Guidelines for assessing and improving District Determined Measures (DDMs)

Overview
Once districts have completed the preliminary stages of selecting, administering, scoring, and analyzing their DDMs of student learning, growth, and achievement, they will inevitably want to make adjustments to those measures based on lessons learned from the development and implementation processes. It will be important for districts to communicate to all stakeholders that DDMs are about creating a common framework to explore student learning.

Over time, districts will learn about which DDMs provide quality information about students and when a revision or replacement is necessary. ESE fully expects districts' DDMs to change as a reflection of lessons learned from implementation. This brief outlines the process and strategies that districts can use to improve DDMs over time.

This document includes guidance around improving the quality of DDMs and implementing a continuous cycle of improvement. As districts work to improve the quality of DDMs, both as tools to determine student growth and as evidence towards determining an educator’s impact on student learning, the continued refinement of these instruments is essential.

Suggested Next Steps
The recommendations in this brief can be helpful to districts as they proceed through the following stages of DDM development:

- Developing a long-term plan to modify or replace DDMs as needed.
- Developing a long-term plan to investigate issues of fairness.

A Note about Reliability
Along with validity, districts often express concern about measuring the reliability of their DDMs. Reliability is the degree to which scores might be different if the same assessment was given again to the same group of students.

As described in the Implementation Brief on Investigating Fairness, the amount of random error present in an assessment is what determines an assessment’s reliability. An assessment with high reliability has low random error, while an assessment with low reliability has high random error. Random error cannot be completely eliminated, but it can be decreased. Increasing the number of assessments, calibrating raters, or increasing the number of items on a single assessment decreases random error and improves reliability. As a general rule, the more students who take an assessment, the more reliable the assessment. As a result districts are encouraged to include all possible students in the content area of a given DDM.

Since a Student Impact Rating is based on the performance of many different students, on different measures, across multiple years, it is unlikely that a clear pattern and trend of student growth would be caused by random error. As a result, ESE encourages districts to focus on questions of validity (i.e., content alignment and utility) rather than on reliability.
Validity and Assessments

Many district and teacher leaders ask, “Given the practical realities facing our district, how do we ensure that we select valid DDMs?” Focusing on the quality of DDMs is important to provide the most meaningful and accurate results possible. However, this question also reflects two important misconceptions:

- First, validity is a characteristic of an interpretation or use of the assessment – not a characteristic of the assessment itself. The same assessment may be appropriately used to support one decision and not another.
- Second, validity is not a predetermined, fixed standard – instead it is a reflection of the level of confidence one has in a particular claim. As a result, a measure can always be improved.

For example, consider a vocabulary assessment where a student is given a list of 10 words to study at the beginning of a unit (pre-test) and then must spell and use them correctly at the end of the unit of instruction (post-test). A teacher who administers this type of vocabulary test will have clear evidence that a student who performs well on the post-test knows how to spell and use the 10 words after a week of concentrated study. In other words, the teacher might conclude that this test is a valid measure of whether a student can spell and use those specific 10 words at the end of the unit.

The teacher has less evidence that a student who performs well on the test will be able to spell and use the 10 words correctly at a later point in the year, for example, in her or his writing. Therefore, the test is less valid for this purpose. Finally, the teacher has almost no evidence that a student is able to spell and use words other than the 10 included on the test. As such, the vocabulary test is not a valid measure of a student’s ability to spell and use words generally. An interpretation of the validity of the vocabulary test or any assessment depends entirely on the conclusion the teacher is attempting to reach.

Validity is the extent to which the conclusions an educator is making from the results of the measure are well founded.

Many educators are concerned that, without extensive training in assessment, they are not able to assess the validity of assessments for use as DDMs. In actuality, teachers are best positioned to draw conclusions about the validity of the measures used in their classrooms because they are knowledgeable about the curricula.

If a measure is well aligned to content (has content validity) and provides meaningful information about student growth – the two essential characteristics of DDMs described in Technical Guide B – there is sufficient validity evidence to support the use of the measure as a DDM. A valid assessment must measure what the educator intends to measure.

Using DDMs in Educator Evaluation

Since the question of validity begins with interpretation and use, districts must think carefully about the conclusions drawn from the results of a DDM. The Massachusetts Educator Evaluation Framework calls for the use of DDM results and Student Growth Percentiles, where available, in an intentional way. A direct measure of student growth (either a DDM or median SGP from a state assessment) is designed to support the interpretation that a student demonstrated more (or less) growth than academic peers in the assessed area.

An individual measure of student growth is not a measure of an educator’s impact. There are always many factors that will influence any individual student’s learning, e.g., student effort, home environment, maturation, etc. It is only after evaluators and educators look at multiple measures of student growth across multiple years – and apply professional judgment to consider the measures and the learning context – that a determination is made about an educator’s impact on students.

This distinction is important because it highlights the dual role of each DDM as both a measure of student growth and one piece of evidence used to inform an educator’s Student Impact Rating. These two purposes of a DDM should guide the continuous improvement of measures over time. Ultimately all improvements should be aimed at improving a district’s confidence in using a DDM as both a measure of student growth and as evidence to support a claim about educator impact.

Cycle of Continuous Improvement

So how do educators approach the continuous improvement of DDMs? The graphic on the next page describes the steps districts should complete. Over time, each district will develop measures, procedures,
and structures that support meaningful use of DDM results.

Selecting DDMs: The step of selecting DDMs is not limited to the first year of implementation. Selecting the correct measure is the most important step for ensuring a valid interpretation of results. Districts should have an agreed-upon process in place to guide the revision of DDMs over time. Districts are encouraged to annually review measures used in the previous year and consider which measures need to be replaced or improved.

The use of new measures in subsequent years will not affect an evaluator’s ability to determine a Student Impact Rating informed by multiple years of data. Since each DDM is designed to measure student growth within a single school year, the use of identical measures from one year to the next is not required.

DDMs assigned to an educator may need to be changed for a variety of reasons, including changes in educator grade/subject or course, shifts in curricula or district priorities, or because a new measure becomes available. In fact, it is the expectation that many DDMs will be modified or replaced, especially in the early years of implementation, as each district determines where their agreed-upon measures are benefitting teachers and students and where different choices should be made.

ESE has produced guidance and resources to support districts in the selection of DDMs. (See call out box below.) As districts continue to develop new measures, ESE encourages cross-district sharing and collaboration.

ESE Resources for Selecting DDMs
- Technical Guide B
- Implementation Brief Series
- Using Current Assessments in DDMs
- DDMs and Assessment Literacy Webinar Series
- Example Assessments

Administering DDMs: Once a district has identified appropriate measures, the next step is to develop processes to ensure they are administered consistently and fairly. Problems with administration can interfere with the proper interpretation of DDM results. Inconsistencies in how a measure is administered can introduce differences in how groups of students perform based on factors unrelated to student ability and growth. These issues will lower confidence in the use of the results.

For example, student scores on an assessment of math skills may yield incomparable measures of growth if some students are allotted 15 minutes to complete the assessment, while others are allotted 30 minutes. (These considerations must be balanced against appropriate accommodations, which are designed to eliminate inequalities on the basis of disability or background.)

Standardized administration protocols are important to provide educators with a clear set of instructions for how to properly administer the DDM and should be part of the DDM development process. At a minimum, districts should be able to answer the following questions about each DDM:

- **When should the assessment be administered?** This may be defined by a time of year, before or after a particular unit or series of units, or at the beginning or end of a course.
- **How should the assessment be administered?** How much time – if it is a timed assessment – are students given to complete the assessment? How will results impact students? What tools or resources are students allowed to use to complete the assessment? What prior experiences should students have with the form and style of the assessment?
- **How will accommodations be documented?** How will the district ensure that all students receive appropriate accommodations, e.g., students with disabilities, English-language learners? Will modified versions on the assessment be available? If so, what are the criteria for using them?
How are deviations to the administration schedule addressed? What happens if a student misses an assessment? What happens when a student enters mid-year? What happens if the assessment is not given at the right time?

How is the assessment scored? What scoring rubrics, answer keys, and growth parameters are used to support consistency in scoring and parameter setting?

Collect Results: After educators have administered and scored DDMs, the results must be collected and organized. Decisions about data collection and use will depend on district priorities and available resources. Student Impact Ratings are the only data related to DDMs that will be reported to ESE. Decisions about how data connected to DDMs is collected and used are determined at the district level.

Districts should consider capacity and data use priorities when determining which data to aggregate at the central level. For example, some districts may establish structures or use existing structures, such as professional learning communities, to provide educators with opportunities to collaboratively review and discuss student results on DDMs with their colleagues and evaluators.

These conversations could be around student scores on DDMs or how students performed against the district’s parameters for high, moderate, and low growth for each DDM. Districts that take this approach may not find it necessary to collect student scores at the central level, but rather collect only the aggregated designations of high, moderate, and low growth for each educator’s DDMs. Conversely, some districts may prioritize central student score collection in order to facilitate student growth analyses across the schools in the district.

ESE encourages districts to continue to share effective practices around managing DDM results. To see one district’s approach to collecting and organizing results from DDMs using a straightforward Excel worksheet developed by district leadership, see Part 9 of the DDMs and Assessment Literacy Webinar Series. ESE is continuing to research ways to further support districts as they manage the data resulting from DDM implementation. Send your ideas to EducatorEvaluation@doe.mass.edu.

Sharing Feedback with Educators: The Educator Evaluation Framework is designed to provide meaningful feedback to educators. Educators and their evaluators can engage in powerful conversations about student growth by reviewing DDM results. Likewise, teams of educators can use DDM results to engage in collaborative inquiries into which practices most improve student outcomes.

For example, if the results of a writing DDM indicate that an educator’s students struggle to include details in their writing, she/he might work with her/his evaluator to identify opportunities to develop potential new instructional strategies, such as observing a colleague or receiving support from a literacy coach.

The educator might also confer with colleagues to look at the student results from all students who completed the DDM. This action can determine whether all students across the school or district struggled to include details in their writing or whether the results were unique to the educator’s specific set of students.

This type of analysis can help educators tease apart areas where changes in educator practice are needed, e.g., my students did not perform as well as students in other classes on the part of the assessment. It can also reveal more global instructional shifts or changes to the assessment should be made, e.g., all students seemed to struggle with this concept.

One important use for DDMs is to look across the students who are demonstrating low growth to determine what additional supports would help those students, or if students needed to be re-grouped for re-teaching.

Advanced Topic: Item Difficulty

The following section describes how to review item difficulty to design assessments that provide meaningful feedback to educators. This content is more advanced than the rest of the brief and is designed to introduce one strategy for looking at the quality of individual DDMs for districts that are ready to take on an interesting challenge. Computing item difficulty is not a requirement for DDMs.

Reviewing student performance on items of varying difficulty can yield meaningful feedback for educators. For example, if a disproportionate number of students struggle with certain “very easy” items, educators may identify core skills that need to be reviewed or re-taught.

Educators may also look at patterns within the measures themselves. For example, for assessments with individual items, educators can compute each question’s item difficulty. Item difficulty is computed simply by dividing the number of total points awarded to students on a given item by the total number of possible points.
Given the varied nature of DDMs, an “item” can take many forms, such as a single test question, a single performance task, or a single line from a rubric used to score a piece of writing. Item difficulty is expressed as a decimal between 0 and 1. An item difficulty very close to 1 means that most students earned all the possible points on the item, i.e., it is a very easy item, while an item difficulty close to 0 means very few students earned any points on the item, i.e., it is a very difficult item.

Examples:
1. 20/25 (80% or .8) students answer a multiple-choice item correctly. The item is of “moderate difficulty.”
2. 30 students complete a written response item, scored with a rubric from 0 to 4. Students scored as follows: 2 = 15 students; 3 = 10 students; 4 = 5 students.

The total number of possible points was 120 (4 points x 30 students). The total number of earned points was 80. The difficulty for this item is 80/120 (.66 or 66%). The item is “hard.”

Assessments should include items of varying difficulty. Understanding which items are more or less difficult for students to answer correctly provides important feedback to educators about areas where additional instruction may be needed. For example, one common problem with classroom assessments is that they are not typically designed with enough hard items to be able to determine if high performing students have demonstrated high growth.

These students hit the “ceiling,” i.e., the assessment suffers from a “ceiling effect.” By adding one or two very hard items to the assessment, it may make it possible to measure student growth more accurately at the high end of the scale. The scale below provides a good rule of thumb for interpreting item difficulty.

- **Easy items (> .9)**
  - Potentially yield little information because most students respond correctly
  - Important to have some items like this to prevent floor effects
- **Moderate items (.7-.9)**
  - Begin to provide information while still allowing most students to be successful
- **Hard items (.5-.7)**
  - Allow for distinctions between students
- **Very hard items (< .5)**
  - Potentially frustrating to students
  - Important to have some items like this to prevent ceiling effects

**Reviewing and Modifying DDMs**

The final step of the cycle of continuous improvement is to review and modify DDMs based on feedback from educators and perform an analysis of student results. As stated above, it is highly likely that DDMs will be revised or replaced, especially in the first several years of implementation.

For instance, an analysis of student results and consultation with educators could reveal that a DDM is sound in structure and content, but might be more meaningful if parameters for high, moderate, and low growth are adjusted. Likewise, a district might find that a DDM requires revisions to the administration protocol to promote improved consistency of implementation across educators.

Feedback from educators should be a driver of DDM review and modification. For example, the callout box below describes efforts taken by the Lincoln Public Schools to collect feedback regarding whether the district’s DDMs proved to be both “meaningful and manageable.” This feedback is used to identify areas of improvement.

Continued analysis of DDMs may reveal issues related to fairness, validity, and comparability. The reviewing and modifying phase of continuous improvement should include an examination of bias and a review of comparability within and across grades and subjects. For more strategies designed to improve DDMs, read the Implementation Brief on [Investigating Fairness](#).

**Lincoln Public Schools**

Lincoln Public Schools used the prompts below to facilitate group discussions with educators regarding the DDMs the district piloted in 2013-14. The educators’ feedback directly contributed to revisions to the district’s DDM Implementation Plan.

**Teacher Input for 2014-15 Plan for District-Determined Measures (DDM)** During this pilot year, teacher leaders and teachers have worked collaboratively to try out assessments that provide evidence of student growth. Our focus is to ensure these measures are “meaningful and manageable.” We are seeking the teachers’ perspective on the value of these measures and we will use their input for the development of the 2014-15 district plan.
Key Questions from the Massachusetts ESE initiative:

1. Does the measure effectively assess what is important for students to know and be able to do?
2. Does the measure provide valuable information to the teacher and the district about student growth?

From our district:

3. What kind of information from the initial baseline measure has been useful in planning instruction going forward?
4. In what ways do the results on the summative benchmark measure show growth when compared to the baseline measure?
5. Are there aspects of the baseline and/or benchmark measures that need revision? And if so, what revisions would make the measures more valuable? Should the measure be replaced by a different, more useful assessment?
6. What are some of the logistical considerations that will need to be addressed to make these measures both meaningful and manageable?

Frequently Asked Questions

How do I determine the validity of a DDM? Validity is the extent to which the conclusions an educator is making from the results of the DDM are well founded. When assessing the validity of a DDM, districts should be guided by the key principles outlined in Technical Guide B. A DDM must assess what is most important for students to learn and be able to do and what the educator intends to teach. A DDM must also provide valuable information to educators about their students and to the district about its educators. If a DDM does not meet these two goals for all like classes, there is little validity evidence for its use and districts might consider using a different measure.

How can a DDM be modified as part of a continuous improvement process without compromising the collection of trend data required to determine Student Impact Ratings? Each DDM is a measure of student learning, growth, or achievement within a school year. In the context of Student Impact

Reviewing for Accessibility and Universal Design

As part of the review and modify step, districts may want to analyze student results to identify DDMs or elements of DDMs that present barriers that interfere with students’ ability to demonstrate their knowledge. When reviewing assessment items for accessibility, educators should be mindful of the specific concept being assessed by the item and look for potential barriers that present challenges beyond the intent of the assessment. For example:

- Vocabulary words on a social studies assessment item that present challenges to students beyond the intent of the assessment.
- Cultural references in a language arts reading comprehension item that present challenges beyond the intent of the assessment.
- Writing skills required on a science assessment item that measure skills beyond the intent of the assignment.
- Reading comprehension skills required on a math assessment item that measure skills beyond the intent of the assignment.

While some assessment items are designed to measure integrated skills, educators will need to be clear about the nature of those skills. This review process can help refine the accuracy of assessment data and focus the intent of instructional practice.

ESE Resources for Accessibility

- Implementation Brief: Considerations for Special Education
- Implementation Brief: Considerations for English Language Learners

Ratings, a trend in student growth is established by examining the results of multiple measures across multiple school years. In this case, the term “trend” does not mean examining an increase or decrease in scores on a particular assessment from year-to-year.

As a result, aligning educators to the same measures for multiple consecutive years is not required. The evaluation framework must accommodate changes in measures from year-to-year for a variety of reasons, including continuous improvement, changes in educator grade/subject or course, and shifts in curricula or district priorities.