Technical Guide B:
Measuring Student Growth & Piloting District-Determined Measures

September 2013

Contents

Appendices

Section 1. Introduction

Overview of Technical Guide B

Massachusetts Educator Evaluation Framework

Massachusetts's educator evaluation system employs two independent but connected ratings to identify the intersection of educator practice and student impact:

- The *Summative Performance Rating* is the final step of the 5-step evaluation cycle designed to place educators in a more central role in the evaluation process. This rating assesses an educator's practice against four statewide Standards of Effective Teaching Practice or Administrator Leadership, as well as an educator's progress toward attainment of his/her goals.

- The *Student Impact Rating* is separate but complementary to the Summative Performance Rating; it is informed by trends (at least two years) and patterns (at least two measures) in student growth as measured by statewide growth measures, where available, and district-determined measures (DDMs).

Technical Guide B addresses the measures used to determine the Student Impact Rating. No single measure of effectiveness will determine an educator's Student Impact Rating. Multiple measures (at least two per educator) across multiple years (at least two years) are required to determine the educator's Student Impact Rating.

In turn, the Student Impact Rating is juxtaposed on the educator's Summative Performance Rating to determine the focus and duration of the educator's growth or improvement plan. The system is designed to triangulate evidence across multiple measures–including measures of impact on student learning and measures of educator practice–thus maximizing the identification of opportunities for growth and development. As a result of this design, no single measure by itself presents high stakes to the educator. Rather, it is the triangulation of evidence across the various measures that is significant.

> **District-Determined Measures**
>
> Per 603 CMR 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.

ESE Guidance

In August 2012, the Massachusetts Department of Elementary and Secondary Education (ESE) published Part VII of the Massachusetts Model System for Educator Evaluation, "*Rating Educator Impact on Student Learning Using District-Determined Measures of Student Learning, Growth, and Achievement,*" which introduces the use of DDMs in educator evaluation. In April 2013, ESE released *Technical Guide A,* a supplement to Part VII, which provided guidance to districts on how to evaluate and identify DDMs.

The following document, *Technical Guide B,* builds on both Part VII and Technical Guide A. However, this guide serves a different purpose than Technical Guide A. Here, Technical Guide B focuses on the piloting of DDMs and shares concrete examples that demonstrate how to incorporate key assessment concepts in measuring impact on student learning, growth, and achievement. Reading this guide will prepare districts to:

- select and pilot meaningful and informative DDMs that measure student growth; and

- identify clear next steps for piloting DDMs during the 2013-2014 school year.

The guide is designed to be timely, practical, and highly accessible. Thus, it does not present statistical methods for developing local growth models. ESE is happy to collaborate with districts that are ready to explore this level of technical development. We welcome you to contact us at EducatorEvaluation@doe.mass.edu.

- Section 1 offers an overview and background information about the goal of measuring impact on student learning.

- Section 2 discusses key questions and considerations for selecting and piloting DDMs.

- Section 3 describes approaches to measuring student growth with specific examples.

- Section 4 outlines clear steps for piloting DDMs.

- Section 5 provides an update on available and upcoming resources from ESE.

- Appendices to the guide provide additional information and resources. The appendices include:

  - Appendix A includes Commissioner Chester's August 2013 memo outlining supports, the timeline, and pilot expectations for DDMs.

  - Appendix B provides a deep dive into each suggested pilot step.

  - Appendix C includes links to resources for modifying and scoring DDMS.

  - Appendix D shares district examples of educator professional practice S.M.A.R.T. goals that focus on DDM work.

  - Appendix E is an example of criteria used for identifying content to measure.

  - Appendix F is a sample survey used for gathering information from educators on existing assessments.

  - Appendix G provides a quick overview of validity and reliability.

Measuring Impact on Student Learning in Educator Evaluation

From the earliest stages of the development of the new Educator Evaluation framework, educators have agreed that student learning must be a central part of the evaluation process if it is to play a meaningful role in improving the effectiveness of educators. In the words of the 40-member task force that developed and proposed the framework, "Educator effectiveness and student learning, growth and achievement are inextricably linked."[1] Specifically, the ability to improve educator effectiveness relies on having meaningful and accurate information about student progress.

Student learning has always been at the heart of education: educators across the Commonwealth ask themselves every day, "What did my students learn? How much did they learn? And how do I know?" District-determined measures (DDMs) ensure that all educators have meaningful, timely information to assess student learning. ESE's approach to DDMs builds on work that most teachers and many districts already do: use measures (e.g., tests, assignments, tasks, portfolios) to evaluate student performance and growth, as well as to calibrate learning from one class to another. The statewide shift to using measures that are comparable across grades/subjects or courses within and across schools creates a powerful opportunity for collaborative conversations about effective instructional practices.

Piloting DDMs During the 2013-14 School Year

As an effort to foster collaboration at state and local levels, all districts will be piloting at least one potential DDM in the same five areas during the 2013-14 school year. The required minimum pilot areas are:

- Early grade (K-3) literacy
- Early grade (K-3) math
- Middle grade (5-8) math
- High school writing to text
- Traditionally non-tested grades and subjects (e.g., fine arts, music, p.e.)

ESE carefully selected these five minimum pilot areas to align to high priority statewide goals and initiatives. In support of this challenging work, ESE will provide support and technical assistance to help districts integrate implementation of the new educator evaluation system and implementation of the shifts in curriculum and instruction embedded in the new ELA and math Curriculum Frameworks. In particular, ESE will make available a series of resources, technical assistance, and professional development aimed at helping districts build comprehensive plans for integrating these two key initiatives, with a focus on the five required DDM pilot areas. For more information, please refer to the full text of Commissioner Chester's memorandum ,"District-Determined Measures: Supports, Expectations, and Timelines" (Appendix A).[2]

---

[1] "Building a Breakthrough Framework for Educator Evaluation in the Commonwealth," March 2011 available at http://www.doe.mass.edu/edeval/breakthroughframework.doc.

[2] Also available at http://www.doe.mass.edu/news/news.aspx?id=7640.

The goal of measuring impact on student learning is straightforward. It requires that districts:

- clarify the important learning objectives for each grade and subject;
- identify desired learning gains at each grade and subject;
- identify methods of measuring student learning; and
- distinguish between high, moderate, and low learning gains.

Including a measure of impact on student learning in educators' evaluation requires significant cultural shifts, for many, from how evaluations were once conducted. The shifts include:[3]

1. **It is no longer just about the teaching. It is also about the learning.**

   This focus on the "learning" refers to determining the impact that educators have on student learning. It is important, however, to remember why: data on student learning is feedback to inform teaching. The incorporation of student learning should always bring an equally important focus on the opportunity for educator learning.

2. **It is not just about the effort. It is also about the results.**

   Many educators have had the experience of looking at student results on a test or samples of student work, and thinking with surprise, "I thought they *got* that!" While frustrating, these moments are critically important as they shine a light on the gaps between effort and results. In making student learning central to educator evaluation, the difference between effort and results becomes part of the conversation between educators and evaluators. This may be new for many educators and not always comfortable. It takes courage and commitment for both educators and evaluators to engage in these conversations. It is this shared examination of educator practice as it relates to student results that makes evaluation a meaningful and effective process for accelerating the professional growth of educators and student learning.

This guide supports districts to meet the goal of measuring impact on student learning by first introducing key questions and considerations for selecting and piloting measures (Section 2). It then describes approaches to measuring growth with examples of assessment types (Section 3). Finally, the guide provides a brief overview of recommended steps for piloting measures (Section 4) which are then discussed in more detail in Appendix B.

> As this guide focuses on supporting districts to meet the piloting requirements for the 2013-14 school year, the examples and explanations are designed to be highly relevant to the five minimum pilot areas. That said, the concepts and suggested steps in this guide are also applicable to DDMs for administrators and specialized instructional support personnel (SISPs). ESE will continue working to identify examples, share best practices, and provide resources and technical assistance for measuring the impact of educators in grades/subjects and courses outside of these five areas, including administrators and SISPs.

---

[3] Descriptions of the cultural shifts are adapted from remarks made by Karla Baehr, former superintendent and Deputy Commissioner at ESE, at the MASS Summer Conference on July 18, 2013.

Section 2. Key Questions and Considerations

Assessments that are valuable to educators measure, as fairly and accurately as possible, the extent to which students have learned the most important content and skills taught and yield data that can then be used to inform instruction. Two fundamental questions should be the guideposts for districts as they choose DDMs as a part of their assessment of student learning:

1. **Is the measure aligned to content?**[4]

   ▪ Does it assess what is most important for students to learn and be able to do?

   ▪ Does it assess what the educators intend to teach?

2. **Is the measure informative?**

   ▪ Do the results inform educators about curriculum, instruction, and practice?

   ▪ Does it provide valuable information to educators about their students, helping them identify whether students are making the desired progress, falling short, or excelling?

   ▪ Does it provide valuable information to schools and districts about their educators?

Keeping these two questions at the forefront of the work will ensure that measures start on the right track. In the words of a Massachusetts teacher: "If the measure doesn't give me information that's useful and relevant to my instruction, I'm not interested." All other technical considerations are part of a developmental process to refine and strengthen the measures over time.

Given the evaluation context, it can be tempting to prioritize the question, "What measures are best for determining an educator's impact on student learning?" It is important to remember, however, that including student learning in evaluation is also about the learning for educators. For the process to be effective, the measures used must provide valuable feedback to educators.

The highest priorities for DDMs, therefore, are that they are **aligned to content and informative**. The results should help educators recognize where students are succeeding as well as where they are struggling and to identify where to adjust practice. Furthermore, the results should also help schools and districts recognize where educators–including teachers, support personnel, and administrators–are succeeding and struggling and identify where to adjust support.

Districts engaged in identifying and piloting measures for use in educator evaluation are considering–and rightly so–a variety of factors, including validity and reliability. (See Appendix G for a brief description of these terms.) While these are important qualities, *a narrow focus on empirically evaluating validity and reliability should not come at the expense of the **usefulness** of the assessment*.[5] Massachusetts educators have a wealth of experience in assessing student learning, including educator-developed quizzes, tests, short-cycle assessments, performance tasks, assignments, and end-of-course exams. This experience will be invaluable to successful implementation of DDMs.

---

[4] This is referred to as "content validity" in assessment literature.

[5] Since the publication of Technical Guide A, ESE has received feedback from the field expressing the perception that ESE overemphasized the importance of technical quality in the DDM decision-making process. We are appreciative of that feedback and want to make it clear that Technical Guide A was designed to increase assessment literacy. It was not the intent of ESE to suggest that these considerations are more important than measures that are meaningful and useful, i.e., aligned to content and informative.

Five Considerations When Piloting DDMs

Once educators are confident that the measures are well-aligned to local curricula and find the results informative, districts are ready to consider aspects beyond these two priorities. The next five considerations reflect next steps in evaluating the quality of DDMs in preparation for piloting (testing the measures). ESE acknowledges that this is not an easy process and that it will not be done perfectly the first time. By piloting DDMs, district leaders, administrators, and teachers gain a greater understanding together of where future attention needs to be directed. Over time, working through the five considerations outlined below will allow districts to make important refinements that result in stronger assessments. Throughout this developmental process, districts should always return to the two focus questions: Is the measure aligned to content? Is the measure informative to educators at the classroom-, school-, and district-level?

1. **How effectively does the assessment measure growth?**

    The regulations define DDMs as "measures of student learning, growth, or achievement." This represents the next important consideration for selecting and piloting DDMs. An assessment that measures student growth is able to determine the "change in an individual student's performance over time."[6] A variety of methods and types of assessments can be used to measure growth. In Section 3, this guide introduces key concepts about measuring growth and demonstrates how to incorporate the consideration of this question into a pilot.

What does "growth" mean in the context of DDMs?

District-determined measures are ultimately used to determine whether an educator's impact on students was low, moderate, or high. The determination of a Student Impact Rating asks the fundamental question: How much progress did this educator's students make relative to the district's expectations for learning for one school year? Assessments that measure learning gains relative to prior achievement are best suited for answering that question. The MCAS *growth model*, the method ESE uses to measure student progress on statewide assessments, which produces *student growth percentiles* (SGP),[7] is a sophisticated statistical model for measuring growth. **The focus on growth in the context of DDMs does not require sophisticated statistical approaches.** For example, the use of end-of-course exams to judge learning in relation to course objectives or the repeated administration of a writing assignment to evaluate improvement in writing skills are appropriate DDMs for measuring growth. Section 3 of this guide describes approaches to measuring growth and specific examples of appropriate DDMs.

---

[6] Definition from the http://www.doe.mass.edu/mcas/growth/GrowthPresentation.pdf

[7] Find more information on the Massachusetts Growth Model or Student Growth Percentiles at http://www.doe.mass.edu/mcas/growth/.

2. **Is there a common administration protocol?**

Clear instructions should accompany each measure so that educators administer it fairly and consistently across all students, classrooms, and schools that use the same DDM. This means that all educators use a common procedure for administering the assessment, which includes giving the same instructions to all students. For example, are students expected to complete a test during a 55-minute class period or can the test be completed over three days? There may be a script that educators read to students. Are there agreed-upon responses to common student questions about the directions? How much clarification is allowed? An administration protocol should also include agreed-upon options for accommodations for students with disabilities. (See Appendix B for details on administration.)

3. **Is there a common scoring process?**

Educators should score assessments consistently. This includes both the scoring method and the scoring directions. For example, educators may score an assessment using an answer key or a scoring rubric. Who is responsible for scoring the student responses–their teachers, another educator, or a combination? Scorers need to be trained to ensure consistency across results.[8] When the measure is used at multiple schools, this training should be done at the district level to calibrate across scorers. A key strategy for promoting calibration of raters is the development of "anchor papers," i.e., real examples of student work that represent different performance levels on a rubric. Scorers compare student responses to anchor papers or anchor works to promote consistent and reliable scoring and interpretation of the rubric. Anchor papers can also be used to show students what a final product looks like at a given proficiency level. (See Appendix B for details on scoring.)

4. **How do results correspond to low, moderate, or high growth and impact?**

The key connection between student outcomes on a DDM and the Student Impact Rating is the idea of "one year's growth." A rating of moderate on a given measure would indicate that on average, the educator's students made one year's progress in one year's time. (603 CMR 35.09(3)) When measuring growth, a typical growth score on a DDM will often be determined by the difference between where the students started (baseline) and ended (final results). What growth results corresponds to one year's worth of learning? What would indicate that on average, the students made less than a year's growth? Made more than a year's growth?

Note that DDMs do not necessarily need to measure growth over the whole school year or course (in the case of courses that span less than a full school year). Assessments of a portion of the school year or course could potentially be used as DDMs. When an educator uses a DDM that measures less than a full school year or course, the educator's second DDM should measure growth over the whole school year or course. Districts might also consider using more than one additional measure to determine an educator's Student Impact Rating, e.g., an educator might use three or more DDMs if they each measure content and/or skills that cover less than a full year or course of instruction.

---

[8] Districts are encouraged to identify, modify, or develop DDMs that cover higher-level skills and knowledge which may be best assessed using short and extended constructed response items and tasks and related scoring rubrics. In these cases, districts should use multiple raters and have clear protocols for how they address situations when different raters arrive at different ratings.

Evaluators will use the student results of DDMs to determine educator Student Impact Ratings, measuring whether an educator's impact on student learning is low, moderate, or high. This will *typically* happen in three distinct phases:

Phase 1

Determine how individual student scores correspond to low, moderate, or high outcomes for particular DDMS.

Example: Students who gained more than 30 points from the pre to post test may receive a rating of high.

Timeline: Begins 2013-14 where DDMs are piloted.

Phase 2

Determine how aggregate student scores correspond to low, moderate, or high educator impact for a particular measure.

Example: The impact score may be the median student's individual rating.

Timeline: Begins 2014-15 where DDMs are implemented.

Phase 3

Combine information from Phase 2 across at least two years and two measures (a minimum of four data points) to determine whether an educator's Student Impact Rating is low, moderate, or high.

Example: A teacher with three ratings of moderate and one rating of high may receive a rating of moderate.

Timeline: Begins 2015-16 where trends and patterns are established, i.e. at least two DDMs for at least two years.

Phase 1 will not always be necessary. For example, if a district used a graduation rate as a DDM, the students either did or did not graduate. Therefore, it would not be possible for individual students to show low, moderate, or high growth. Rather, the district might determine what degree of change in graduation rate from one year to the next would represent low, moderate, or high growth across all students relative to a 4-year or 5-year rate.

While piloting during the 2013-14 school year, the primary focus is on the first phase. ESE does not expect districts to collect Year 1 data for Student Impact Ratings during the piloting year and 2013-14 pilot data will not contribute to the determination of any educator's Student Impact Rating.[9] However, districts should begin to formulate an approach to Phase 2, e.g., determine *how* student outcomes will translate to low, moderate, or high educator impact for *each measure.*

---

[9] Per the regulations, no educator will be issued an impact rating until DDMs are identified and trends (data across at least two years) and patterns (data across at least two measures) have been established. See 603 CMR 35.11 (4)(c) and 603 CMR 35.02.

5. **Are DDMS "comparable across schools, grades, and subject matter district-wide?"[10]**

This definition of "comparable" combines two different types of comparability: 1) across schools and 2) across grades and subject matter. The key distinction between the two is whether the educators using the DDM(s) are in the same role (i.e., responsible for teaching the same grade/subject or course) or different roles (i.e., responsible for teaching different grades/subjects or courses).[11]

- *"Comparable across schools" pertains to educators in the same role.* ESE recommends that educators responsible for teaching the same grade/subject or course in a district use identical DDM(s) to ensure comparability across schools. For example, all full first-year high school chemistry teachers or all 5th grade teachers in a district should typically administer the same DDM(s). The use of identical DDMs creates opportunities for educators to discuss results and share ideas for adjusting instruction, and allows for better calibration in scoring across educators.

  However, there may be good reasons why a district would choose measures that are not identical across educators in the same role within or across schools. For example, some larger districts with elementary schools that serve different student populations have significantly different school-level goals. Such districts may choose to select DDMs that differ from one school to another in order to better align to school goals. For example, 5th grade teachers in School A may use an ELA and a math DDM, while 5th grade teachers in School B use the same math DDM but replace the ELA measure with a science measure because School B has a stronger STEM focus.

- *"Comparable across grades and subject matter" pertains to whether DDMs permit educators in different roles to demonstrate a similar range of impact on student learning.* DDMs provide the opportunity to compare student outcomes across different grades, subjects, and educator roles, but these comparisons must be fair and meaningful and result in accurate information. DDMs should not have significantly different levels of rigor such that it is more challenging for high school students to demonstrate growth in chemistry than it is for 5th grade students to demonstrate growth in ELA. The rigor of a measure relies on a variety of factors including its alignment to content, the range of difficulty of tasks or items, and the scoring criteria. Given the variation in assessment practices across different grades and subjects, it is natural that the rigor of DDMs will vary at first. The pilot year is a critical time to refine and adjust measures to work toward comparability across grades/subjects and courses.

An extended explanation of comparability is available in Appendix G.

---

[10] The regulations refer to DDMs as "comparable across schools, grades, and subject matter district-wide" in 603 CMR 35.09(2)a.

[11] Educators responsible for more than one subject/course do not need to use multiple DDMs for <u>each</u> subject/course. Example: Mrs. Lawrence and Mr. Colby are both elementary art teachers in the same district. Mrs. Lawrence teaches only visual art; she has two visual art DDMs. Mr. Colby teaches visual and performing art; he uses one visual art DDM and one performing art DDM. Mr. Colby's visual art DDM is identical to one of Mrs. Lawrence's. (Note: Mr. Colby *could* use two visual arts DDM and not use a performing art DDM. However, ESE encourages the use of DDMs that fairly represent an educator's practice and responsibilities.)

Key Questions and Considerations At-A-Glance

Key Questions When Selecting DDMs

These are the two most important questions and should always be at the forefront when identifying and piloting district-determined measures.

1. Is the measure aligned to content?
2. Is the measure informative?

Considerations When Piloting DDMs

These five questions represent key considerations that districts should keep in mind when piloting district-determined measures.

1. How effectively does the assessment measure growth?
2. Is there a common administration protocol?
3. Is there a common scoring process?
4. How do results correspond to low, moderate, or high growth and impact?
5. Are DDMs "comparable across schools, grades, and subject matter district-wide?"

Section 3. Measuring Student Growth with DDMs

Student growth scores provide greater insight into student learning than is possible through the sole use of single-point-in-time student achievement measures. This is because students do not enter a classroom with the same level of knowledge, skills, and readiness. Achievement scores provide valuable feedback to educators about student attainment against standards, but taken by themselves may not be a sufficient reflection of student progress.

In addition, DDMs that measure growth help to "even the playing field" for educators–allowing educators who teach students who start out behind have a similar opportunity to demonstrate their impact on student learning as educators who teacher students who start out ahead. Many teachers worry that they will be "punished" for teaching the neediest students–that is, they will always receive a low Impact Rating if they have a high number of low-achieving students. Measuring growth ensures that educators teaching low-achieving students will be just as likely as educators teaching high-achieving students to receive an Impact Rating of moderate or high. (In fact, measures of growth actually create opportunities to identify where low-achieving students are making tremendous gains). Therefore, assessments that measure growth allow schools and districts to rate educators based on *their impact* on student learning, and not on which students they happen to be teaching.

This section describes several different practical approaches to measuring growth. It is designed to help districts consider how to select, modify, or create DDMs to ensure they measure growth in ways that are accessible to educators.

Approaches to Measuring Student Growth

Student growth measures answer the fundamental question of, "Where did my students start and where did they end?" Although measuring growth is sometimes perceived as synonymous with pre- and post-testing, there are many approaches to determining student growth. The goal of this section is to familiarize the reader with the following range of options:

- Pre-Test/Post-Test
- Repeated Measures Design
- Holistic Evaluation
- Post-Test Only

The approaches outlined here do not represent a comprehensive list. Rather than introducing complex statistical and psychometric methods found in many statewide and commercial assessments, this guidance focuses on four approaches that lend themselves to locally developed measures.[12]

---

[12] Again, ESE is happy to engage with interested districts on the more technical statistical and psychometric approaches. We welcome you to contact us at EducatorEvaluation@doe.mass.edu.

Pre-Test/Post-Test

Description of the approach

One straightforward way to measure growth is to measure student knowledge before and after instruction, and then use the scores to determine a student's growth. Pre- and post-tests can be identical measures administered twice or comparable versions. For example, a music teacher could evaluate a student's knowledge of scales using a performance task at the beginning of the year and then again at the end of the year. (If the teacher asked students to perform the same four scales, this would be an example of identical assessments; if she asked the students to perform different scales, this would be a comparable version of the same assessment.) Many commercial interim assessment programs like MAP, Galileo, and Acuity provide psychometrically strong assessments that districts can administer using a pre-test/post-test approach.

How to measure growth

Once students have completed both the pre- and post- tests, a straightforward way to estimate growth from these assessments would be to compute the difference between the scores. For example, if a student scored an 80 on the final exam and a 40 on the pre-test, one could compute that the student grew 40 points. However, a difference score is not the only way one could use a pre-test score to compute a growth score, and is not always the most appropriate.[13] The important thing is that the method of computing a growth score best reflects comparability between students beginning at different levels of achievement.

Considerations

As one teacher wisely points out, "students should not take a pre-test that is not valuable to their learning. We don't want this to be an exercise in measuring growth for DDMs but a useful learning experience in its own right." This can be a delicate balance to strike, but is a critical point. For example, a writing assessment that uses an identical prompt may result in more accurate growth scores, but students may not benefit from repeating the exact same writing assignment. Thus, the prompt for the two writing assessments may be different. However, the prompts should be comparably rigorous and the results scored using an identical rubric.

When using pre-tests, improvement between the assessments should be highly indicative of classroom learning for that grade level and subject area. For example, assessments of basic/foundational skills such as reading comprehension may be an important prerequisite to learning the course material, such as U.S. history, but unless the curriculum explicitly addresses those basic skills, it is not appropriate to measure growth in those skills. Districts should pay close attention to assessments that include measures of higher-level thinking, problem solving, and creativity. While it is important to provide opportunities for students to demonstrate and develop these skills, growth in such areas should only be measured if it was an explicit purpose of the curriculum, i.e., educators expected to see students be better able to solve problem or think more creatively after instruction.

---

[13] For example, one could compute the percentage increase instead of a difference. Consider the example of completing push-ups. If one used percentage increase, a student who went from a score of 10 to 20 would receive the same growth score as a student who scored moved from 40 to 80. A percentage increase may be a more appropriate approach than using a difference score.

One advantage to this method is that pre-test results can inform curriculum planning and instruction. For example, if a science teacher found from a pre-test that students had a good understanding of simple machines but several misconceptions around forces, they could choose to spend extra time on addressing these misconceptions.

Repeated Measures:

Description of the approach

Many educators already use short measures repeatedly throughout the year. A DDM can look at the pattern across the repeated administrations of a short and authentic assessment throughout the year or course. For example, early reading teachers may already complete weekly running records to track the number of errors that a student makes when reading a text; or an assistant principal may monitor daily attendance records while working to reduce student absences (an *indirect* repeated measure). These repeated measures serve a similar function to a pre- and post-test by illustrating change over time in student learning or performance.

How to measure growth

One of the most straightforward ways to analyze the results from repeated measures is to look at the results graphically. This presentation can provide educators with information on when and how students made progress that is not available through other methods. A growth score can be generated using methods ranging from simple to sophisticated. For example, the assistant principal could look at the difference between the average attendance for October and April to see if a targeted intervention made a significant impact on the level of absences during that time. Districts can contact the ESE for support with more sophisticated approaches that are feasible for district-level use by emailing EducatorEvaluation@doe.mass.edu.[14]

Considerations

Authentic performance tasks that (a) closely match the learning an educator wants students to demonstrate, and (b) are designed to show improvement over time as students gain more sophisticated knowledge and/or improve skills are the best candidates for use as a repeated measures approach DDM. Avoid repeated measures on which students may demonstrate improvement over time simply due to familiarity with the assessment; for example, students could make large gains over the year on a weekly quiz of state capitals without significant growth in knowledge of geography.

Repeated measures have the advantage of being shorter and deeply integrated into the curriculum. This can make them a useful option for young students. Another advantage of this method is that teachers receive continuous feedback about students, and can modify their practice accordingly.

---

[14] A more sophisticated analysis of attendance during this timeframe would be to compute a regression slope using Excel that could further illustrate the impact of various factors on attendance, yielding valuable information to the assistant principal about a specific intervention.

Holistic Evaluation:

Description of the approach

A holistic evaluation of student growth combines aspects of a pre- and post-test model with the regularity of a repeated measures approach. The critical difference is a nuance of the scoring method: like some pre/post or repeated measures, a holistic evaluation uses a rubric to score student work but the rubric differs in criteria and use. Rubrics used to score pre- and post-test or repeated measures are typically designed to be used with one piece of student work at a time, or a collection of student work that's being examined as a whole product (such as a portfolio). A "growth rubric" for a holistic evaluation is designed to *compare* two or more examples of student work. The rubric criteria would include language that describes growth, as shown in Figure 2.

How to measure growth

The key component of this method is a rubric that looks at growth across multiple samples that are collected systematically, spanning the course or year. Unlike the repeated measures or pre-test/post-test, there is no mathematical calculation that results in a growth score, such as taking the difference between two scores or calculating the average growth between administration at select points in time. Rather, the rater determines what level of growth the work demonstrates by applying the growth criteria to multiple work samples. This rubric should include detailed descriptions of what growth looks like across the examples and not the quality at any individual point. The Tennessee Fine Arts Portfolio Model includes a well-known example of this type of rubric.

Considerations

Using a holistic evaluation approach can avoid important validity concerns that are raised by scoring performance assessments individually and then computing a growth score from the individual scores. For example, if an educator administered a writing pre-test, a student may receive a low score on following writing conventions because she/he used quotation marks incorrectly. If the student did not use dialogue on the post-test, she/he might receive a higher score on following writing conventions because there were fewer errors. When the scorer computed a growth score from the difference between the two writing samples, it would appear that the student improved in this area. However, when looking across the writing samples, there would be no clear evidence of growth. In contrast, a holistic evaluation rubric could include a growth criterion that requires the scorer to identify examples of writing conventions that were used incorrectly in early examples and then used correctly in later writing samples, indicating growth over time.

The following example demonstrates an assignment that might lend itself to holistic evaluation. Figure 1 reproduces a series of drawings of a butterfly done by Austin, a first grader from Anser Charter School in Boise, Idaho. The assignment integrated standards from both science and art curricula.[15] Figure 2 is an example of a single criterion from a growth rubric (developed by ESE). The rubric language requires the scorer to identify examples of details that are added or improved across the versions but does not specify which details, e.g., butterfly body proportions, wing patterns, etc.

---

[15] Used with permission from Expeditionary Learning's Center for Student Work. For more information, including the associated content standards, please see http://elschools.org/student-work/butterfly-drafts.

Example: Student Work Sample and Example Growth Criteria for a Holistic Evaluation Rubric

The purpose of the assignment shown in Figure 1 was to improve the students' ability to "think like scientists—to observe carefully and record their observations…"[16] in alignment with key content standards in both art and science. Although the teacher would provide feedback to Austin after each iteration, the holistic growth rubric (shown in Figure 2) would be used at the end to determine what level of growth Austin demonstrated across the five drawings rather than an individual drawing.[17] It is clear from these work samples that Austin demonstrated strong growth in the level of detail and accuracy from picture to picture. For example, the "High Growth" descriptor includes, "There are multiple examples of details that build and elaborate on previous versions." Each of Austin's drawings added detail to the wings' overall shape, markings, and proportion.
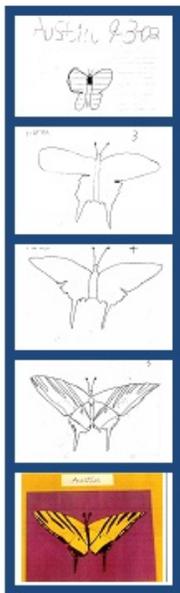
Figure 1. Student Work

Figure 2. Example Growth Criteria from Holistic Evaluation Rubric

Growth Criteria: Improving the Level of Detail			
No Growth	Low Growth	Moderate Growth	High Growth
One or two are true	*One or two are true*	*All are true*	*All are true*
* No new details across versions	* There are a few details included across all versions	* There are many examples of added details across all versions	* On average there are multiple details added across every version
* New details are added, but not included in future versions	* There are many added details included, but they are not included consistently, or none are improved or elaborated upon	* At least one example of a detail that is improved or elaborated in future versions	* There are multiple examples of details that build and elaborate on previous versions
* A few new details are added that are not relevant, accurate, or meaningful	* There are many added details, but several are not relevant, accurate, or meaningful	*Details are consistently included in future versions	* The added details reflect the most relevant and meaningful additions
		*The added details reflect relevant and meaningful additions	

---

[16] Quotation used with permission from Expeditionary Learning Center for Student Work. http://elschools.org/student-work/butterfly-drafts

[17] Note that these content standards may not be sufficiently representative of the curriculum to be a strong DDM; the example is intended to illustrate the kind of work that would lend itself to a growth rubric.

This approach may be particularly useful when using multifaceted performance assessments that require considerable interpretation when being evaluated. In this context, a multifaceted performance assessment may involve only one task but has many different elements that could be measured. For example, one photograph may be an example of high quality because it demonstrates an excellent use of depth of field while another uses framing.[18] However, many equally high quality images do not emphasize depth of field or framing at all. A holistic growth rubric allows the scorer to consistently identify growth in student work that may have very different strengths and weaknesses.

The approach would also be valuable when the students in a particular grade, subject, or course have quite varied levels of ability but are working on a common task. For example, if a high school student with moderate disabilities is working on an ELA writing assignment that requires use of evidence from text, a rubric that has sufficient criteria to fully capture the range of student ability may become unwieldy. In this case, a holistic evaluation rubric may be better able to show growth along a wide spectrum.

Holistic rubrics can be challenging to construct. Districts need to be proactive about ensuring the quality and inter-rater reliability of these rubrics. For example, teams of reviewers can rate selected examples together and discuss differences in scores. The goal from these discussions would be to clarify definitions and arrive at a consistent interpretation. (See Appendix B for suggested resources to support the development of rubrics.)

Post-Test Only

A DDM that only measures student performance at or near the end of the grade or course is a post-test only. While some state and commercial assessments use sophisticated statistical methods for calculating growth with a post-test only (such as the MCAS), this approach is unlikely to be feasible or appropriate for a locally-developed academic measure.

More broadly speaking, a post-test only could also refer to a measure that is only available annually that is appropriate for administrators or support personnel, such as graduation and promotion rates.

Description of the approach

Some measures of achievement or performance can only be administered once. For example, there is no pre-test for graduation rate. In these cases, it may be defensible to only use a post-test. In such instances, it is critical to include additional information that helps predict the expected level of achievement, such as the difference from the previous year's graduation rate.

How to measure growth

How growth is computed will depend on the type of information included with the post-test. For example, using the graduation rate example, a simple difference from the previous year's graduation rate would be one approach. However, if there were considerable variation in graduation rate across multiple years, or across multiple schools, it may be more appropriate to compute a difference from the average graduation rate over time or across a large sample.

---

[18] Depth of field is the area of an image that is in focus. This can be used to highlight the key part of the image. Framing is the use of parts of image to create borders within a picture.

Considerations

Since post-test only measures are not based on changes within a discrete sample, and growth must reflect change across different samples, there are multiple threats to the validity of these measures. For example, if a town's demographics were changing dramatically, the previous year's graduation rate may not be a fair baseline for comparison.

Another key type of assessment that typically falls into the Post-Test only category is "credential assessments." Students in Vocational Technology programs complete assessments associated with the attainment of industry-recognized credentials. These assessments are often a critical component of successfully completing a course or program. ESE is collaborating with WestEd to provide future guidance on the use of these assessments as DDMs.

Applying the Different Approaches to Measuring Growth

The purpose of describing these approaches is not to ensure that districts use the correct labels when deciding on an approach to measure growth. Some very valuable approaches may not be easy to classify, such as a measure that is given at the beginning, middle, and end of year. Whether this is called a pre-test/post-test or a repeated measure is not important; what matters is whether the measure provides useful information to educators and reinforces a meaningful learning experience for students.

Different approaches to measuring growth do not always result in different learning experiences for students. In some cases, the difference may be how the results are scored. One type of measure, such as a writing assignment, can determine student growth in a variety of ways.

Example: Grade 9 history essay

Pre-Test/Post-Test: students complete the essay assignment at the beginning and end of the year; rater(s) score the two essays using the same rubric, then calculate student growth by taking the difference between the rubric score on the baseline and final essays.

Repeated Measure: students build a portfolio composed of five essays over the course of the year. The rater(s) score each of the five essays using the same rubric and look at the growth over the year. This may be easily incorporated into instruction during the year and a key learning experience for both the students and the teacher, as it provides immediate feedback to the student, as well as targeted information to the teacher about the effectiveness of his/her instruction. Furthermore, using multiple samples of student work often results in a more comprehensive picture of student growth. The tradeoff is increased workload, particularly if multiple raters score each assignment. Educators can start by graphing the five scores for each student to see progress.

Holistic Evaluation: students build a portfolio comprised of five essays over the course of the year. During the year, the teacher grades each essay and provides feedback, but does not measure growth until all five essays are completed. At the end of the year, the rater uses a rubric with language that examines growth across the five essays. Rather than comparing five rubric scores, this results in one growth score. This approach is the exact same learning experience for students as the repeated measures approach, but a different process for measuring growth. It is important that rubrics used for holistic evaluation carefully define evidence of growth.

Post-Test Only: in this example, a post-test only would not be an appropriate approach to measuring growth, as there would be no baseline for student work.

Note: all of these approaches require a significant investment in collaboration, training, and practice to ensure that scoring is consistent across raters (inter-rater reliability).

Measuring Growth with Specific Assessment Types

Districts have to weigh many options when determining which measures to use as DDMs. A highly relevant measure of student achievement that is closely tied to the curriculum but does not fully capture growth may be a better measure than a poorly constructed one that does. Certain types of assessments are often used as measures of achievement but also have the capacity to measure student growth when carefully designed and scored. The regulations specifically identify three particular types of measures that fall into this category when defining DDMs: portfolios, unit assessments, and capstone projects. The following section describes considerations to support district decision-making with regard to the use of such assessments.

Portfolios

Not all portfolios serve the same purpose. Portfolios can be an excellent achievement measure or an excellent growth measure. For example, an art student applying to art programs would likely create a portfolio of work composed of his/her best pieces across a variety of media to demonstrate ability. On the other hand, an art teacher applying for a new job would be more likely to create a portfolio of student work composed of pieces that show how the students grew over the course of the year to demonstrate how his/her instruction helped student growth. If a portfolio is to be used as a DDM that measures growth, it must be designed to capture progress rather than to showcase accomplishments.

Unit Assessments

Unit assessments are one of the most common measures currently used in districts. Some educators are already using unit assessments to measure growth . For example, one high school math teacher in a Massachusetts district began administering her end of unit tests as pre-tests and found the results to be invaluable in regrouping students for the new unit. This teacher, aware that the material had not been taught to her students, began her administration of the pre-test this way: "Today we're going to be beginning a new unit. I'd like to have a good sense of what you all currently know so I know the best way to teach you. <hands out the assessment> Don't worry – there may be some things on here you don't understand yet – but this won't be graded so just try your best."

It is important to note that a pre-test and post-test on one unit is unlikely to represent a significant portion of the curriculum for which the educator is responsible and may not be, on its own, a strong DDM. However, it is possible to measure growth by examining a series of units. ESE is developing a guide that provides a detailed example of how a district could generate a DDM comprised of multiple curriculum-embedded performance assessments from ESE's Model Curriculum Units. [19]

---

[19] ESE designed the Model Curriculum Units to provide districts with example curriculum aligned with the Curriculum Frameworks. A key aspect of the Model Curriculum Units is the inclusion of performance assessments that require students to demonstrate understanding in realistic settings. Sample Model Curriculum Units are currently available at http://www.doe.mass.edu/candi/model/sample.html.

End-of-Course Exams

End-of-course exams, like unit assessments, are frequently used by schools and districts. They have the advantage of being well-aligned to content and informative with regard to student achievement. Districts can use end-of-course exams to measure growth with a variety of approaches. For example, knowing students' prior achievement in social studies might provide an estimate of growth in the current U.S. History class. However, it may be difficult to draw accurate conclusions about the U.S. History teacher's impact on those students without a careful analysis of the constructs being measured across the two exams. Districts should be cautious when using this type of data in an educator's evaluation.

Educators could also construct a measure that captures baseline information within the same year or course and produces scores that could be compared to the end-of-course exam. A simple approach would be to use an alternative version of the end-of-course exam as a pre-test. A more complex approach would be to identify key knowledge or skills that are taught during the year or course and develop a test/task or scoring method that allows educators to measure growth. For example, some science exams may require students to conduct an experiment; in that case, educators could examine student growth in the skills necessary to successfully conduct an experiment, e.g. setting up the conditions, testing a hypothesis, or documenting and analyzing outcomes. The baseline test/task should be as similar as possible to the end-of-course exam to have more confidence in interpreting growth results.

Capstone Projects

Capstone Projects are large-scale student projects that represent a culmination of the work completed in a course. These projects often include extensive planning, problem solving, creative thinking, research, and/or formal presentations. Capstone projects could include original research papers, extended science experiments, or a theatrical show. Considering the time and effort put into these projects, they are important to consider for use as a DDM.

One of the advantages of using these projects as a DDM is that they offer the chance to be authentic measures closely tied to the curriculum. One of the important considerations when choosing a DDM is that it assesses content that is instructionally sensitive. That is, students must demonstrate knowledge or skills explicitly taught as part of the curriculum. When projects reflect research, presentation, or artistic skills, it is important to ask if these skills were taught as a part of the class. Another consideration is that these projects need to reflect student work, and not the work of other students, parents, teachers, or support staff. It can be difficult to ensure that students receive comparable help and support when projects include diverse, complex, and/or multifaceted tasks over long periods of time.

Perhaps the biggest challenge in using capstone projects as DDMs is the difficulty with measuring growth. The goal of DDMs is that they measure student growth over the year. Often there is no clear pre-test for a capstone project, although they may be excellent and comprehensive measures of achievement that are an important learning experience for students and well-aligned to curriculum. This does not mean that capstone projects cannot be used as DDMs, but that districts need to think carefully about how they are measuring a student's ability relative to where she/he started. For example, if the educator's purpose is to teach students the process of conducting research, an option might be to have the students write a brief description at the beginning of the year of what steps they would take to develop a research project, then compare that to their level of knowledge of the process at the end of the year, as captured by their success in developing the capstone project.

Section 4. Piloting DDMs

Piloting provides an opportunity to receive important feedback about how DDMs work in practice and how they may need to be revised. "Piloting" simply means testing, analyzing results, making adjustments using what was learned, and testing again. When piloting, districts should remember:

- Piloting measures should always include collaboration with the educators responsible for administering and scoring the assessment; these educators have vital insight and the process will foster educator ownership.

- Pilots are an iterative process. Improving DDMs requires administering the test, analyzing the results, and reflecting on the process to make smart modifications.

- The pilot will be most informative if districts have a clear idea of what they want to learn from the pilot based on the starting point for each piloted measure. The process, which may result in unexpected lessons, will–and should–raise more questions to be explored in the next assessment cycle.

ESE suggests the following steps for piloting DDMs:

Preparing to Pilot	
Step	**Key Questions**
1. Create a Team	- Who will be on the team? - What are the key responsibilities?
2. Determine Content to Be Measured	- What standards and objectives are the most important to measure?
3. Identify Potential Measure to Pilot	- What measures are already in use in the district? - Do components need to be developed or modified before piloting?
4. Prepare to Pilot Measure	- What is the piloting timeline? - How will the data be collected and stored?
Piloting	
Step	**Key Actions**
5. Test	- Administer the measure - Score the results - Collect the data
6. Analyze	- Analyze student results - Analyze the administration and scoring processes
7. Adjust	- Modify components as needed
8. Repeat	- Test, analyze, and adjust

These steps are described in more depth in Appendix B, including key questions, considerations, and examples.

Section 5. ESE Support

Existing ESE Resources

All ESE resources can be found at http://www.doe.mass.edu/edeval/ddm/.

- Part VII: Rating Educator Impact on Student Learning Using District-Determined Measures of Student Learning.

  Part VII provides an overview of district-determined measures and related concepts. It will be most valuable for districts beginning to learn about this work. *(August 2012)*

- Technical Guide A: Considerations Regarding District-Determined Measures.

  Technical Guide A is a supplement to Part VII and is designed to increase assessment literacy by introducing foundational assessment concepts. It will be most valuable to districts interested in learning more about technical assessment concepts. *(April 2013)*

- Assessment Quality Checklist and Tracking Tool.

  The Assessment Quality Checklist and Tracking Tool is an interactive tool, built in Microsoft Excel, that organizes and catalogs information about individual assessments into a district-wide tracker of all potential DDMs. It will be most valuable to districts working to identify and select measures across the district. *(April 2013)*

- District-Determined Measures and Assessment Literacy Webinar Series. ESE has developed an eight-part webinar series on District-Determined Measures & Assessment Literacy. This series is targeted for district teams engaged in the work of identifying and selecting DDMs (e.g., district and school-based curriculum and assessment leaders). Resources from these webinar include the **recorded webinar** and **materials** from each session. *(March 2013-January 2014)*

- Technical Assistance Sessions and Related Materials. ESE is hosting a set of three technical assistance and networking sessions. These sessions are intended to build on the Assessment Literacy Webinar Series and provide participants an opportunity to engage with colleagues from other districts around critical planning and implementation questions related to the piloting and eventual implementation DDMs. ESE and national assessment and evaluation experts are also available during these sessions to provide support and clarifying information to help districts pilot DDMs in 2013-14. All materials from the technical assistance sessions are posted to the website.

- ESE Educator Evaluation Team. The Ed Eval team is committed to learning from and with educators about their successes, challenges, and needs. ESE welcomes you to contact us at EducatorEvaluation@doe.mass.edu.

- Educator Evaluation Newsletter. The monthly newsletter is designed to be a timely resource that provides key information, updates, and answers to frequently asked questions. Find issues at: http://www.doe.mass.edu/edeval/communications/newsletter/. ESE encourages educators to sign up to receive the newsletter by email. Sign up here: http://edeval-newsletter-signup.org/.

Forthcoming ESE Resources

ESE is committed to providing support to districts working to integrate implementation of the new educator evaluation system and implementation of the shifts in curriculum and instruction embedded in the new ELA and math Curriculum Frameworks. ESE plans to provide the following resources, technical assistance, and professional development:

- *APPLY BY SEPTEMBER 30TH:* A new grant opportunity, the Professional Practice Innovation Grant, to support district initiatives that build the capacity of teachers and administrators to implement the rigorous standards of the Curriculum Frameworks, in ways fully aligned with the educator evaluation system. Interested districts can learn more here: http://www.doe.mass.edu/grants/grants14/rfp/213.html.

- Additional Model Curriculum Units, which include curriculum-embedded performance assessments. Available in August, September, and November. Information about the Model Curriculum Project and currently available Model Curriculum Units are available at http://www.doe.mass.edu/candi/model/.

- Guidance on the use of curriculum-embedded performance assessments as part of a DDM-strategy (to be published this fall).

- Quick Reference Guide: Integrating the Curriculum Frameworks and Educator Evaluation Implementation (to be published this fall).

- *MARK YOUR CALENDAR:* A Curriculum Summit **November 6-8** that will include breakout sessions related to the five required DDM pilot areas. Look for the registration link in late September. Follow-up professional development will be delivered regionally in early winter.

- Professional development for evaluators on how to focus on shifts embedded in the new ELA and math Curriculum Frameworks during classroom observations.

Additionally, ESE will provide forthcoming support to ensure districts are positioned to well-implement DDMs and determine Student Impact Ratings:

- *REGISTER NOW:* Technical Assistance and Networking sessions on **September 19th** in three regions focused on DDM pilot plans in the five focus areas. To register, please visit: http://www.surveygizmo.com/s3/1333164/Registration-DDM-Technical-Assistance-amp-Networking-Session-2-September-19th-2013

- Example DDMs in the five required pilot areas to provide options for districts to pilot (will be made available in September 2013).

- Model collective bargaining language related to the Student Impact Rating and DDMs (under development).

- Guidance on constructing local growth scores and growth models.

- Guidance on determining Student Impact Ratings.

- Professional development for evaluators on how to administer and score DDMs and use them to determine high, moderate, or low growth, focused on the five required DDM pilot areas.

- Educator Evaluation Spring Convening 2014.

Section 6. Conclusion.

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. DDMs that are aligned to content and are informative support the primary purpose of the new system: to provide timely, useful feedback to teachers to improve their practice and better support student learning.

When engaging in this work, do not let "perfect" be the enemy of "good." Since district-wide implementation of many DDMs is new, it will take time to resolve every detail, such as nuanced questions about details of administration. The key focus is the use of measures that are meaningful and useful to educators at the classroom-, school-, and district-level.

In the words of Commissioner Chester: "As educators, we are committed to having a positive impact on the learning, growth, and achievement of all students. This is our ultimate and most fundamental goal. Educators across the content areas have strived to answer an essential question: Does my actual impact meet or exceed the impact I aspire to have on my students? While no single metric can definitively answer this question, there are ways to build a more complete picture of how an educator's practice intersects with her/his impact on students. The Massachusetts evaluation system, with its two independent but linked ratings, is designed to home in on this intersection and to provide all educators with opportunities for growth and development tailored to their individual needs."[20]

---

[20] From his August 2013 memorandum on district-determined measures. See Appendix A for full text.

Acknowledgements

ESE would like to thank the following contributors for their time and thoughtfulness in reviewing and providing feedback and input on this document:

- Patrick Daly, Director of Academic Services, North Reading School District
- Laura Kohn, Consultant, Education First
- Jamie LaBillois, Executive Director, Norwell Public Schools
- Pauline Parker, Director of Curriculum and PD, Lower Pioneer Valley Edu. Collaborative
- Kristan Rodriguez, Assistant Superintendent and additional school-level educators, Chelmsford Public Schools (CPS)
- Megan Whitmer, former special education teacher, Fayette County Public Schools, KY
- Ross Wilson, Asst. Supt. for Teacher & Leadership Effectiveness, Boston Public Schools

ESE is also appreciative of the following partners who contributed to our thinking around DDM policy and implementation during the development of this guidance:

- Russell Johnston, Superintendent, West Springfield
- Glenn Koocher, Executive Director, Massachusetts Association of School Committees (MASC)
- Michael Long, Attorney, Massachusetts Association of School Superintendents (MASS)
- Dan Murphy, Director of Educational Policy and Programs, American Federation for Teachers, MA (AFT-MA)
- Tom Scott, Executive Director, Massachusetts Association of School Superintendents (MASS)
- Paul Toner, President, Massachusetts Teachers Association (MTA)

ESE's Ed Eval Team is grateful for the collaboration among colleagues who reviewed drafts and provided critical feedback, including:

- Bob Bickerton
- Commissioner Chester
- Kathy Flanagan
- Deputy Commissioner Ingram
- Deb Steenland
- Lucy Wall

Finally, we acknowledge ESE's Ed Eval Team for their hard work on this document:

Samantha Warburton (lead author), Claire Abbott, Susan Berglund, Kat Johnston, Ron Noble, Craig Waterman

Appendices

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

MEMORANDUM

To:	**Superintendents and Race to the Top Charter School Leaders**
From:	**Mitchell D. Chester, Ed. D., Commissioner**
Date:	**August 15, 2013**
Subject:	**District-Determined Measures: Supports, Timeline, and Pilot Expectations**

Through this memorandum I am providing you with an update on supports, timelines, and expectations for implementing the Student Impact Rating of the educator evaluation system. This update includes revisions to timelines that we are making in response to feedback we have heard regarding the challenges of implementing District Determined Measures (DDMs).

Introduction

As educators, we are committed to having a positive impact on the learning, growth, and achievement of all students. This is our ultimate and most fundamental goal. Educators across the content areas have strived to answer an essential question: Does my actual impact meet or exceed the impact I aspire to have on my students? While no single metric can definitively answer this question, there are ways to build a more complete picture of how an educator's practice intersects with her/his impact on students. The Massachusetts evaluation system, with its two independent but linked ratings, is designed to home in on this intersection and to provide all educators with opportunities for growth and development tailored to their individual needs:

- The *Summative Performance Rating* is the final step of the 5-step evaluation cycle. This rating assesses an educator's practice against four statewide Standards of Effective Teaching or Administrator Leadership Practice, as well as an educator's progress toward attainment of his/her goals.

- The *Student Impact Rating* is separate but complementary to the Summative Performance Rating; it is informed by trends (at least two years) and patterns (at least two measures) in student learning as measured by statewide (MCAS) growth measures, where available, and district-determined measures (DDMs).

In the 2012-13 school year, RTTT districts and charter schools completed implementation of the Summative Performance Rating for at least 50 percent of educators. In the 2013-14 school year, all RTTT districts and charter schools will implement Summative Performance Ratings for their remaining educators and all non-RTTT districts will implement with at least 50 percent of their educators.

While many districts have made steady progress in selecting and piloting DDMs – some districts are already well underway with this work – realization of the Student Impact Rating will continue over the next several years. The remainder of this memorandum provides an updated timeline and expectations for selecting, piloting, and implementing DDMs.

Implementation Plan Timeline Extension
In direct response to requests for additional time to do this work thoughtfully and in collaboration with your educators, the Department of Elementary and Secondary Education (ESE) revised its timeline for implementing the Student Impact Rating. In April, I announced that for districts that have yet to begin DDM implementation, the 2013-2014 school year will be a year for identifying and piloting DDMs. At a minimum, districts must pilot DDMs in five areas during the 2013-14 school year.

<div align="center">

April 2013 Timeline Revision

</div>

	2013-14	2014-15	2015-16
Original Timeline	Collect data (Year1)	Collect data (Year 2), **Issue 1st Student Impact Ratings**	Collect data, Issue Student Impact Ratings
Revised Timeline (as of April 2013)	Research and pilot potential DDMs	Collect data (Year1)	Collect data (Year 2), **Issue 1st Student Impact Ratings**

On July 19, 2013, I announced at the Massachusetts Association of School Superintendents' Executive Institute my intention, subject to approval by the Board of Elementary and Secondary Education, to provide further modifications to the due dates for DDM implementation plans. These modifications include an opportunity for districts to apply for a one-year extension of the implementation of DDMs for grade(s)/subject(s) or course(s) for which the district will not have identified DDMs by the second half of the 2013-14 school year. Such extensions will be conditioned on ongoing district progress in implementing DDMs and a clear plan for utilizing the extension (more details below). Blanket extensions will not be granted.

An extension will allow extra time to districts that are unable to fully implement DDMs during the 2014-15 school year. Districts should use this process judiciously. In fact, I am aware that many districts will not need to seek an extension, as they are on pace to begin DDM implementation during the 2014-15 school year. In all cases, I expect districts to continue identifying and piloting DDMs with due diligence during the 2013-14 school year.

July 19, 2013 Timeline Revision

2013-14:

- RTTT districts and charter schools determine and report Summative Performance Ratings for all educators.

- Non-RTTT districts determine and report Summative Performance Ratings for at least 50 percent of educators

- All districts and RTTT charter schools pilot DDMs.

  - September 2013 – By the end of September all districts and RTTT charter schools must submit a list of the DDMs to be piloted during the 2013-14 school year.

    - Access the submission template here: http://www.doe.mass.edu/edeval/ddm/.

    - Note: This list must include at least one potential DDM to be piloted in each of the following five areas (as identified in the April memorandum):
      a) Early grade (K-3) literacy
      b) Early grade (K-3) math
      c) Middle grade (5-8) math
      d) High school writing to text
      e) Traditionally non-tested grades and subjects (e.g., fine arts, music, p.e.)

  - December 2013 – In December, ESE intends to release the June 2014 submission template and DDM implementation extension request form.

  - February 2014 – The previously announced deadline for all districts and RTTT charter schools to submit the final plan for implementing DDMs was February 21, 2014. Based on feedback from the field, I have shifted this deadline to June 1, 2014 to allow more time for piloting.

  - June 2014 – By June 1, 2014, all districts and RTTT charter schools must submit their final plans for implementing statewide (MCAS) growth measures and DDMs for the purpose of collecting Year 1 Student Impact Rating data during the 2014-15 school year.

    - Note: ESE recognizes that some districts may pilot DDMs that require an end-of-year or end-of-course assessment to be administered after the June submission. Districts will have the opportunity to update their implementation plans, as needed, based on pilot results.

    - Note: In conjunction with the June submission, districts may request an extension of the time to implement DDMs for specific grades/subjects or courses for which DDMs have not yet been identified. Such requests will be evaluated on a district-by-district basis and must be supported by a rationale that includes a statement of the efforts made to date to identify DDMs for the specified

grade(s)/subject(s) or course(s), as well as a plan for how the district will use an additional year to research and pilot potential DDMs. Any district seeking a one-year extension of the implementation of DDMs for specific grades/subjects or courses must present a plan that shows the district is implementing DDMs on schedule for some grades/subjects or courses, including the five required 2013-14 DDM pilot areas (initial implementation in 2014-15, Student Impact Ratings calculated at the conclusion of the 2015-16 school year). Consideration of an extension is not consideration of a delay, pause, or moratorium on implementation of the new educator evaluation system.

For all districts, whether or not they pursue an extension for some grades/subjects or courses, evaluations conducted during the 2014-15 and 2015-16 school years must be used to continually improve instruction; meaningfully differentiate performance using the four performance levels (exemplary, proficient, needs improvement, and unsatisfactory); measure student learning through the implemented DDMs; include frequent observations that result in clear, timely, and useful feedback to drive support and professional development; and result in Summative Performance Ratings as well as Student Impact Ratings where trends and patterns have been established using statewide growth measures and DDMs.

2014-15:

- All districts and RTTT charter schools determine and report Summative Performance Ratings for all educators.

- All districts and RTTT charter schools implement DDMs and collect Year 1 Student Impact Rating data for all educators (with the exception of educators who teach the particular grades/subjects or courses for which a district has received an extension).

2015-16:

- All districts and RTTT charter schools determine and report Summative Performance Ratings for all educators.

- All districts and RTTT charter schools implement DDMs, collect Year 2 Student Impact Rating, and determine and report Student Impact Ratings for all educators (with the exception of educators who teach the particular grades/subjects or courses for which a district has received an extension).

Support for Implementing Evaluations in Tandem with New Curriculum Frameworks

I understand that the work of building a coherent district strategy for the rollout of new educator evaluation systems in conjunction with the implementation of new ELA and Mathematics Curriculum Frameworks is challenging. To this end, during the 2013-14 school year, ESE will provide support and technical assistance to help districts integrate

implementation of the new educator evaluation system and implementation of the shifts in curriculum and instruction embedded in the new ELA and math Curriculum Frameworks. In particular, ESE will make available a series of resources, technical assistance, and professional development aimed at helping districts build comprehensive plans for integrating these two key initiatives, with a focus on the five required DDM pilot areas. The supports include:

- Sample DDMs in the five required pilot areas by the end of August to provide options for districts to pilot.

- Technical Assistance and Networking sessions on September 19th in three regions focused on DDM pilot plans in the five focus areas.

- Additional Model Curriculum Units, which include curriculum-embedded performance assessments, in August, September and November.

- Guidance on the use of curriculum-embedded performance assessments as part of a DDM-strategy.

- Professional development for evaluators on how to focus on shifts embedded in the new ELA and math Curriculum Frameworks during classroom observations.

- Professional development for evaluators on how to administer and score DDMs and use them to determine high, moderate or low growth, focused on the five required DDM pilot areas.

- A Curriculum Summit in November that will include breakout sessions related to the five required DDM pilot areas.

- A new grant opportunity, the Professional Practice Innovation Grant, to support district initiatives that build the capacity of teachers and administrators to implement the rigorous standards of the Curriculum Frameworks, in ways fully aligned with the educator evaluation system. Interested districts can learn more here: http://www.doe.mass.edu/grants/grants14/rfp/213.html.

Additionally, ESE will provide forthcoming support to ensure districts are positioned to well-implement DDMs and determine Student Impact Ratings:

- Technical Guide B will be released shortly to assist districts in the practical application of assessment concepts to piloting potential DDMs and measuring student growth.

- Model collective bargaining language related to the Student Impact Rating and DDMs is under development.

- An ongoing Assessment Literacy webinar series are available. Register here: http://www.doe.mass.edu/edeval/ddm/webinar.html.

- Guidance on constructing local growth scores and growth models will be released.

- Guidance on determining Student Impact Rating will be released.

Three Suggestions for Piloting DDMs:

As you continue planning your potential DDM pilots, please keep these ideas in mind:

- First, collaboration with and among educators is paramount. Involve your educators in developing and/or evaluating potential DDMs. Harkening back to the Introduction section, educators are invested in the pursuit of information that will help them gauge their impact on students. Effective educators across the Commonwealth ask themselves, "Did my students learn today? How much did they learn? How do I know?" The Student Impact Rating and DDMs scale-up processes that great teachers have long had in place by establishing common measures across a district.

- Second, take advantage of a no-stakes pilot year to try out new measures and introduce educators to this new dimension of the evaluation framework. Districts are strongly encouraged to expand their pilots beyond the five required pilot areas. Fold assessment literacy into the district's professional development plan to stimulate dialogue amongst educators about the comparative benefits of different potential DDMs the district could pilot. Consider how contributing to the development or piloting of potential DDMs can be folded into educators' professional practice goals.

- Finally, let common sense prevail when considering the scope of your pilots. I recommend that to the extent practicable, districts pilot each potential DDM in at least one class in each school in the district where the appropriate grade/subject or course is taught. There is likely to be considerable educator interest in piloting potential DDMs in a no-stakes environment before year 1 data collection commences, so bear that in mind when determining scope.

Early Successes

I want to acknowledge the strong early work several districts have engaged in to be well positioned for piloting and eventual implementation. For example, Norwell Public Schools has already piloted DDMs for a wide array of educator roles, including teachers of art, music, and physical education, as well as guidance counselors, nurses, and school psychologists. Educators in Norwell received rigorous assessment literacy and data analysis professional development that enabled them to engage deeply in the process of refining their potential DDMs. West Springfield has partnered with the Lower Pioneer Valley Educational Collaborative to coordinate a DDM development project with several area districts. Their initial focus is to enlist educators to assist with the development of a common scoring rubric for a 10th grade writing to text DDM. Finally, recognizing the benefits of leveraging expertise across districts to identify potential DDMs, Chelmsford Public Schools has coordinated a curriculum leaders working group that include representatives from twenty-five districts. Additional details about these examples will be included in Technical Guide B.

Appendix B. Steps for Piloting a DDM

This appendix presents a deep dive into each suggested step of the pilot process. These steps are ESE's recommendation for a process that applies to each DDM that a district pilots. The majority of these steps apply to a team focused on a single measure. However, district-wide implementation is often a key shift in using an assessment as a DDM. For example, a district's pilot of a DDM may be focused on scaling up the implementation of a measure from use in only select classrooms or schools to district-wide use, with a focus on consistency in administration and scoring. (The administration protocol and scoring process should be common across all educators who administer the measure.)

The descriptions of the suggested steps are not intended to be comprehensive. ESE will continue to build on this information through the assessment literacy webinar series, technical assistance sessions, and other resources under development. Please see Section 5 for a complete list of the forthcoming ESE supports.

Step 1. Create a Team

Districts are required to pilot measures in five areas, at minimum, during the 2013-14 school year. This means districts and schools will pilot multiple DDMs concurrently. Therefore, districts should have a team engaged in district-wide DDM work. (Many districts have already created this team.) The beginning of the school year is a critical time to clearly define the roles, responsibilities, and expectations of each team member. In addition to the comprehensive team, there may also be teams with primarily school-level educators that are unique to individual measures being piloted.

Key Questions	Considerations

▪ Who will be on the team?
 	
  ▪ A team should work across the district and include members to represent the teacher voice. An ideal team incorporates knowledge of: content, special populations (students with disabilities and ELLs), feasibility (understanding how the assessment will impact a classroom), purpose (understanding how the assessment data will be used to inform curriculum, instruction, educator practice, and educator evaluation), and assessment development.
 	
▪ What will the team's responsibilities be? Responsibilities may include:
 	
  ▪ Providing professional development
 	
  ▪ Selecting measures to pilot
 	
  ▪ Creating a comprehensive action plan and timeline for piloting
 	
  ▪ Providing logistical and/or scheduling support to school-level educators involved in piloting measures
 	
  ▪ Ensuring that pilot steps are completed according to schedule
 	
  ▪ Collecting feedback from educators
 	
  ▪ Working with and across school-level educators to ensure consistent administration and scoring of measures
 	
  ▪ Analyzing DDM components and student results across all piloted DDMs to evaluate district-wide comparability
 	
▪ Will the team include partners from other districts, collaboratives, or other organizations, including vendors?
	▪ Some districts may choose to work as part of a regional consortium of districts piloting the same DDMs. For example, West Springfield has partnered with the Lower Pioneer Valley Educational Collaborative (LPVEC) to coordinate a DDM development project with several area districts, including superintendents, directors of curriculum, assessment experts from LPVEC, and school-level educators. Their focus is to enlist educators to assist with the development of a common scoring rubric for a 10th grade writing to text DDM.
	
	▪ Districts should also determine who will be engaged in piloting individual measures and explicitly identify roles and responsibilities. Teams focused on a single measure should include school-level educators who administer and score the measure; members may also overlap with the comprehensive team. For example, a team piloting a Grade 2 early literacy DDM may include the Director of Curriculum, one Grade 2 teacher from each of the district's elementary schools, and a principal from one of the elementary schools. It is most critical for these teams to include content knowledge.
	
	The team for a single DDM *may* be responsible for identifying the measure, developing or modifying components, collecting feedback from other educators on the administration and scoring, and/or creating a timeline for piloting, such as when the DDM will be administered, scored, analyzed, and modified.
	
	▪ Be clear about the decision-making process, particularly when working with educators piloting a DDM across multiple schools.

Step 2. Determine Content to Measure

Determining the content to measure is the first, and most critical, task in choosing DDMs. Clarity around the standards, content knowledge, and skills ensure that DDMs closely reflect the content taught. DDMs should reflect a meaningful sample of student learning that spans across the grade/subject or course. Technical Guide A provides multiple resources to support districts, including information about different assessment domains, curriculum maps, and an alignment tool. Many district teams are already well underway with this work.

Teams should start by considering the full range of what is taught in a particular grade/subject or course. From that full sample, teams determine what content is both representative of the most important content and measurable.

Key Questions	Considerations

▪ What content is most central to the curriculum and relevant to the school, district, and state goals?
   ▪ What are the most primary standards for this grade or subject/course?
   ▪ What are the important lesson objectives from the curriculum?
   ▪ What are the instructional shifts needed to effectively implement the new MA curriculum frameworks?
   ▪ What skills do students need to be college and career ready?
▪ Of the knowledge, skills, and abilities taught, what can be assessed?
▪ What content knowledge, skills, or behaviors are most directly attributable to an educator? For example, the rate of appropriately sequenced course enrollment may be a more appropriate measure for a high school guidance counselor than graduation rate, which is influenced by many educators.
	▪ Be mindful not to focus on what's easy to measure at the expense of what's important to learn.
	▪ A DDM does not need to cover the whole curriculum, but what is measured by the DDM should represent the most important concepts or skills taught.
	▪ Consider focusing on the instructional shifts needed to effectively implement the new MA curriculum frameworks.
	▪ Districts should work with specialized instructional support personnel (SISPs) such as school counselors, school nurses, or school psychologists to identify the responsibilities that are most critical to their role as a foundation to selecting DDMs. For example, a district may decide that the "return to class" rate is a critical responsibility for school nurses and should be an indirect measure.

ESE in partnership with WestEd convened panels of Massachusetts educators over the summer of 2013 to develop **Core Course Objectives** (CCOs) to inform the identification of example DDMs for over 100 non-tested grades/subject and courses. The CCOs are statements that describe different elements of essential content (knowledge, skills, or abilities) pulled from a larger set (MA Curriculum Frameworks, where available). The CCOs were identified using four key criteria: scope, assess-ability, centrality, and relevance. See Appendix B for detailed descriptions of these criteria, and learn more here: http://www.doe.mass.edu/edeval/ddm/identify-course.html.

Step 3. Identify Potential Measures to Pilot

Once educators have agreed on what content a DDM should measure, the next step is to identify the DDM. Districts, schools, or individual educators may already have an assessment that is well-aligned to this content or a measure that can be well-aligned with some modification.

Learn what is already in use. Many districts are beginning with the assessments that educators already use and find valuable. This often requires *learning* what assessments educators use that are either common across or unique to schools or classrooms. Districts who have not done so should survey their own educators to learn what measures are currently used across grades and subjects/courses district-wide. A number of districts have issued a survey to their educators to collect information on assessments used across all grades/subjects and courses.[21]

Key Questions	Considerations
▪ What measures do educators currently use across the districts? ▪ What objectives do current assessments measure? Are they well-aligned to the critical content identified in Step 2?	▪ ESE encourages districts to focus on scaling the use of assessments that are already used and valued by educators within the district. ▪ Many districts have noted a lack of alignment between some commercial assessments and local curriculum.

Learn what other options exist. Existing assessments that districts have identified to pilot include locally developed measures such as research papers, common writing assessments, portfolios, and capstone projects. Districts are also choosing commercial assessments, such as NWEA's MAP, Galileo, DIBELS, DRA, and Fountas & Pinnell. Teams in need of options should refer to ESE's list of example DDMs, which span many grades/subjects, courses, and assessment types.

> **Approach to Selecting DDMs.** Districts have three approaches when selecting DDMs:
>
> ▪ **Borrowing:** intact measures (e.g., a scoring rubric for a performance task) that are available to the public at no cost and do not require permission to use; a district pursuing this option would identify one or more existing open source assessments that closely align to the local curriculum.
>
> ▪ **Building:** released items, item pools, or sample items that a district can use to create customized tests for a particular grade or course; a district pursuing this option would identify items from multiple sources that can be used as a starting place for building a new district-specific DDM.
>
> ▪ **Buying:** commercial assessments that can be purchased from a vendor; a district pursuing this option would purchase measures closely aligned to local curricula.
>
> Whether borrowed, built, or bought, the assessment should be aligned to local curricula and informative to educators at the classroom-, school-, and district-level.

---

[21] This survey was created by the Massachusetts Teachers Association (MTA). See Appendix F.

Step 4. Prepare to Pilot

Once the district has decided on a measure to pilot, the next step is to prepare for piloting the measure. This includes determining what components need development or modification prior to administration, identifying focus questions, and creating a timeline for the next steps.

Piloting a *new* measure will result in valuable information for revisiting and refining each of these components where it is possible to do so. (These components of most commercial assessments cannot be modified.) Therefore, districts should prioritize development or modification that is critical to do prior to using the measure but should *not* expect to perfect all components of the DDM before administration and scoring.

Components of a DDM. In their final form, each DDM should have:

- *Instrument*, e.g., a set of test questions, a writing prompt, etc.

- *Directions for administering the measure*, e.g., how long the students have to complete the work, required materials, etc.

- *Student directions*, e.g., what students can and cannot say, do, or use; how much time they have, etc.

- *Scoring method*, e.g., answer key, rubric, etc.

- *Scoring directions*, e.g., single or multiple raters, a process for resolving differences in raters' scores for a student, etc.

Determine what components need development or modification. Within the options of borrowing or building, there is a range of potential development and modification. For example, a district that is using a freely available assessment (the "borrow" approach) may need to make decisions about local administration. When a district or group of districts is building a new measure, many of the components may need to be developed. However, the survey of educators described in Step 2 will likely have identified measures currently being used in a single school or classroom that could be scaled up for district-wide use. In this case, the primary development work may be developing a common administration protocol and/or a common scoring process. (See Key Considerations 2 and 3 on page 6.) A district may also decide to use an existing local assessment that does not currently measure growth; as part of the 2013-14 pilot, the district might decide to develop a pre-test or modify the scoring process so the DDM measures growth.

Key Questions	Considerations
▪ What components need development or modification before administration? ▪ Who is responsible for developing or modifying items, directions, rubrics, etc.? ▪ By when must the measure be ready to administer?	▪ First ensure content alignment and informativeness, then focus on Considerations 1-3 from Section 2 (pp. 5-8): measuring growth, common administration protocols, and common scoring process.

Applying the Key Questions and Considerations to Developing or Modifying a Measure

The two key questions from Section 2 are directly applicable to the work of developing or modifying a measure. The examples below illustrate modifying an existing measure; additional information about developing a new measure is available in Technical Guide A. This corresponds to the first DDM component: the instrument.

Key Questions	Examples
▪ Is the measure well-aligned to content? 　▪ Is the measure aligned to local curricula and to state and local priorities? 　　*See Step 2 on page B-3 for additional questions about content alignment.*	An end of year history exam is not well-aligned with the Common Core instructional shifts, because it does not include writing that is grounded in evidence from the text. As a result, two written response questions are added to the exam that require students to use evidence from the assigned reading
▪ Is the measure informative? 　▪ Do items/tasks allow students with different levels of ability to demonstrate their knowledge or skills? That is, will high and low achieving students have opportunities to be challenged and to succeed?	The identified DDM is a multiple-choice test that focuses on identifying multiple scientific properties and theories. As a result, the assessment does not give educators information about how well students can *apply* these skills. The district decides to reduce the number of recall-based multiple-choice questions and add multistep problem-solving tasks that demonstrate how successfully students apply higher order thinking skills.

Determining Administration Protocols

DDMs are intended to be "comparable across schools, grades, and subject matter district-wide." Part of that comparability is common administration procedures (see Appendix G). This includes directions for administering the measure and directions for students taking the test or performing the task(s). It is also important that districts provide clear directions about how to address common issues such as student absences and the need for accommodations or modifications.

Key Questions	Examples
▪ **When:** At what time is the assessment given? Is it a specific day or a window of time? How many times will the measure be administered?	An end-of-course exam may be administered on the same day or in the same week across a district. A running record may be administered once a month. The construction of an art portfolio may take place over the school year.
▪ **Environment:** What resources do students have? Are there clear directions to the teachers? How much time is allocated for the assessment?	Students may be allowed to use their notes on a chemistry exam or their calculators on an algebra exam. Students may be prohibited from helping one another on a writing assignment completed over the course of two weeks.

▪ **Student Instructions:** What questions can the students ask? ▪ What modifications or accommodations can be offered to students with disabilities or English language learners?	A group of teachers reviewing a performance assessment in a Massachusetts district described the assessment as unfair "because the lack of clarity and specifics in the project guidelines means that an 'A' student will get it, but a struggling student will require more guidance to be successful."[22]

Determining a Scoring Process. The protocols also need to provide clear guidance about how DDMs are scored, including how that data is stored and collected.

Key Questions	Examples

▪ **Scoring:** Who scores the results?
   ▪ If scored using a rubric, who is trained and who conducts the training?
   ▪ How is consistency across raters ensured?
	If educators are concerned that raters may rate writing samples differently, the district should ensure that: all raters are using the same rubric or scoring key; there is training and practice using that rubric; anchor papers are provided; and raters work toward calibration (e.g., raters discuss discrepancies in ratings and resolve those discrepancies).

▪ **Data Collection:** Who is collecting scores? Where is the data stored? ▪ Is the original student work being kept to refer to if problems are found later, i.e., responses and/or scores need to be examined for verification?	Districts currently use a variety of systems that range from simple (such as a Microsoft Excel tool) to sophisticated (such as software supported by a vendor). ESE is currently exploring current practices and options.
▪ **Computing:** If a growth score is computed from multiple scores, who is responsible?	In Norwell, the Executive Director of Instruction supported teachers by creating an Excel tool that allowed educators to input raw data and compute standardized scores that allowed for combining and comparing across measures.

---

[22] Brown, C. and Mevs, P. (2012). *Quality performance assessment: Harnessing the power of teacher and student learning.* Boston, MA: Center for Collaborative Education.

Piloting

Key Steps	Considerations
Test ▪ Administer the measure ▪ Score the results ▪ Collect the data	Districts will need to decide how many educators will pilot the measure. Commissioner Chester recommends, "…to the extent practicable, districts pilot each potential DDM in at least one class in each school in the district where the appropriate grade/subject or course is taught."[23] The number of administrations will vary by measure.
Analyze ▪ Analyze student results ▪ Analyze process	Districts should consider how to engage educators to provide feedback on the process of administering and scoring the measure. For example, a district that is piloting a measure with 20 teachers across a district may ask all of them to respond to a survey but only two might work on modifications to the instructions for administering the measure. The student results should also inform what modifications are made, particularly to the instrument. ESE will provide additional detail during the Assessment Literacy Webinar Series and Technical Assistance sessions to support districts in this work.
Adjust ▪ Modify DDM components as needed	The questions and considerations outlined on pages B-6 and B-7 will be revisited at this point. Districts may choose to adjust certain components that were not a focus prior to the first administration. ESE will provide additional detail during the Assessment Literacy Webinar Series and Technical Assistance sessions to support districts in this work.
Repeat ▪ Test, analyze, and adjust	This is an iterative process. Depending on the measure, this may happen once or more than once throughout the 2013-14 school year. The focus questions during the pilot are likely to change from one iteration to the next.

See Appendix C for suggested resources to support this work.

---

[23] From a memorandum dated August 15, 2013. Full text is available in Appendix A of this document.

Appendix C. Resources for Selecting and Piloting DDMs

Scoring Rubrics

These websites provide a list of potential resources specifically aimed at writing high quality rubrics, including an overview of rubric creation, examples, rubric banks, and samples:

- http://pandora.cii.wwu.edu/cii/resources/teaching_tips/assessment_grading.asp

- http://pandora.cii.wwu.edu/cii/resources/teaching_tips/rubrics.asp

Tools for Preparing and Piloting (From the Quality Performance Assessment)

The Quality Performance Assessment (QPA) initiative, launched by the Center for Collaborative Education, has published a guide and comprehensive series of tools designed to support schools and districts using performance assessments. Many of these tools may be applicable to a wide variety of measures, such as:

- a Calibration Protocol to calibrate scoring of student work;

- a Training With Anchors Protocol "to score compositions reliably using anchor papers, rubrics, scoring guidelines, and score reports;'

- a Tuning Protocol for Tasks to receive feedback and fine-tune tasks; and

- an Assessment Validation Checklist and Protocol to help school-level educators review assessment plans for effective design.

Interested educators may register to gain free access to these tools and protocols. A complete list of tools is available here: http://www.qualityperformanceassessment.org/publications-products/. See the QPA website for more information: http://www.qualityperformanceassessment.org/.

Appendix D. Linking Professional Practice S.M.A.R.T. Goals to DDM Piloting

Classroom educators are critical to the piloting process, as they will be responsible for administering and scoring the measure, analyzing the results, and providing feedback on the usefulness, clarity, feasibility, and so forth. One practical strategy to ensure educator engagement is to have educator professional practice goals that focus on their role in piloting DDMs.

Below, ESE shares a few samples of goals that have been developed by districts moving forward with this work in smart and strategic ways. These examples have not been refined by ESE and are still under development locally, but can be thought of as a starting point.

Chelmsford Public Schools: Draft DDM S.M.A.R.T. Goals

For CPS Made Assessment: By the spring of 2014, the _____ department will **develop** and pilot (Quantity) (Content Subject) common assessments in grade(s) _____, that meet the criteria established by the Department of Elementary and Secondary Education for District Determined Measures (DDMs), as measured by an updated department data inventory.

For Purchased or Use of State Exemplar Assessments: By the spring of 2014, the _____ department will **purchase/adopt** and pilot (Quantity) (Content Subject) (Name Assessment) in grade(s) _____, that meet the criteria established by the Department of Elementary and Secondary Education for District Determined Measures (DDMs), as measured by an updated department data inventory.

Make sure action plan defines:

- Training on developing DDMs
- The dates of development/purchase
- The dates of implementation
- Where the data will be housed and how it will be analyzed
- How the data will be used (placement, MTSS, instructional data meetings)

Norwell Public Schools: Template for Professional Practice Goals

Administrators: Establish a professional competency relative to student growth measures through guided discussions with faculty in regards to initial piloting and data norming, data analysis, revisions (as needed) to existing or previously developed measures, and the establishment of year one data collection activities in all grade levels and/or content areas as measured by two completed and standardized student growth measures in two distinct areas for all grade levels/content areas and one representative sample agenda from guided discussions/meetings.

Teachers: Establish a professional competency relative to student growth measures through guided discussions with grade level/department/content area colleagues and my primary evaluator, initial piloting, revisions (as needed) to existing or previously developed measures as measured by two completed and standardized student growth measures in two distinct areas of my curriculum and one representative sample agenda/representative notes from guided discussions/meetings.

Appendix E. Criteria for Core Course Objectives

Core Course Objectives are statements that describe different elements of core, essential content (knowledge, skills, or abilities) pulled, created, or synthesized from a larger set of curriculum standards. As a set, they are intended to be representative of a larger set of course objectives in important ways, but to reflect what many educators and other content experts working together agree are most critical in that content area, grade, or course. Like performance objectives or learning targets, core course objectives identify key learning objectives to which assessments can be aligned and, thereby, enable students to demonstrate what they know and can do in relation to course objectives that have been taught; educators can then use that information to improve instruction.

Scope

The Core Course Objective describes an overarching learning goal. It has enough scope (in breadth and/or depth) to subsume or integrate other smaller, more particular skills or concepts and to be addressed through a variety of student activities and assignments, including those requiring complex thinking skills. An example is a grade 8 mathematics course objective focused on defining, evaluating, and comparing functions; a number of more detailed course objectives are subsumed by this standard.

Assess-ability

The Core Course Objective describes knowledge, skills, or abilities that are assessable. It is clear, precise, and focused enough that it can be mapped to an appropriate assessment, which may include a post-instruction test of achievement (state or commercially available) delivered in paper-pencil or computer-supported format, or a locally-developed measure, such as a portfolio assessment, observation rubric, performance task, or culminating project.

Centrality

The Core Course Objective describes a critically important concept, skill, or ability that is central to the subject/grade or course. Some Core Course Objectives may describe high-priority knowledge or skills that are essential to student achievement at a particular grade level and foundational to success at higher grade-levels. An example is a grade 1 ELA course objective related to initial literacy (learning to read). Other Core Course Objectives may represent an enduring major idea, deep concept, or thematic question that is core to the subject and essential for all students to learn. An example is a grade 4 social sciences course objective focused on the rights and responsibilities of citizenship, a thematic concept that will be studied every year in Commonwealth social studies classrooms.

Relevance

The Core Course Objective represents knowledge, skills, and abilities that are consistent with Massachusetts' values and goals. While instructional priorities may vary from district to district, it is each panel's collective task to ensure that their final set of Core Course Objectives reflects the values and goals of educators from across the Commonwealth who are teaching that particular subject/grade or course.

For more information, please visit: http://www.doe.mass.edu/edeval/ddm/identify-course.html.

Appendix F. Sample Survey for Identifying Currently Used Assessments

This survey was developed by the Massachusetts Teachers Association and has been used in a number of districts across the state. Districts who have not identified the assessments that educators are currently using in classrooms and schools across the district may find this survey to be a helpful starting point or template.[24]

![MTA Massachusetts Teachers Association] District-determined Measures Survey

This is a confidential survey. The anonymous data is collected by the Massachusetts Teachers Association as part of the district's educator evaluation implementation. The faculty of each school will be provided with the aggregate results from the school and the district as a whole. It should take no more than 10 minutes to complete the survey.

QUESTION 1 - ASSESSMENTS

What ASSESSMENTS of student learning do you currently use in your classroom, caseload practice or school? (You will be asked below to list the different assessments. Do not include MCAS or MEPA in any of your answers.)

Type of Assessment	Never	Weekly	Monthly	Annually
Standardized tests	❍	❍	❍	❍
Textbook-based tests	❍	❍	❍	❍
Teacher-developed tests, e.g. quadratic equations, *Catcher in the Rye,* photosynthesis	❍	❍	❍	❍
Mid-Year & Finals common across subject, e.g. all students in the grade or subject take the same test at the same time.	❍	❍	❍	❍
Mid-year & Finals common across schools, e.g. all students in the grade or subject across all schools in the district take the same test at the same time.	❍	❍	❍	❍
Performance Assessments, e.g. lab experiment, speaking a second language, research project, serving a volleyball, solving a math problem	❍	❍	❍	❍
Writing Tasks, e.g. compositions, essays, reports, explanation of solving a math problem, lab report	❍	❍	❍	❍
Portfolios, collection of student work over time	❍	❍	❍	❍
Exhibitions, e.g. oral report, art show, musical performance, DECA competition, science fair, capstone project	❍	❍	❍	❍
Other	❍	❍	❍	❍

---

[24] Reproduced with permission from the Massachusetts Teachers Association.

Question 2 – Open Response

Please list the textbook titles of the textbook-base tests you currently use in your classroom, caseload practice or school:

QUESTION 3 – OPEN RESPONSE

Please list the unit titles for the teacher-developed unit tests you currently use in your classroom, caseload practice or school:

QUESTION 4 – OPEN RESPONSE

Please list the performance assessments you currently use in your classroom, caseload practice or school:

QUESTION 5 – OPEN RESPONSE

Please list the types of writing task assessments you currently use in your classroom, caseload practice or school:

QUESTION 6:

What data about student engagement, success, well-being, and/or behavior do you currently use in your classroom, caseload practice or school? Only check those that you use.

- ⭕ Attendance rates
- ⭕ Tardiness rates
- ⭕ Promotion rates
- ⭕ Graduation rates
- ⭕ College sending rates
- ⭕ Course taking patterns
- ⭕ In-school Suspension rates
- ⭕ Out-of school Suspension rates
- ⭕ Expulsion rates
- ⭕ Detention rates
- ⭕ Honor Roll
- ⭕ Awards, e.g. perfect attendance, good citizenship
- ⭕ Participation grade
- ⭕ Homework grade
- ⭕ Conduct grade
- ⭕ Industry/Effort grade
- ⭕ Other

QUESTION 7 – SCHOOL -- SELECT YOUR SCHOOL – SCHOOLS LISTED.

QUESTION 8 – SUBJECT(S) TAUGHT (SELECT ALL THAT APPLY.)

- ○ Arts (Art, Music, Drama)
- ○ English language arts
- ○ Foreign Languages
- ○ Health/Physical Education
- ○ History/Social Studies
- ○ Mathematics
- ○ Science
- ○ Technology (including vocational courses)
- ○ Other

QUESTION 9 - YOUR ROLE:

- ○ Classroom Teacher
- ○ Special Subject Teacher: art, music, library, computer, physical education, reading
- ○ Special Education Teacher
- ○ Counselor: Guidance, School Adjustment, School Psychologist, Case Manager
- ○ Therapist: Speech & Language, Occupational, Physical, Nurse
- ○ School-based Administrator
- ○ District-wide Administrator
- ○ Other _____

QUESTION 10 – GRADE (CHECK ALL THAT APPLY.):

- ○ Pre-K
- ○ Kindergarten
- ○ Grade 1
- ○ Grade 2
- ○ Grade 3
- ○ Grade 4
- ○ Grade 5
- ○ Grade 6
- ○ Grade 7
- ○ Grade 8
- ○ Grade 9
- ○ Grade 10
- ○ Grade 11
- ○ Grade 12
- ○ Ungraded

Thank you for taking the time to complete this survey.

Appendix G. A Quick Guide to Validity, Reliability, and Comparability

Validity and reliability are concepts that are frequently discussed in the context of assessments, particularly when the results will be used in educator evaluation. While they are both important, districts' first priorities are that the measures chosen closely align to content and provide useful information about student growth. That said, validity and reliability do not have to be intimidating concepts. The following section presents a brief introduction to these concepts. For more detail, see Technical Guide A.

Validity

Simply put, the test actually measures what it claims to measure. If it does, then the inferences, conclusions, and decisions made on the basis of test scores are appropriate and meaningful. For example, if a student reads a paragraph at his/her grade level for a reading comprehension test but has a reading ability significantly below grade level, a low comprehension score may be indicative of low decoding ability. It would not be valid to conclude that the student had low reading comprehension or that the teacher had a low impact on the student's reading comprehension. DDMs will need to:

- be aligned with the content of what districts expect to be taught in each classroom. This alignment does not require assessing every learning standard in the curriculum; it does require a sufficient connection to the most important standards.

- provide all students with fair and appropriate access to demonstrate what they know and can do. Tests must be free of bias. Accommodations available during instruction must also be available for the DDM; accommodations should not undermine what the test is intended to measure.

- be able to capture the full range of student performance, i.e., the DDM is not too easy or too hard for the vast majority of students, and reported in ways that are instructionally useful/actionable.

Reliability

Reliability answers the question, *"If I administered and scored the test again, would I get similar results?"* Reliability is the degree to which a test is consistent and stable in measuring what it is intended to measure, producing similar results under consistent conditions. For example, there should be enough items measuring a certain skill to ensure that the student does not get the questions right or wrong by chance; e.g., a single long division problem would not provide reliable information about how well a student had learned 5th grade math or how well she can solve long division problems. DDMs will need to:

- provide sufficient coverage of the standards and sufficient information about how a student can perform on each standard.

- be administered in a consistent manner across the grade/subject and across years.

- be scored precisely (according to pre-established criteria) and in a consistent manner, including when scoring involves judgments made by multiple scorers.

Comparability

DDMs must be "comparable across schools, grades, and subject matter district-wide." (Per [603 CMR 35.09(2)a](#))

"Comparable across schools" *pertains to educators in the same role.* ESE recommends that educators responsible for teaching the same grade/subject or course in a district use identical DDM(s) to ensure comparability across schools. However, there may be good reasons why a district would choose measures that are not identical across educators in the same role within or across schools.

- For example, all full first-year high school chemistry teachers or all 5th grade teachers in a district should typically administer the same DDM(s). The use of identical DDMs creates opportunities for educators to discuss results and share ideas for adjusting instruction, and allows for better calibration in scoring across educators.

- Educators with responsibilities that partially overlap may have one identical DDM and one that is different, such as a high school science teacher with four periods of chemistry and another with two periods of chemistry and two periods of biology.

- Some larger districts with elementary schools that serve different student populations have significantly different school-level goals. Such districts may choose to select DDMs that differ from one school to another in order to better align to school goals. For example, 5th grade teachers in School A may use an ELA and a math DDM, while 5th grade teachers in School B use the same math DDM but replace the ELA measure with a science measure because School B has a stronger STEM focus.

Comparable across grades and subject matter" *pertains to whether DDMs permit educators in different roles to demonstrate a similar range of impact on student learning.* DDMs provide the opportunity to compare student outcomes across different grades, subjects, and educator roles, but these comparisons must be fair and meaningful and result in accurate information.

- DDMs should not have significantly different levels of rigor such that it is more challenging for high school students to demonstrate growth in chemistry than it is for 5th grade students to demonstrate growth in ELA. The rigor of a measure relies on a variety of factors including its alignment to content, the range of difficulty of tasks or items, and the scoring criteria. Given the variation in assessment practices across different grades and subjects, it is natural that the rigor of DDMs will vary at first. The pilot year is a critical time to refine and adjust measures to work toward comparability across grades/subjects and courses.

- It is critical to consider whether a year's worth of student growth is comparable across different DDMs. This is also an opportunity to adjust for natural variation in levels of rigor. For example, if educators discover through piloting that a particular DDM is less rigorous than typical measures, districts might raise the standard for student outcomes to be considered more than a year's worth of growth.

- Ultimately, educator Student Impact Ratings must be consistent across educators, which relies on comparable measures and a comparable translation to student growth. For example, if a School A 5th grade teacher from the earlier example earned a rating of Moderate Impact based on an ELA and a math DDM, that should be comparable to a Moderate Student Impact Rating for a School B 5th grade teacher whose rating was based on the same math DDM and a science DDM.