantly higher than one year's growth relative to academic peers in the grade or subject.
- (b) A rating of moderate indicates one year's growth relative to academic peers in the grade or subject.
- (c) A rating of low indicates significantly lower than one year's student learning growth relative to academic peers in the grade or subject.”

The MCAS Student Growth Percentile model provides a concrete example of how to identify high, moderate, and low growth.²² The model uses a 99-point percentile scale to rank students' academic growth compared to their academic peers: others who scored similarly on previous MCAS tests. This rank is called a student growth percentile (SGP). Student growth on MCAS is defined as high, moderate, or low based on these score ranges:

MCAS SGP Range	Description
1 - 39	Low growth
40 - 60	Moderate growth
61 - 99	High growth

By rank-ordering the SGPs for all the students an educator serves and finding the median (midpoint), one can determine whether his or her median student exhibited low, moderate, or high growth. This can serve as a measure of the educator's impact on student learning.

In theory, the principle of administering a district-wide assessment, dividing all students into three groups by how much they improved to identify high, moderate, and low growth, and then determining whether a particular educator's median student grew at a high, moderate, or low rate can be applied more broadly. The callout box on the next page describes the steps to take for this approach when a pre- and post-test are available.

²² For a more detailed explanation of the Student Growth Percentile (SGP), see <http://www.doe.mass.edu/mcas/growth/>.

Rating Educator Impact on Student Learning

However, this approach may not work well in all circumstances. It can be difficult to calculate growth (change relative to a baseline) if the pre-test is measured differently than the post-test or if the assessment is a more qualitative one such as a portfolio or capstone project. Further, if all educators in a district are achieving high rates of growth among their students, this approach may overemphasize small differences between them, since it typically requires a relative ranking of educators. It also will be ineffective for distinguishing between low, moderate, or high impact for educators who are “singletons,” that is, the only educator of their type in a district (a high school principal in a district with only one high school, or the only theater teacher in a district). The approach also may not be valuable for indirect measures such as graduation or grade promotion rates where an external benchmark may be more meaningful.

The Department will continue to work with early adopter districts and other stakeholders to examine options for determining high, moderate, and low growth on different kinds of district-determined measures. Technical Guide B will report on the results of the Department’s work with the field and continuing research and will offer more substantive guidance on defining high, moderate and low growth.

A guiding principle for that work is this: district-determined measures should help educators secure valuable feedback about student learning and growth so that they can use the feedback to adjust instruction and curriculum, along with their own leadership and teaching practices. The recommendations the Department will develop will be designed to support classroom and school use of assessment data.

Key Takeaways

- To rate an educator’s impact on student learning, districts must determine whether an educators’ students have achieved high, moderate, or low growth. Moderate growth is one year’s growth in one year’s time.
- One way to do this is to rate the growth of the median student an educator serves.
- More detailed guidance from ESE will be available in winter 2012–13.

**A potential approach for identifying high, moderate, and low growth
when pre- and post-assessment results exist**

First, at the district level:

- a) Compute, for each of the district's test takers, the difference between the pre-assessment results and the post-assessment results.
- b) List all district test-takers' improvement scores in order of highest to lowest.
- c) Divide the list into three equal bands, each representing one-third of the students. Identify the cut scores for each band.

Then, at the school or classroom level:

- d) For each class, list each student and his/her improvement score from highest to lowest.
- e) Divide each class list into three bands (high, moderate, and low) using the cut scores identified for each band at the district level.
- f) Compute for each class list the proportion of students who performed in each band, as well as the score of the educator's median (middle) student
- g) Identify which band the median student's improvement score falls in for each class and for each educator (in cases where the educator has taught more than one class).
- h) The results from step (g) determine student growth for that educator's students on the measure: students in the middle third can be considered to have achieved moderate growth; students in the top third can be considered to have achieved high growth; while students performing in the bottom third can be considered to have achieved low growth.

Rating Educator Impact on Student Learning

Identifying Trends and Patterns in Student Growth

In Massachusetts, an educator's impact on student learning is determined neither by a single year of data nor by a single measure of student learning. It must be based on a *trend* over time of at least two years; and it should reflect a *pattern* in the results on at least two different assessments. Patterns refer to consistent results from multiple measures, while trends require consistent results over at least two years.²³ Thus, identifying an educator's impact on student learning requires:

- **At least two measures each year.** Establishing a pattern always requires at least two measures during the same year. When unit tests or other assessments of learning over a period of time considerably shorter than a school year are used, consideration should be given to using more than two assessments or balancing results from an assessment of a shorter interval with one for a longer interval.
- **At least two years of data.** Establishing a trend always requires at least two years of data. Using the same measures over the two years is not required. For example, if a sixth-grade teacher whose typical sixth-grade student showed high growth on two measures that year transfers to eighth grade the following year, and her typical eighth-grade student that year shows high growth on both new measures, that teacher can earn a rating of high impact based on trends and patterns. That said, if an educator changes districts across years, his/her rating in the previous year cannot be used to construct a trend because of the confidentiality provisions of the regulations.

Rating Educator Impact on Student Learning

When arriving at an impact rating, ESE recommends that the rating of impact be based on the weight of evidence: What impact rating do the four (or more) data points suggest? While expected to exercise professional judgment in determining the impact rating, the evaluator should use this rule of thumb: If more than half of the data points point to the same impact, then that should be the educator's overall impact rating. When this is not the case, there is no clear pattern and trend, and ESE recommends that moderate impact be the default absent compelling evidence pointing to a different conclusion.

The charts below offer examples of applying this approach. The first example is for a eighth grade English language arts teacher whose measures of impact on student learning include the MCAS Student Growth Percentile on the grade 8 English language arts test, a writing assessment with a pre- and post-test, and a unit assessment on oral presentation skills. The first two are used as measures in both years, while the oral presentation assessment was added in the second year.

Year	Measure	Student Results
Year 1	MCAS SGP, grade 8 English language arts	High growth
Year 1	Writing assessment	Moderate growth
Year 2	MCAS SGP, grade 8 English language arts	High growth
Year 2	Writing assessment	Moderate growth
Year 2	Unit assessment on oral presentation skills	Moderate growth

²³ 603 CMR 35.02

Rating Educator Impact on Student Learning

Here, results from two district-determined measures used in Year 1 are combined with results from three district-determined measures used in Year 2, making a total of five data points. Three of the data points indicated moderate impact on student growth; two indicated high impact. Applying the rule of thumb, the evaluator concludes that the trend and pattern in student results for this educator points to a rating of impact on student learning of moderate, because 60 percent of the data points (three of five) indicated moderate impact. Had the student results on one of the measures been switched from moderate to high, then the rule of thumb would suggest an impact rating of high.

The second example is for a fifth-grade math specialist with two measures of impact on student learning: the MCAS student growth percentile on the grade 5 mathematics assessment and a unit assessment on fractions.

Year	Measure	Student Results
Year 1	MCAS SGP, grade 5 mathematics	Low growth
Year 1	Unit assessment on multiplication and division of fractions	Moderate growth
Year 2	MCAS SGP, grade 5 mathematics	Moderate growth
Year 2	Unit assessment on multiplication and division of fractions	Low growth

In this case, half of the data points indicate low growth, while half indicate moderate growth. No single rating appeared on more than half of the data points, so the rule of thumb points to a rating of moderate as the default when there is no clear trend or pattern.

To be clear, the rule of thumb is a guideline, not a requirement. As they determine an educator's impact on student learning, evaluators may take into consideration compelling evidence in circumstances where the data is mixed.

Key Takeaways

- In Massachusetts, the rating of an educator's impact on student learning must be based on a trend of two or more years of data and a pattern in the results across at least two different assessments.
- Evaluators are expected to use professional judgment when determining a rating where evidence is mixed. A rule of thumb is that if more than half the data points indicate a particular rating, that should be the educator's overall rating. Otherwise, the Department recommends that the rating defaults to moderate, absent compelling evidence otherwise.

Rating Educator Impact on Student Learning

Using the Impact Rating

In the new educator evaluation framework, educator impact ratings serve a number of purposes. They determine whether an experienced educator with Proficient or Exemplary summative performance ratings will be on a one- or two-year Self-Directed Growth plan. When there is a discrepancy between an educator's summative rating and impact rating, the impact rating serves as a spur to explore and understand the reasons for the discrepancy. Equally importantly, disaggregated student results on the district-determined measures that yield the impact rating can play a critical role in several steps in the 5-Step Cycle of evaluation. Finally, when combined with an educator's summative rating, impact ratings provide a basis for recognizing and rewarding Exemplary educators.

Placing Educators on an Educator Plan

Summative Performance Rating	Exemplary	1-YEAR SELF-DIRECTED GROWTH PLAN	RECOGNITION		
	Proficient		2-YEAR SELF-DIRECTED GROWTH PLAN		
	Needs Improvement	DIRECTED GROWTH PLAN			
	Unsatisfactory	IMPROVEMENT PLAN			
		Low	Moderate	High	
		Impact on Student Learning (rating based on trends and patterns in results of multiple measures of student performance, including MCAS Student Growth Percentile Scores if available)			

The impact rating affects the Educator Plan for educators with professional teacher status or administrators with more than three years of experience as administrators.

An experienced educator who receives a summative evaluation rating of Exemplary or Proficient and an impact rating that is either moderate or high is placed on a Self-Directed Growth Plan for two school years. The educator becomes "eligible for additional roles, responsibilities, and compensation, as determined by the district and through collective bargaining, where applicable."²⁴

An experienced educator who receives a summative evaluation rating of Exemplary or Proficient and an impact rating of low is placed on a Self-directed Growth Plan for one school year. The educator and evaluator are expected to "analyze the discrepancy in practice and student performance measures and

²⁴ 603 CMR 35.06(7)(a)(1)c

Rating Educator Impact on Student Learning

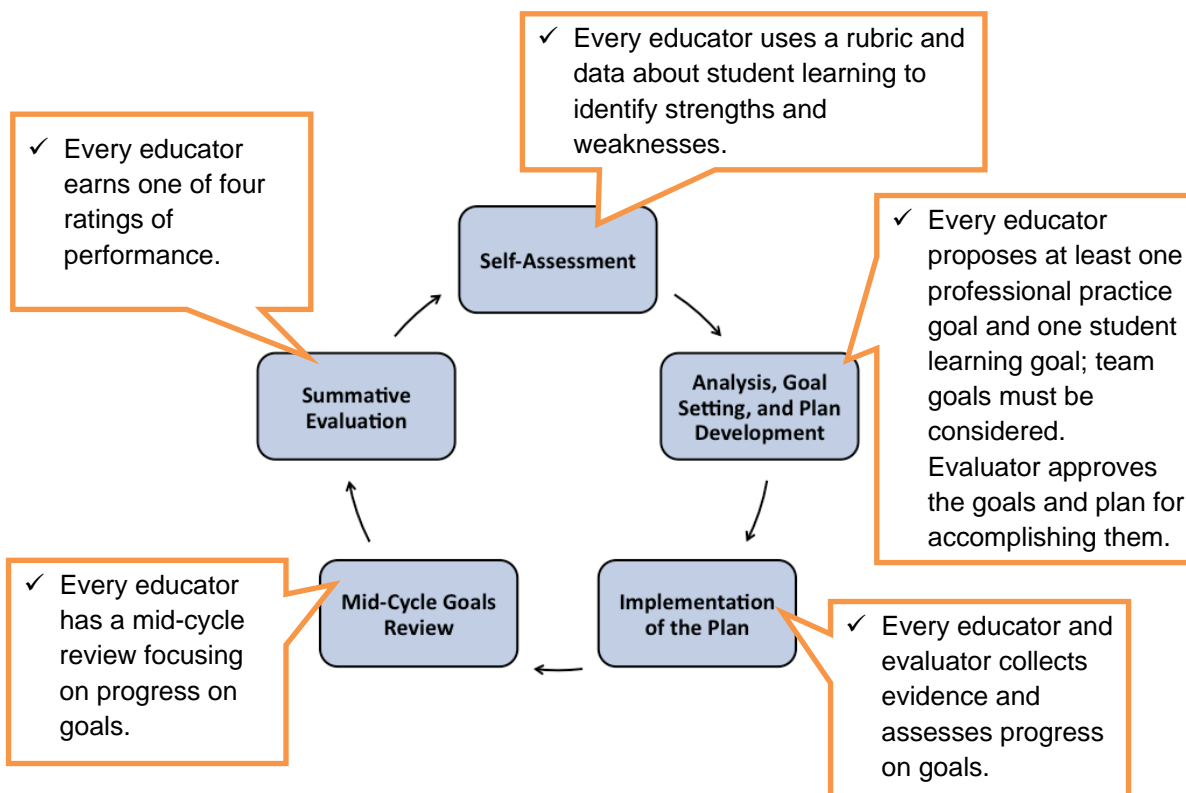
seek to determine the cause(s) of such discrepancy. The Educator Plan is expected to include one or more goals directly related to examining elements of practice that may be contributing to low impact.”²⁵

In addition, the regulations require that the evaluator's supervisor review the performance rating with the evaluator when a notable discrepancy occurs. When there are significant discrepancies between impact ratings and summative performance ratings, the evaluator's supervisor may note these discrepancies as a factor in the evaluator's own evaluation.²⁶ The supervisor is expected to look for consistency in application of Standards and Indicators as detailed in the district's rubrics, and evaluator's professional judgment. These checks and balances are designed to promote deeper inquiry and help foster continuous improvement through close attention to all three elements that determine evaluation ratings: educator practice, educator impact on student learning, and evaluator consistency and skill.

District-Determined Measures in the 5-Step Cycle of Evaluation

Student performance on district-determined measures has an important role to play in the 5-Step Cycle, as does the rating of educator impact. In the self-assessment and goal setting stages of the 5-Step Cycle, disaggregated results from the state and/or district-determined measures will help the educator assess his or her performance and target areas of practice that may need improvement. These results, along with the impact rating derived from those results, can also be used to help formulate goals to improve practice and student learning.

The 5-Step Cycle in Action



²⁵ 603 CMR 35.06(7)(a)(2)a and c

²⁶ 603 CMR 35.09

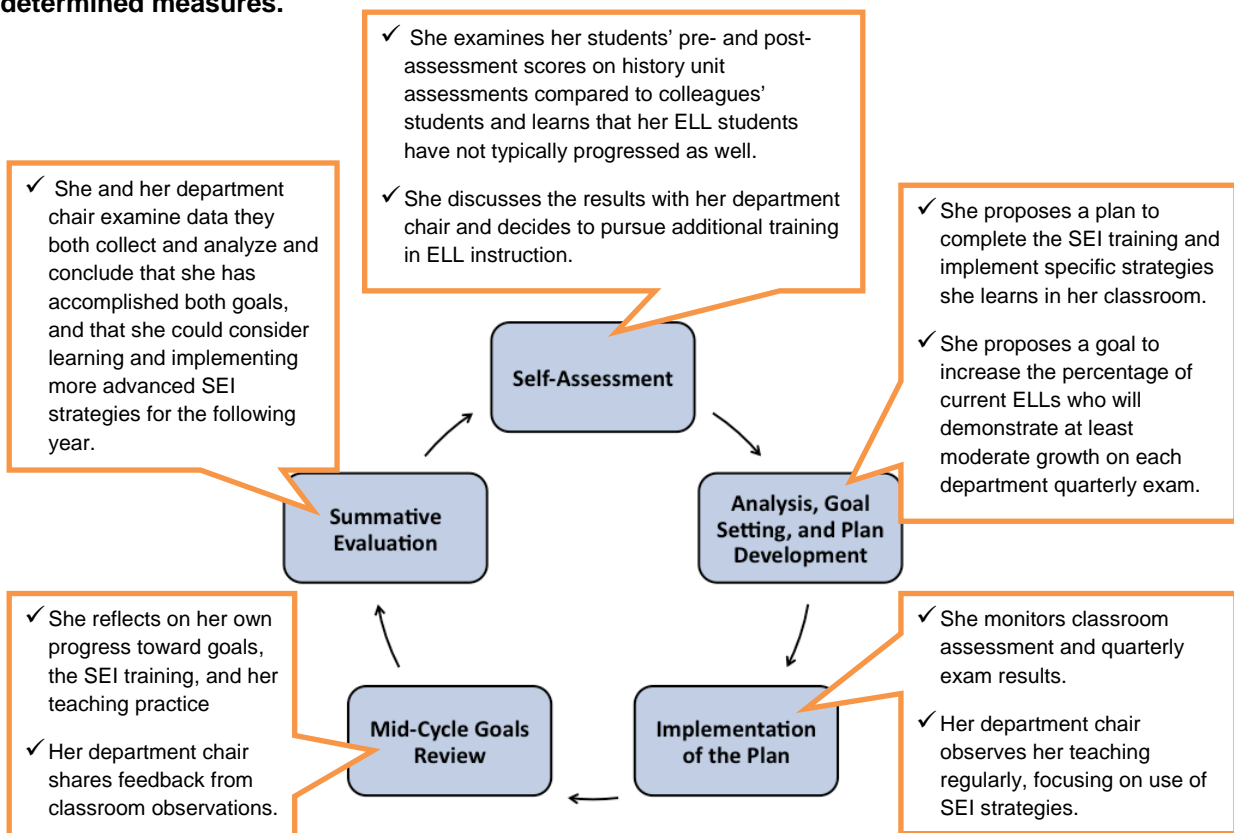
Rating Educator Impact on Student Learning

Below is an example of how the impact rating and results from district-determined measures can be valuable sources during the 5-Step Cycle.

Ms. Arletty is a ninth-grade social studies teacher whose summative rating has been Proficient and whose rating of impact on student learning has been moderate. She is on a two-year Self-Directed Growth Plan. She and her colleagues administer common unit assessments linked to the Massachusetts curriculum frameworks and a common year-end assessment in the form of a capstone paper that requires students to demonstrate mastery of the core concepts covered throughout the year and critical analysis and persuasive writing skills. The capstone papers address key concepts and skills that are shared district-wide and are assessed against a common rubric. A district-wide eighth-grade capstone project assessing comparable skills provides baseline data for each student.

In reviewing student data during her self-assessment, Ms. Arletty notices a pattern of less improvement among the English language learners (ELLs) in her class when compared to other ELL students in the ninth grade. Ms. Arletty recognizes that she has limited experience teaching ELLs. She decides to set a professional practice goal of implementing more effective strategies for sheltering English in her classroom. She also creates a student learning goal that focuses on her English language learners' performance on the department's quarterly exams.

Ms. Arletty: A ninth-grade history teacher whose students show moderate growth on district-determined measures.



Note that the combination of assessment formats and content addressed in the common assessments for Ms. Arletty provides a rich source of data on student performance that inform other goals or steps in the 5-Step Cycle of evaluation, as well.

Rating Educator Impact on Student Learning

Student results on district-determined measures and the ratings they yield can be powerful tools for drilling down on what is working well, and what can be improved upon in both practice and student performance. Both are designed to inform self-assessment, collaborative inquiry, conversation, and action, resulting in an Educator Plan with goals that lead to improvement in professional practice and student learning.

Recognizing Educators With High Impact on Student Learning

By the time their first educators are rated as having a high, moderate, or low impact on student learning, districts will need to identify the way(s) they will recognize and reward educators “whose summative performance rating is exemplary and whose impact on student learning is rated moderate or high.” They are to be “recognized and rewarded with leadership roles, promotion, additional compensation, public commendation or other acknowledgement.”²⁷

To aid districts in this work, ESE is currently developing performance-based licensure endorsements—called the Career Ladder endorsements—with the intent of providing teachers with career growth opportunities while simultaneously providing them the opportunity to stay in the classroom. Possible endorsement titles include mentor, instructional leader, and data specialist. Regulations will be brought to the Board of Elementary and Secondary Education (BESE) in the 2012-2013 school year, with an expected launch in the 2013-2014 school year. With the advent of the Career Ladder endorsements, districts will have new tools for recognizing and rewarding educators “whose summative performance rating is exemplary and whose impact on student learning is rated moderate or high” in order to use their expertise to strengthen teaching and learning beyond a single classroom.

Key Takeaways

- The rating of educator impact on student learning, along with the summative performance rating, affects which type of educator plan an educator is on.
- Where there is a discrepancy between the student impact rating and the performance rating, the impact rating serves as a spur to explore and understand the reasons for the discrepancy.
- Disaggregated student results on the district-determined measures that yield the impact rating can also play a critical role in several steps in the 5-Step Cycle of evaluation.
- When combined with an educator’s summative rating, impact ratings provide a basis for recognizing and rewarding exemplary educators.

²⁷ 603 CMR 35.08(7)

Rating Educator Impact on Student Learning

Looking Ahead: Reporting the Rating of Impact to ESE

Educator evaluation ratings will be collected in the Education Personnel Information Management System (EPIMS) beginning with ratings for the 2012–13 school year for some educators in Race to the Top districts. Ultimately, all districts will report seven data elements annually:

- 1) overall summative evaluation or formative evaluation rating
- 2) rating on Standard I
- 3) rating on Standard II
- 4) rating on Standard III
- 5) rating on Standard IV
- 6) rating of Impact on Student Learning
- 7) professional teacher status

Details of reporting requirements will be made available by fall 2012.

Identifying and Beginning to Address District Capacity and Infrastructure Needs

The new educator evaluation regulations are ushering in a variety of changes to how educators are evaluated, impact on learning is assessed, feedback on performance is provided, and professional development is pursued. To take full advantage of these changes, districts will need to build upon, and in some instances, rethink many of their processes, systems, and infrastructure for managing human resources, data collection, and assessment.

Districts participating in Race to the Top are establishing labor-management evaluation implementation working groups, consisting of district administrators for curriculum and instruction, assessment, human resources, and teams of role-alike educators. These working groups are the starting point for assessing human resource and technology capacities and considering the steps required to implement the new regulations in ways that will support professional growth and student learning. Decisions and processes will be needed to address these and other areas:

- Attribution and responsibility level classifications;
- Roster verification (including guidelines for student mobility, tardiness and/or chronic absenteeism; educator attendance, etc.)
- Determining trends and patterns;
- Determining high, moderate, or low growth on district-determined measures;
- The role of evaluator judgment;
- When and how to use school-wide combined results for an educator's impact rating;
- Dealing with educators serving small numbers of students.

The Department plans to work with stakeholders over the next year to develop recommended policies and practices for these and other steps. Districts can use these recommendations and resources as starting points for their own work.

Identifying and Addressing District and School Needs

Educator training

Districts should introduce the concept of using district-determined measures in educator evaluation to educators before they engage in the process of assessing, selecting, adapting, developing, or implementing measures. To support this work, the ESE-developed training module series will include an introduction for school leadership teams to district-determined measures and rating educator impact.

Depending on the assessment choices a district makes, different training elements will be more or less important. For example, in grades using portfolios and/or capstone projects as a measure, educators may need to develop their understanding of required components and the use of common rubrics. Processes for calibrating evaluator judgments on a performance task may need to be learned and established to ensure consistency in scoring.

Addressing District Capacity and Infrastructure Needs

Evaluator Training and Ongoing Support

Evaluators need access to quality training and support if they are to lead effective implementation in their districts, schools, and departments. Evaluators should be trained in the application of measures, the district's protocols for attribution and roster verification, the role of educator judgment in the process of determining the educator impact rating, and how to recognize significant discrepancies and know what to do about them. Evaluators will need to learn how the MCAS Student Growth Percentile works and what to do when the numbers of test-takers is small. They may need additional professional development to assist them in collaborating with educators to use results of district-determined measures to strengthen goal setting and plan development and providing them with pertinent suggestions for high quality professional development targeted to identified needs.

Technology

The new regulations will require that additional data be collected, stored, and analyzed to facilitate the determination of educator impact ratings. Districts will need systems for:

- Tracking student performance on district-determined measures over time;
- Roster verification and attribution;
- Monitoring and analyzing long-term trends;
- Monitoring for patterns of discrepancy between summative and impact ratings

Technology tools can dramatically enhance these efforts. To do so, current district data and reporting systems may need to be adapted or enhanced. At a minimum, it is likely that new reports and methods of storing and analyzing data will be needed. ESE expects to support networking efforts that will enable districts to share promising tools and approaches for enhancing their effective use of technology to support educator evaluation.

Accessing More Resources From ESE

Professional Development

ESE is completing development of a series of training modules to prepare school leadership teams to implement the new Massachusetts educator evaluation system in their schools. Each training module is three hours in length and includes interactive learning activities for school leadership teams. School leadership teams are encouraged to take each module back to their school for the professional development of their school staff. Suggested homework assignments described at the conclusion of each module are intended to help participants extend and apply their learning and are designed to take about an hour. Modules are described here: <http://www.doe.mass.edu/eval/implementation/modules/>.

The module on Rating Impact on Students will be released during the 2012–13 school year and will provide an overview of this guidance on district-determined measures of student learning and ratings of educator impact on student learning. The module is designed to help school leadership teams develop a common understanding of the student impact rating and how it fits together with the 5-Step Cycle. Participants will also receive implementation tips and strategies designed to help schools utilize state assessments and district determined measures when determining each educator's impact rating.

Addressing District Capacity and Infrastructure Needs

Exemplar District-Determined Measures

As described earlier, ESE will be working with practitioners and assessment experts to identify, develop, and pilot district-determined measures in the arts, English language arts, mathematics, physical education/health, science, technology and engineering, social studies/history, foreign languages, and vocational and business education. The Department will keep districts informed of opportunities to participate in this effort and to use the results, once available.

Multi-State Item Bank of Assessment Questions

As part of its Race to the Top Grant, ESE received funds to develop a teaching and learning system. The system, called Edwin, will first and foremost be a tool for educators to improve instruction. It will include a variety of resources, such as digital vignettes of practice, an item bank of assessment questions built from released MCAS questions tagged to common core standards, and a test builder engine that can be used to create interim and formative assessments from the item bank.

ESE is in conversation with a number of other states to create a common resource of assessment items from across the country. If successful, this effort will result in a vastly enhanced assessment resource for Massachusetts educators, a deeper pool of assessment items to draw upon from a wider array of subjects and grades than those currently included in prior year MCAS test item releases. For example, New York's involvement would make available assessment items from its Regents exams, which include tests in a variety of world languages, and discrete subjects in the sciences, history and social studies, and mathematics. Georgia and Rhode Island have been developing exemplars of Student Learning Objectives that could be used by special educators who work with students with severe disabilities.

This item bank will enable districts to develop measures in a variety of subjects and grades that could potentially be used as district-determined measures. A pilot of the multi-state item bank is planned for 2012–13. ESE will explore its usefulness as resource with the content specialists and districts that will be identifying, reviewing, and potentially developing and piloting the exemplar district-determined measures during 2012–13.

Curriculum-Embedded Performance Assessments

Teachers and other educators are working with curriculum specialists and others under the Massachusetts Race to the Top grant to develop curriculum units aligned with the most up-to-date Massachusetts curriculum frameworks. This effort will yield nearly 100 curriculum units with lesson plans, and curriculum-embedded performance assessments spanning English language arts, mathematics, science, and history/social science in grades prekindergarten–1,. The first units are expected to be available during 2012–13 for some districts to begin piloting. Districts may consider using these as starting points for district-determined measures.

MA Model Curriculum Units at a Glance

The development of Model Curriculum Units involves teams of teachers delving deeply into the MA Curriculum Frameworks to identify and target groups of standards. Using the standards, the teachers chosen to serve on these teams are engaging in rich discussions to define the following elements of high quality curriculum design:

- Transfer Goals
- Enduring Understandings
- Essential Questions
- Knowledge
- Skills
- Curriculum Embedded Performance Assessments (CEPA)
- Other forms of assessment e.g. formative, summative, quick writes, observation, running records, etc.
- Learning Plan
- Lesson plans

The lesson plans illustrate a well-structured, ordered approach to instruction, including text and digital resources. Moreover, lesson plans include elements of lesson study through which teachers are expected to reflect upon the successes and challenges of each lesson

Each unit includes curriculum-embedded performance assessments aligned with the targeted standards and designed to elicit rich evidence of student learning. Analysis of this data in conjunction with other types of assessment will serve to inform instructional next steps.

All participating educators will actively engage in a pilot of the units. Part of this process will include the collection of feedback on the curriculum and instructional components of the unit, collection of student data, collaboration with colleagues, and adjustments to the unit.

Planning Collective Bargaining

The procedures for conducting educator evaluation are a mandatory subject of collective bargaining in Massachusetts.²⁸ As indicated in the District-Level Guide published in January 2012, all districts will be engaging in collective bargaining in order to implement the framework for teachers, caseload educators and administrators represented by bargaining agents.

Many of the early adopter districts see the new framework as a welcome opportunity for labor and management to engage deeply and constructively in the conversation, collaboration and negotiation required to establish a sound foundation for implementing new practices. They understand that formal negotiations are only one step in a much longer process of collaboration that will be needed to build, monitor, update, and revise an educator evaluation process that is fair, transparent, credible, and leads to educator growth and development.

ESE will continue to work with Level 4 and early adopter districts and other stakeholders to learn more about implementation challenges, and will work with state associations representing teachers unions, superintendents and school committees to draft model contract language addressing rating educator impact on student learning by summer 2013.

That said, the Department again urges districts to conduct bargaining in a way that permits the parties to return to educator evaluation periodically over the next several years. Unlike the regulations promulgated in 1995, these educator evaluation regulations explicitly envision modification through phased implementation and ongoing learning from the field. Districts may want to consider the use of side letters, memorandum of understanding, "re-opener clauses," and other mechanisms for facilitating the work that lies ahead.

Key Takeaways

- Districts should plan for providing training for educators and evaluators on the new evaluation framework, and potentially also for technology support.
- ESE will make available training materials on the evaluation framework, as well as assessment tools that can be used as the basis for district-determined measures.
- ESE will work with state associations to draft model contract language addressing rating educator impact on student learning by summer 2013.

²⁸ M.G.L. c 71 s 38. See Appendix E.

Immediate Next Steps

As indicated in earlier sections, the focus of the initial work on rating educator impact on student learning will be on identifying and exploring possible district-determined measures of student learning. That said, important groundwork can be laid now to support future efforts. Steps to consider include:

- 1) Identify a team of administrators, teachers and specialists to focus and plan the district's work on district-determined measures.
- 2) Consider having one or more educators apply to serve on one of ESE's content specialist teams charged with identifying, revising, and/or developing exemplary district-determined measures.
- 3) Assess educators' understanding of the basics of how the MCAS Student Growth Percentile is derived and how it can be used to understand student growth and progress; develop a plan for ensuring educator understanding.
- 4) Complete an inventory of existing assessments used in the district's schools.
- 5) Assess where there are strengths to build on and gaps to fill.
- 6) Discuss with the district's educational collaborative its interest and capacity to assist member districts in the work of identifying and assessing assessments that may serve as district-determined measures.
- 7) Consider submitting to ESE in 2012–13 a locally developed district-determined measure for review and feedback by the appropriate ESE-supported statewide content and assessment team.
- 8) Plan a process for piloting district-determined measures.
- 9) Create (or augment) the district's communication plan to ensure that educators, school board members parents, and other stakeholders understand the role that district-determined measures will play in the new evaluation framework as well as their timetable for implementation.

Developing district-determined measures is an opportunity for districts to clarify their instructional priorities, align their measures of growth to those priorities, and expand the range and quality of assessments used. But most importantly, done well, the implementation process can create greater coherence district-wide in instructional improvement efforts. Careful thought in the early phases of implementation to these considerations will benefit the district for years to come.

Appendix A. What the Regulations Say

Excerpts from 603 CMR 35.00 (Evaluation of Educators) Related to District-Determined Measures and Rating Educator Impact on Student Learning

Rating Educator Impact on Student Learning

35.09 Student Performance Measures

- 1) Student Performance Measures as described in 35.07(1)(a)(3-5) shall be the basis for determining an educator's impact on student learning, growth and achievement.
- 2) The evaluator shall determine whether an educator is having a high, moderate, or low impact on student learning based on trends and patterns in the following student performance measures:
 - a) At least two state or district-wide measures of student learning gains shall be employed at each school, grade, and subject in determining impact on student learning, as follows:
 - i) MCAS Student Growth Percentile and the Massachusetts English Proficiency Assessment (MEPA) shall be used as measures where available, and
 - ii) Additional district-determined measures comparable across schools, grades, and subject matter district-wide as determined by the superintendent may be used in conjunction with MCAS Student Growth Percentiles and MEPA scores to meet this requirement, and shall be used when either MCAS growth or MEPA scores are not available.
 - b) For educators whose primary role is not as a classroom teacher, appropriate measures of their contribution to student learning, growth, and achievement shall be determined by the district.
- 3) Based on a review of trends and patterns of state and district measures of student learning gains, the evaluator will assign the rating on growth in student performance consistent with Department guidelines:
 - a) A rating of high indicates significantly higher than one-year's growth relative to academic peers in the grade or subject.
 - b) A rating of moderate indicates one year's growth relative to academic peers in the grade or subject.
 - c) A rating of low indicates significantly lower than one year's student learning growth relative to academic peers in the grade or subject.
- 4) For an educator whose overall performance rating is exemplary or proficient and whose impact on student learning is low, the evaluator's supervisor shall discuss and review the rating with the evaluator and the supervisor shall confirm or revise the educator's rating. In cases where the superintendent serves as the evaluator, the superintendent's decision on the rating shall not be subject to such review. When there are significant discrepancies between evidence of student learning, growth and achievement and the evaluator's judgment on educator performance ratings, the evaluator's supervisor may note these discrepancies as a factor in the evaluator's evaluation.

Evidence Used (35.07)

(1)(a)3-5

- 3) Statewide growth measure(s) where available, including the MCAS Student Growth Percentile and the Massachusetts English Proficiency Assessment (MEPA); and
- 4) District-determined Measure(s) of student learning comparable across grade or subject district-wide.
- 5) For educators whose primary role is not as a classroom teacher, the appropriate measures of the educator's contribution to student learning, growth, and achievement set by the district.

Definitions (35.02)

...**District-determined** Measures shall mean measures of student learning, growth, and achievement related to the Massachusetts Curriculum Frameworks, Massachusetts Vocational Technical Education Frameworks, or other relevant frameworks, that are comparable across grade or subject level district-wide. These measures may include, but shall not be limited to: portfolios, approved commercial assessments and district-developed pre and post unit and course assessments, and capstone projects.

...**Impact on Student Learning** shall mean at least the trend in student learning, growth, and achievement and may also include patterns in student learning, growth, and achievement.

...**Patterns** shall mean consistent results from multiple measures.

...**Trends** shall be based on at least two years of data.

Basis for the Rating of Impact on Student Learning

35.07(2) Evidence and professional judgment shall inform:...(b) the evaluator's assessment of the educator's impact on the learning, growth and achievement of the students under the educator's responsibility.

Consequences

35.06(7) (a)(2) For the (teacher with professional teacher status or administrator with more than three years in a position in a district) whose impact on student learning is low, the evaluator shall place the educator on a Self-directed Growth Plan.

- a) The educator and evaluator shall analyze the discrepancy in practice and student performance measures and seek to determine the cause(s) of such discrepancy.
- b) The plan shall be one school year in duration.
- c) The plan may include a goal related to examining elements of practice that may be contributing to low impact.
- d) The educator shall receive a summative evaluation at the end of the period determined in the plan, but at least annually.

35.08(7) Educators whose summative performance rating is exemplary and whose impact on student learning is rated moderate or high shall be recognized and rewarded with leadership roles, promotion, additional compensation, public commendation or other acknowledgement. (as determined through collective bargaining).

Timetable and Reporting

35.11

- 4) By September 2013, each district shall identify and report to the Department a district-wide set of student performance measures for each grade and subject that permit a comparison of student learning gains.
 - a) The student performance measures shall be consistent with...35.09(2)
 - b) By July 2012, the Department shall supplement these regulations with additional guidance on the development and use of student performance measures.
 - c) Until such measures and identified and data is available for at least two years, educators will not be assessed as having high, moderate, or low impact on student learning outcomes...
- 5) Districts shall provide the Department with individual educator evaluation data for each educator...including...(c) the educator's impact on student learning, growth, and achievement (high, moderate, low)

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

Appendix B. Technical Guide A (District-Determined Measures)

Preliminary Table of Contents

- 1) Focusing on Growth
 - a) What Makes a Good Measure?
 - b) Measuring Growth with District-Determined Measures
- 2) Determining Measurement Needs
 - a) Linking Measures to the New Curriculum Frameworks
 - b) Assessing Social-emotional Learning
 - c) Types of Measures: Advantages and Disadvantages
 - i) Multiple Choice
 - ii) Constructed or Open Response
 - iii) Performance Assessments
 - iv) Portfolios
 - v) End-of-Year and End-of-Course Assessments
 - vi) Interim or Benchmark Assessments
 - vii) Unit Assessments
 - viii) Other
 - d) Examples of Existing Measures
 - e) Matching District-Determined Measures to Educator Roles
- 3) Selecting Appropriate Measures
 - a) Choosing an Existing State, Local, or Commercially Available Measure
 - b) Building a New Measure
 - c) Building Portfolio Assessments
 - d) Piloting District-Determined Measures

Appendix C. Technical Guide B (Rating Educator Impact on Student Learning)

Preliminary Table of Contents

- 1) Matching Students to Educators
 - a) Attribution
 - i) Responsibility Levels
 - ii) Policies for Attribution: Recommendations
 - (1) Highly Mobile Students and Chronic Student Absence
 - (2) Long-term Educator Absence
 - (3) Student Schedule Changes
 - (4) Educators Linked to Small Numbers of Students
 - (5) Other
 - b) Roster Verification: Educator and Evaluator Responsibilities
 - i) Elements of a Strong Roster Verification System
- 2) Deciding on a Rating of Educator Impact Based on Trends and Patterns
 - a) Determining High, Moderate and Low Growth
 - i) Growth over Full Year and Shorter Intervals
 - b) Identifying Trends and Patterns
 - c) Combining Results to Determine Trends and Patterns
- 3) The Impact Rating and the 5-Step Cycle of Evaluation
- 4) Reporting Ratings of Educator Impact to ESE

Appendix D. Looking Ahead: Attribution and Roster Verification

i. Attribution

To rate educators' impact on student learning, districts will need to determine how to attribute students to the educators who serve them. A potential approach to attribution is one that categorizes each student-educator link into one of three levels of responsibility—primary, shared or limited—based on the level of interaction between the student and educator in the subject area being measured. Below are brief descriptions of each of the responsibility levels as they would apply to teachers.²⁹

- **Primary Responsibility** indicates that the teacher is principally responsible for that student's learning in the content or standards being assessed.
- **Shared Responsibility** indicates that the teacher is partially but not fully responsible for that student's learning in the content or standards being assessed, i.e., that responsibility is distributed among two or more educators.
- **Limited Responsibility** indicates that the teacher has insufficient responsibility for the student's learning for that student's assessment results to be used in determining the impact rating. The student, for example, may have been chronically absent or arrived well after the start of the unit, semester, or year.

An emerging best practice is that educators who share responsibility for specific students should share full responsibility for these students' growth. In other words, assessment results for students for whom an educator has primary responsibility should count equally with results from students for whom the educator has shared responsibility. This approach recognizes that teaching is often a collaborative endeavor and team members' efforts to support shared students are inextricably linked, whether the scenario involves team teaching, pull-out resource supports, or flexible grouping.

Data should be examined during the self-assessment and goal setting steps in the 5-Step Cycle to determine if there are differences in growth between students for whom the educator had primary responsibility and those s/he had shared responsibility. Differences may be worthy of further inquiry and could be important data in identifying goals.

ii. Roster Verification

To ensure that the results upon which trends and patterns are identified are accurate, it is critical that educators have an opportunity to review and confirm the roster of students whose assessment outcomes will comprise their impact rating. For example students for whom the educator has limited responsibility need to be identified and their assessment results removed. This process of ensuring that information on the educator-student link is accurate and complete is vital for maintaining the integrity and credibility of the new educator evaluation system. A strong roster verification process is important for another reason: to ensure that every student is the primary or shared responsibility of at least one educator in the school.

²⁹ Descriptions and examples that apply to administrators, instructional specialists and professional support specialists will be provided in Technical Guide B.

A strong roster verification system should:

- Integrate with ESE data and reporting standards for SIMS, EPIMS, SCS and other systems where applicable, for efficiency and accuracy
- Identify educators using a Massachusetts Education Personnel Identifier (MEPID)
- Account for all educators serving in a position that requires a license, including those who are working under a waiver
- Account for all students attending a school for any amount of time during the school year
- Account for the contributions of multiple educators in a single course
- Capture meaningful differences in instructional time, e.g., a middle school art teacher in one school who teaches students once a week and another who teaches students three times a week.
- Reflect educator and student course and schedule changes throughout the school year
- Enable an educator to review his/her roster for accuracy

Technical Guide B will provide more detailed guidance on the elements of a strong roster verification process, including technical information on online systems of roster verification. ESE will also offer recommendations for such limitations on the use of data as the minimum length of time a student needs to have been in a class or school in order for his/her assessment results to factor into the educator's impact rating.

Appendix E. Educator Evaluation and Collective Bargaining

Excerpts from M G.L. c. 71, § 38.

The superintendent, by means of comprehensive evaluation, shall cause the performance of all teachers, principals, and administrators within the school district to be evaluated using any principles of evaluation established by the board of education pursuant to section one B of chapter sixty-nine and by such consistent, supplemental performance standards as the school committee may require, including the extent to which students assigned to such teachers and administrators satisfy student academic standards or, in the case of a special education student, the individual education plan, and the successful implementation of professional development plans required under section thirty-eight Q; provided, however, that such principles and standards be consistent with the anti-discrimination requirements of chapter one hundred and fifty-two B. The superintendent shall require the evaluation of administrators and of teachers without professional teacher status every year and shall require the evaluation of teachers with professional teacher status at least once every two years. The procedures for conducting such evaluations, but not the requirement for such evaluations, shall be subject to the collective bargaining provisions of chapter one hundred and fifty E.

Performance standards for teachers and other school district employees shall be established by the school committee upon the recommendation of the superintendent, provided that where teachers are represented for collective bargaining purposes, all teacher performance standards shall be determined as follows: The school committee and the collective bargaining representative shall undertake for a reasonable period of time to agree on teacher performance standards. Prior to said reasonable period of time, the school district shall seek a public hearing to comment on such standards. In the absence of an agreement, after such reasonable period, teacher performance standards shall be determined by binding interest arbitration. Either the school district or the teachers' collective bargaining representative may file a petition seeking arbitration with the commissioner of education. The commissioner shall forward to the parties a list of three arbitrators provided by the American Arbitration Association. The school committee and the collective bargaining representative within three days of receipt of the list from the commissioner of education shall have the right to strike one of the three arbitrators' names if they are unable to agree upon a single arbitrator from among the three. The arbitration shall be conducted in accordance with the rules of the American Arbitration Association to be consistent with the provisions of this section. In reaching a decision, the arbitrator shall seek to advance the goals of encouraging innovation in teaching and of holding teachers accountable for improving student performance. The arbitrator shall consider the particular socioeconomic conditions of the student population of the school district. Both the parties and the arbitrator may adopt performance standards established by state or national organizations. The performance standards shall be incorporated into the applicable collective bargaining agreement; provided, however, that any subsequent modification of the performance standards shall be made pursuant to the procedures set forth in this section.