

**Question:** *The focus question for this brief.*

**How do I determine if a student has demonstrated high, moderate, or low growth?**

**District Readiness:** *Work to be completed before answering the focus question of this brief.*

- Engage educators in identifying assessments well-aligned to content and provide valuable information to educators about their students.
- Develop clear protocols to administer assessments fairly and consistently across all students, classrooms, and schools.

**Next Steps:** *Suggested next steps for districts after reading this brief.*

- Provide clear directions for scoring individual student work.
- Provide clear directions for determining a student's growth.
- Set the parameters for high, moderate, and low student growth.

### DDM Development

This Implementation Brief is relevant to the highlighted steps:

- Selecting
- Administering
- **Scoring**
- Analyzing
- Adjusting

This brief is aimed at districts that are ready to administer DDMs and looking for additional guidance. After districts have selected and administered DDMs, districts need to develop a plan for using those assessments to make a determination of whether students have demonstrated high, moderate, or low growth. There are three components that help to determine student growth. Those components are: a clearly written rubric or scoring guide for assessing student work, directions for making a growth determination and finally, parameters for what constitutes high, moderate, or low growth. Districts should engage educators in each step of this process. No single approach will be appropriate for all DDMs used in a district. Likewise, there may be multiple approaches that are equally appropriate for a given assessment. Some approaches may also combine these components. The sections below highlight some of the important considerations for each of these steps.

## Four Key Messages about Student Impact Ratings

These Implementation Briefs are designed to provide targeted guidance focused on timely questions around the implementation of District Determined Measures. These briefs highlight important questions, resources, and approaches for districts to consider, and are not exhaustive resources on the subject. Please continue to share your own examples, suggestions, and questions with us at [EducatorEvaluation@doe.mass.edu](mailto:EducatorEvaluation@doe.mass.edu).

### Use Multiple Measures



DDMs and SGP are part of comprehensive evaluation system

### Focus on Students



The focus of measuring student impact is improving student learning.

### Build Capacity



Developing DDMs builds knowledge about assessment and data use.

### Engage Educators



Educators have expertise developing and evaluating assessments.

## Considerations for Scoring Individual Student Responses:

After students have completed the assessment, the first step to making a determination of high, moderate or low growth for a student is the process of scoring. Scoring is the systematic process of assessing student work. Clear scoring guidance can support student learning and educator practice by explicitly stating the aspects of student work that are essential. Clear scoring guidance also supports fairness and transparency.

**Who Scores Student Work:** Districts will need to determine fair, efficient and accurate methods for scoring students' work. DDMs can be scored by the educator themselves, groups of teachers within the district, external raters, or commercial vendors. For districts concerned about the quality of scoring when educators score their own student's work, processes such as randomly re-scoring a selection of student work to ensure proper calibration or using teams of educators to score together, can improve the quality of the results. When an educator plays a large role in scoring his/her own work, a supervisor may also choose to include the scoring process into making a determination of a Student Impact.

Overly complicated questions may actually cause a student who understands the material better to be more likely to get the item wrong.

**Reviewing Scoring Procedures:** Content experts (e.g., teachers, coaches, department chairs, curriculum coordinators) should review scoring directions to determine clarity and item discrimination. Item discrimination is the degree to which a student's correct answer is associated with understanding the content taught. It is okay for a question to be hard, since difficult questions help make

distinctions between students who understand the material very well. However, overly complicated questions may actually cause a student who understands the material better to be more likely to get the item wrong. This can result in students with different abilities getting the same score. It should be clear to all outside content experts what would constitute a correct response. There are many freely available resources to support quality construction of [multiple choice](#)<sup>i</sup> and [open response](#)<sup>ii</sup> assessments.

**Writing Clear Scoring Guides:** Where appropriate, scoring guides should anticipate different types of partial credit and be as detailed and specific as possible about how scorers should assign partial credit. A lack of specificity can create challenges for scorers. For example, a scoring guide for a math question may state:

- 2 points for a correct answer with student work shown correctly
- 1 point for an incorrect answer with student work shown correctly

These directions are not clear about how to score the correct answer *without student work shown*, or what level of "work" is acceptable for partial credit. The example below addresses these concerns. Districts should anticipate refining scoring guides after scorers have had

experience with the assessments. For example, after using a scoring guide for a fractions assessment, educators may realize that it is not clear how to score correct answers that are not written in the lowest term. These types of revisions are a normal part of continuous improvement of an assessment. Districts should plan to review scoring guides after the first year of implementation.

- 2 points for a correct answer with either a chart or table showing how the student set up the problem.
- 1 point for an incorrect answer, but the work demonstrates setting up the problem with a table or picture. Supporting work may include incorrect numbers or other mistakes.
- 1 point for a correct answer and there is no supporting work or if student work is not organized in a table or chart.
- 0 point for incorrect/no answer and work is not organized in a table or chart.

**Rubrics:** Rubrics are an excellent resource for scoring performance assessments, constructed response and open-ended items. Here, we discuss different types of rubrics. See Figure 1 below for sample rubrics. Analytic rubrics define different characteristics to be observed, often in separate lines of the rubric. A holistic rubric may define the same characteristics, but asks the raters to consider them in a final rating. Since analytic rubrics break a larger observation into smaller components, it can be easier to achieve higher levels of inter-rater agreement. Analytic rubrics also provide clear formative feedback to educators and students about how to improve students' work since each component is clearly defined. Analytic rubrics typically assume an equal weighting of each category. Holistic rubrics have the inherent advantage that they rely on the rater's judgment and may capture important elements lost in an analytic rubric. Acceptable levels of agreement between raters can be achieved with holistic rubrics through the use of anchor papers and other methods for calibrating scores. Additional resources around the use of rubrics can be found [here](#).

### Computing a Growth Score:

**Technical Guide B** introduced four general approaches to measuring student growth. The following section focuses on measuring growth based on these approaches.

**Pre-Test/Post-Test:** Computing the difference between a student's pre-test and post-test is a straightforward way to measure student growth. This is called determining a "gain score." However, using a gain score to measure growth involves the assumption that an improvement of one point is equal across the scale. For example, consider a pre- and post-test that uses a 100 point scale. It is possible that moving from a pre-test score of 10 points to a post-test score of 20 points is much easier than moving from a pre-test score of 90 points to a post-test score of 100 points if the first 50 points reflects basic recall of facts, while the difference between 90 and 100 points reflects using that knowledge in sophisticated ways. Later in this guide, districts can see how to engage educators to consider gain scores at different points on the scale to

determine whether growth is roughly equal across the scale. It may not be possible to fully address this issue in year one, but **this is part of the continuous DDM improvement process that districts will conduct to monitor and improve their DDMs over time.**

**Repeated Measures:** Repeated measures provide additional information to teachers, such as *when* students made improvements across the year. This allows for a more sophisticated way of thinking about how students grew across the year. For example, if a behavior specialist was focused on reducing disruptive behavior with students, they may track outbursts in one observed class a week. This student might have made little progress for several months and then made a break-through during the month of January. This provides a different type of feedback to the educator than the student who made steady progress across the year. For the purposes of determining high, moderate, or low growth, districts can compute a growth score by taking the difference between an average of several observations. Using the previous example, this would mean taking the difference between the number of outbursts in the October observations subtracted from the number of outbursts observed in May.

**Examples of Rubrics designed to Measure Growth**

Analytic			Holistic		
Low Growth	Moderate Growth	High Growth	Low Growth	Moderate Growth	High Growth
0 or 2	3-5	6 or more			
Number of writing mechanics, such as punctuation, capitalization, misspelled word, where student has corrected the mistake in future writing			Little to no improvement in following writing conventions	Average improvement in following writing conventions	High improvement in following writing conventions
0	1	2 or more			
Number of examples of improvement of language usage and sentence formation, such as word order, subject-verb agreement, or run-on sentences student has corrected the mistake in future writing.					

Figure 1. Two different examples of approaches to constructing a growth rubric. Rubrics such as these could be used to make a determination of growth based on examples of student work collected throughout the year or course.

**Holistic Evaluation:** Rubrics are often used to measure student performance. However, they can also be used to measure growth. [Technical Guide B](#) described the use of growth rubrics in a holistic approach to measuring growth. Figure 1 presents two different growth rubrics designed to evaluate the growth of a student’s use of writing conventions across multiple writing pieces completed at different points in the year. This example was designed to highlight that growth rubrics can use either an analytic or holistic approach.

One advantage of this approach is that this rubric can be explicitly designed to translate into levels of high, moderate and low growth. As a result, it combines the process of scoring student work, determining growth, and setting parameters for high, moderate, and low growth. For

example, in the analytic rubric above it has been determined that demonstrating six or more examples of improvement in punctuation, capitalization, or misspelled words would demonstrate high growth. As a result, districts need to check that these categorizations match the districts understanding of high, moderate, and low growth.

**Post-Test Only:** Wherever possible, districts are encouraged to use other approaches to measuring growth than using a post-test only. However, there are situations where this approach is the best option. In a post-test only design, districts use other information to determine growth without administering a pre-test. This is appropriate when a district wants to use a high-quality well aligned measure of student achievement. For example, a district may choose to use the Calculus AP exam in a DDM. The AP exam is a rigorous measure well aligned to the content taught in an AP Calculus class. However, there is not a clear pre-test to the AP Calculus exam. As a result, a district may decide to use pre-calculus grades as a measure of student’s pre-existing skills. The post-test only model acknowledges the inherent differences in different students’ readiness to learn the material, and as a result provides a more fair assessment of an educator’s impact than achievement alone. The table below would be one example of how a district could use results from a previous course to make inferences about growth from the AP exam alone.

Evidence of Growth based on Pre-Calculus Grade and Calculus AP Exam

Previous Grade	Low Growth	Moderate Growth	High Growth
C	1	2-3	4-5
B	1-2	3	4-5
A	1-3	4	5

Figure 2. Example of how a district could use additional information besides a pre-test to make for a determination of low, moderate, or high growth.

It is important to acknowledge that this, like all assessments, is not a perfect solution. Engaging educators in the process of defining the parameters of growth will support later use of the results.

**Making a Determination of High, Moderate, or Low Growth:**

The last step in making a determination of high, moderate, or low growth is to complete the process of setting parameters. Districts should balance quantitative and qualitative approaches in this process. The following qualitative approaches rely heavily on educator’s professional judgment. Quantitative approaches provide important checks and balances on setting appropriate parameters.

**Qualitative Approach:** With both new and existing assessments, districts should engage educators to investigate the instruments themselves that will be used in the assessment. This work can be completed even before the assessment has been administered. For example, districts can ask these educators to imagine students at different levels and what level of growth would match their expectations. Educators can be guided by questions such as, “If a student scores a 5/20 on the pre-test and a 10/20 on the post-test, would that be below, at, or above your expectation of typical growth?” District teams should make sure to ask this question for hypothetical students with a range of ability levels, keeping in mind that a DDM should provide all students an equal opportunity to demonstrate growth. By engaging educators in this process it also supports later conversations around results. For example an evaluator could say, “you

said at the beginning of the year that more than 10 points of growth would be high growth, let's look at how many students demonstrated high growth."

Educators should be guided by their professional understanding of high, moderate, and low growth. There is no fixed number of points that represents each category of growth that will work for all DDMs. For example, average growth could be consistently earning the same score across each assessment if the expectation of performance (i.e., difficulty of the assessments) increased across the year. The two tables below show two students who demonstrated different levels of performance at three times across the year. A district would be justified in defining a student who demonstrated moderate growth as a student who consistently demonstrated the expected levels of achievement corresponding to the different points of the year. That is they performed and grew at the expected level. In contrast, the student who demonstrated lower levels achievement, but made improvement may have demonstrated high growth, even if they didn't reach the expected level by the end of the year. What is important are that these conclusions are supported by educator's professional judgment.

	Below	Approaching	Expectation	Advanced
Fall	√			
Winter		√		
Spring		√		

An Example of High Growth

	Below	Approaching	Expectation	Advanced
Fall			√	
Winter			√	
Spring			√	

Example of Moderate Growth

**Quantitative Approach:** If a district has historical data for a DDM, this can inform the process of setting parameters. For example, if a district has used the measure in past years, it is reasonable to look at the percentage of students that demonstrated different levels of growth. However, some DDMs will be new assessments and districts will not have this data available. Districts may also use results from DDMs to check the quality of parameters. Some different approaches include checking that determinations of growth would be similar for similar DDMs. For example, looking at a student's growth across two different math DDMs, a district could look to see what percentage of students were identified at the same level of growth by both measures. One would expect some agreement between these two measures. Districts can also look to see if the results produce enough variability in results. Given other evidence, a reasonable number of students should be rated as having high, moderate, and low growth. More information about the role of variability in DDMs can be found in the implementation brief on Fairness.

**Validating Results:** After a DDM has been scored, districts can also complete this validity exercise to check if the results from DDMs match educators overall impression of student's growth during that year. A district shouldn't expect perfect agreement, but this process can help reveal potential issues with a DDM. First, identify a team of educators knowledgeable about the students who completed the DDM. Next, randomly select one or two students who were determined to have each of the levels of high, moderate, and low growth. Next the team should look at other evidence of the student's growth, such as student effort, and teacher perspective, to see if the team generally agrees that the determination of growth is consistent with this other evidence. If there are differences, can they identify the cause? For example, a DDM may not allow a high performing student the ability to demonstrate growth. If there is a consistent disagreement between the DDM and the team's observation, this may require changes to the DDM, such as changes to the instrument, administration procedures, scoring process, determination of growth, or any other places to improve the quality of the determination.

## Frequently Asked Questions

**Do the same numbers of students have to be identified as having high, moderate, and low growth?** There is no set percentage of students who need to be included in each category. Districts should set parameters for high, moderate, and low growth using a variety of approaches.

**How do I know what low growth looks like?** Districts should be guided by the professional judgment of educators. The guiding definition of low growth is that it is less than a year's worth of growth relative to academic peers, while high growth is more than a year's worth of growth. If the course meets for less than a year, districts should make inferences about a year's worth of growth based on the growth expected during the time of the course.

**Can I change scoring decisions when we use a DDM in the second year?** It is expected that districts are building their knowledge and experience with DDMs. DDMs will undergo both small and large modifications from year to year. Changing or modifying scoring procedures is part of the continuous improvement of DDMs over time.

**Will parameters of growth be comparable from one district to another?** Different assessments serve different purposes. While statewide SGPs will provide a consistent metric across the Commonwealth and allow for district-to-district comparisons, DDMs are selected and used at the district level and are not intended to be compared across districts. While districts are encouraged to set parameters consistent with other districts, this is not a requirement.

---

<sup>i</sup> [http://www.theideacenter.org/sites/default/files/Idea\\_Paper\\_16.pdf](http://www.theideacenter.org/sites/default/files/Idea_Paper_16.pdf)

<sup>ii</sup> [http://www.theideacenter.org/sites/default/files/Idea\\_Paper\\_17.pdf](http://www.theideacenter.org/sites/default/files/Idea_Paper_17.pdf)