

Massachusetts Educator Evaluation

Rating Educator Impact:

The Student Impact Rating

May 2014

Massachusetts Department of Elementary and Secondary Education

75 Pleasant Street, Malden, Massachusetts 02148-4906

Phone 781-338-3000 TTY: N.E.T. Relay 800-439-2370

www.doe.mass.edu

Contents

Overview 1

The Student Impact Rating 1

Determining Educator Impact for Each Measure 3

Determining a Student Impact Rating 4

Reporting 6

Intersection between the Summative Performance Rating and the Student Impact Rating 6

Appendix A A-1

Overview

On June 28, 2011, the Massachusetts Board of Elementary and Secondary Education adopted new regulations to guide the evaluation of all educators serving in positions requiring a license. The regulations are designed to promote administrators' and teachers' professional growth and development, while placing improved student learning at the center of every educator's work.

The Massachusetts educator evaluation system is designed to allow educators and evaluators to focus on the critical intersection of educator practice and educator impact. Its two independent but linked ratings create a more complete picture of educator performance.

- The **Summative Performance 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 professional practice and student learning goals. This rating is the final step of the 5-step evaluation cycle.
- The **Student Impact Rating** is a determination of an educator's impact on student learning, informed by patterns and trends in student learning, growth, and/or achievement based on results from statewide growth measures, where available, and district-determined measures (DDMs).

Taken together, these two ratings will help educators reflect not only on their professional practice, but also the impact they are having on their students' learning.

This supplemental guidance document is intended to be a useful guide for educators and evaluators in the determination of Student Impact Ratings that meet the regulatory requirements. In addition to a review of the basic components of a Student Impact Rating, sample Student Impact Rating scenarios are included in [Appendix A](#).

Guidance on the determination of Summative Performance Ratings is also [available](#).

The Student Impact Rating

Evaluators are responsible for determining a Student Impact Rating of **high, moderate, or low** for each educator based on patterns and trends using multiple measures of student learning, growth, and/or achievement. Annual data for each educator from at least two measures is needed to establish patterns and trends.

- **Patterns** refer to results from at least two different measures of student learning, growth and achievement.
- **Trends** refer to results from at least two years.

Statewide growth measures (e.g., median MCAS student growth percentiles (SGPs)) must be used as one measure where available. However, while SGPs provide districts with a solid starting point for this work,

Changes in Statewide Assessments

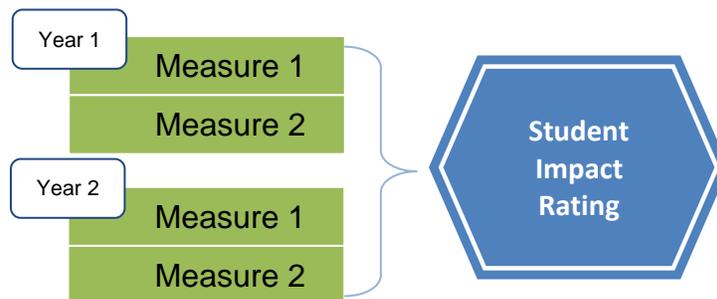
The educator evaluation regulations identify specific state assessments, MCAS and MEPA, for use in educator evaluation. ESE interprets the regulations to apply to statewide assessments that have replaced or will replace these named assessments. ACCESS for ELLs has replaced MEPA as the statewide assessment of English proficiency. Similarly, if Massachusetts implements the PARCC assessment, districts will be expected to use PARCC student growth percentiles as one measure for educators, where available.

they are available for fewer than 20 percent of educators throughout the state. Even where SGPs will be used, they will need to be supplemented with DDMs. As a result, districts will need to identify or develop DDMs for most grades and subjects, as well as for specialized instructional support personnel (SISP) and administrators.¹

Patterns and Trends

Student Impact Ratings must be based on patterns and trends in student learning, growth, and achievement. To establish patterns, evaluators and educators will collect data from at least two measures administered during the same school year. To establish trends, evaluators and educators will collect data from at least two different school years. Therefore, at a minimum, an educator’s Student Impact Rating will be based on four data points (see Figure 1 below).

Figure 1: Components of a Student Impact Rating



Establishing a **pattern** always requires at least two measures from a single school year. Since DDMs may sample from less than a full year of instruction, districts should examine the measures that will inform each educator’s Student Impact Rating to ensure that a determination of impact relative to one year of growth can be made with confidence. For example, if a district is using a rigorous performance assessment from a single unit, balance might be achieved by having these educators use a portfolio assessment that involves applying a growth rubric to student work collected across the school year. Another means of achieving balance in the case of shorter interval DDMs is to match educators with more than two measures.

In the context of Student Impact Ratings, a **trend** is the result of 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 changes in educator grade/subject or course

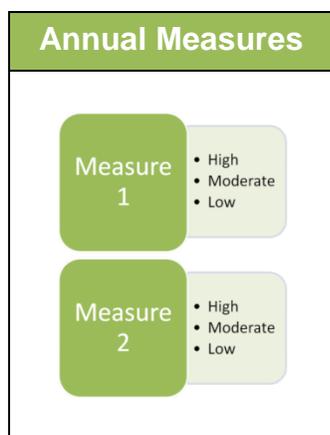
What About Educators Who Change Districts?

If an educator changes districts, the confidentiality provisions of the educator evaluation regulations preclude a trend to be derived using data from the former district the first year and the current district the second year. In such cases, the educator would collect the first year of trend data in his/her first year in the new district.

¹ See the Implementation Briefs on [Indirect Measures and SISP](#) and [Administrators](#) for additional information and guidance on identifying DDMs for these educator roles.

assignment and shifts in curricula or district priorities. For example, if a second-grade teacher’s students demonstrate high growth on two measures in a given year, the teacher does not lose that evidence of impact if he/she is transferred to the fifth-grade the next year. If the teacher’s fifth-grade students, likewise, demonstrate high growth on two measures, the teacher can earn a Student Impact Rating of high based on patterns and trends. A different example is an educator who possesses multiple licenses and may move to a different content area from one year to the next. A trend in impact on student growth could be established for an educator who teaches eighth-grade mathematics one year and sixth-grade science the next by looking at student growth results on measures from each of the two years (See Appendix A for more examples). DDMs should be designed to yield a determination of student growth within a school year and should be comparable. As a result, an educator’s Student Impact Rating may be based on measures from different grade-levels and content areas.

Determining Educator Impact for Each Measure



Determining an educator’s Student Impact Rating begins by investigating how an educator’s students performed on individual growth measures each school year. ESE recommends that evaluators meet annually with each educator to discuss student performance on measures. Educators should have an opportunity to describe the student profile of the class, including learning needs and challenges, and the specific learning context in which the measures were administered. Educators should also be consulted regarding the usefulness of the measures, particularly in the first few years of implementation. The annual meeting at the end of the first year of collecting trend data should serve as a midpoint check-in and will help further the goal of “no surprises” when evaluators determine Student Impact Ratings.² In the case of educators on two-year self-directed growth plans, this meeting could be conducted in conjunction with the formative evaluation.

The process for determining whether an educator’s students, on average, demonstrated high, moderate, or low growth is different for statewide growth measures than for DDMs.

Statewide growth measures: Student growth determinations are based on an educator’s students’ median Student Growth Percentile (mSGP). The educator’s mSGP will result in a designation of **high, moderate, or low impact** of the educator for the statewide growth measure.

Determination of Growth for Teachers and Administrators based on Median SGP

	Low	Moderate	High
Teachers	35 or lower	> 35, but < 65	65 or higher
Administrators	40 or lower	> 40, but < 60	60 or higher

² For educators who will use median SGPs as one measure, the timing of the meeting between the educator and evaluator to discuss the first year of trend data may need to be shifted in order to accommodate the timing of the release of SGPs.

- For more information about using SGPs as a measure to inform Student Impact Ratings, including the rationale for establishing different parameters for administrators and teachers, review the [Implementation Brief on Using Student Growth Percentiles](#).

DDMs: Since a district’s collection of DDMs will be comprised of an array of assessment types and a variety of methods for measuring student growth, the process of determining whether students have demonstrated high, moderate, or low growth will be more involved than for statewide growth measures. The process of setting these **district parameters** for high, moderate and low growth for each DDM should be informed by the professional judgment of educators and other content experts in the district and should reflect reasonable expectations for student growth outcomes on the measure. Engaging educators in the process of setting district parameters will ensure that teachers using the same DDM understand expectations for student growth.

- For more information about establishing district parameters for high, moderate, and low growth, review the [Implementation Brief on Scoring and Parameter Setting](#) or [Part 8 of the DDMs and Assessment Literacy Webinar Series](#).

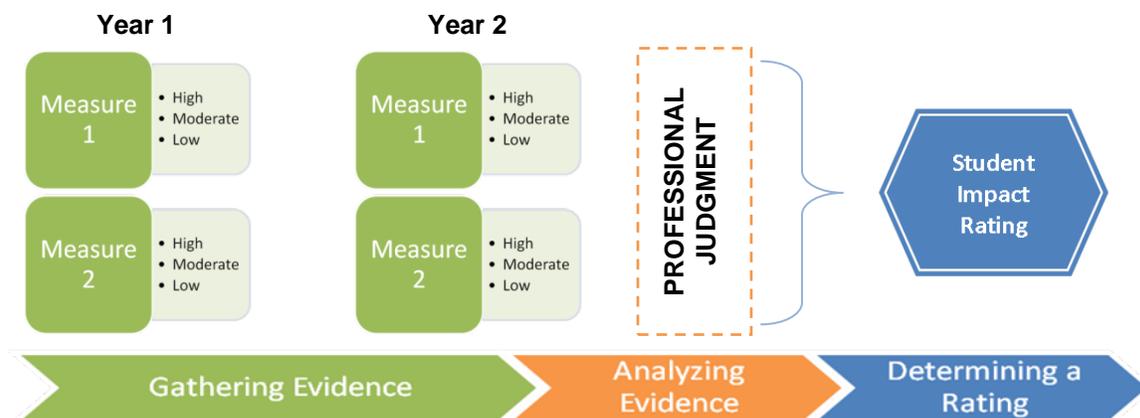
Once a DDM has been administered and scored, the evaluator should consult with the educator and determine whether the educator’s students demonstrated high, moderate, or low growth in comparison to the district parameters for the specific DDM. This consultation may occur during the meeting between the evaluator and educator described above. The evaluator’s determination will result in a **designation of high, moderate, or low impact** of the educator for each DDM.

Evaluators and educators should have a shared understanding of the educator’s designation of impact for all measures administered during a given school year by the beginning of the next school year. ESE expects that SGPs will be released to districts in late summer, while DDM results are likely to be available before the close of school each year.

Determining a Student Impact Rating

Once designations of high, moderate, or low impact have been established for at least two measures in each of at least two years (patterns and trends), the evaluator has enough information to determine the educator’s Student Impact Rating. Figure 2 illustrates the process by which an evaluator determines a Student Impact Rating. Professional judgment plays a key role in the determination of this rating.

Figure 2: Determining a Student Impact Rating



The Role of Professional Judgment

How does an evaluator know how to use pattern and trend data? How do these data translate into an overall Student Impact Rating?

There are no weights or percentages that dictate how an evaluator must interpret pattern and trend data to determine a Student Impact Rating for an individual educator. Rather than adopt a more mechanistic, one-size-fits all approach to supervision and evaluation, the Massachusetts evaluation framework places paramount importance on evidence and the professional judgment of evaluators and educators in the evaluation process.³ While the scoring of individual measures and the determination of whether students demonstrated high, moderate, or low growth must be based on agreed-upon, transparent methods that are appropriate for each measure and fair to students and educators alike, formulaic or numerical processes that preclude the application of professional judgment and dictate how evaluators must use pattern and trend data to determine Student Impact Ratings are inconsistent with the letter and the spirit of the evaluation framework (see sidebar for regulatory requirements).

The use of professional judgment promotes a more holistic and comprehensive analysis of impact, rather than over-reliance on one individual data point or rote calculation of impact based on predetermined formulas. By emphasizing the application of professional judgment, the Massachusetts framework recognizes that student outcomes on state- and district-determined measures should not be considered in a vacuum. Evaluators are encouraged to bear in mind an educator's student population and specific

instructional context. Considerations related to a specific measure or a combination of measures may also factor into an evaluator's determination of an educator's Student Impact Rating. For example, it is likely that as districts try out DDMs, some will prove more meaningful than others. Suppose in the first year of implementation a pre-/post-DDM is found to suffer from a significant ceiling effect because a high number of students earn perfect scores on a pre-test. An evaluator may decide not to consider the results of this DDM as heavily as other measures.

Are DDMs and SGPs measures of educator impact?

DDMs and SGPs are measures of a student's growth. Student results provide information about student growth, but alone do not represent measures of educator impact. SGPs and results from DDMs serve as pieces of evidence, used by evaluators in the determination of an educator's Student Impact Rating. It is only once professional judgment is applied to patterns and trends in student growth; taking into account the measures and the instructional context that a determination of educator impact is made.

Evidence and professional judgment shall inform:

- a) the evaluator's ratings of Performance Standards and overall educator performance [Summative Performance Rating]; and
- b) the evaluator's assessment of the educator's impact on the learning, growth, and achievement of the students under the educator's responsibility [Student Impact Rating]. [603 CMR 35.07\(2\)](#)

³ "Evidence and professional judgment shall inform... the evaluator's assessment of the educator's impact on the learning, growth, and achievement of the students under the educator's responsibility" ([603 CMR 35.07\(2\)](#)).

When professional judgment does not lead to a clear Student Impact Rating, ESE recommends that evaluators employ the following rules of thumb:⁴

1. If more than half of the designations of impact derived from measures point to the same level (high, moderate, or low), then that level should be the educator’s overall Student Impact Rating.
2. When there is no clear conclusion to be drawn from patterns and trends, ESE recommends that moderate impact be the default rating unless there is compelling evidence pointing to a different conclusion.

For examples of this process as applied to individual teachers, specialized instructional support providers, and administrators, please see [Appendix A](#).

Reporting

Districts will implement DDMs and collect the first year of Student Impact Rating pattern and trend data during the 2014-15 school year. Year 2 data will be collected during the 2015-16 school year. Initial Student Impact Ratings of high, moderate or low will be determined following the 2015-16 school year and reported to ESE in October of 2016. Districts that have agreed to determine Student Impact Ratings using three-year trends will report initial Student Impact Ratings based on a two-year trend in 2015-16 and may use a three-year trend thereafter.

Intersection between the Summative Performance Rating and the Student Impact Rating

As described above, a Summative Performance Rating is a rating of educator practice and a Student Impact rating is a rating of educator impact on student learning, growth, and/or achievement. These two ratings are independent, but intersect to provide educators and evaluators with a more complete picture of educator effectiveness. This intersection results in a number of opportunities for educator growth and development.

Figure 3: Intersection of Summative Performance and Student Impact Ratings

Performance Rating	Exemplary	1-yr Self-Directed Growth Plan	2-yr Self-Directed Growth Plan		
	Proficient				
	Needs Improvement	Directed Growth Plan			
	Unsatisfactory	Improvement Plan			
		Low	Moderate	High	
		Impact Rating			

⁴ ESE’s “rules of thumb” were first introduced in [Part VII](#) of the Massachusetts Model System for Educator Evaluation (pgs 27-28) published in August 2012.

Type and Length of Plan: The Summative Performance Rating and Student Impact Rating are used together to determine the type and length of an educator’s Educator Plan. The Summative Performance Rating determines the type of plan and the Student Impact Rating may have a direct impact on the length of the plan. The educator evaluation regulations define four different types of Educator Plans.⁵

- The ***Self-Directed Growth Plan*** applies to educators who earn Summative Performance Ratings of *Proficient* or *Exemplary*. It is developed by the educator. Educators placed on Self-Directed Growth Plans who earn Student Impact Ratings of *moderate* or *high* are placed on two-year plans, while educators who earn impact ratings of *low* are placed on one-year plans.
- The ***Directed Growth Plan*** applies to educators who earn Summative Performance Ratings of *Needs Improvement*. It is developed by the educator and evaluator and is one year or less in duration.
- The ***Improvement Plan*** applies to educators who earn Summative Performance Ratings of *Unsatisfactory*. It is developed by the evaluator and is 30 days to one year in duration.
- The ***Developing Educator Plan*** applies to educators without Professional Teacher Status (PTS), administrators in their first three years in a district, or educators in new assignments (at the discretion of their evaluators). It is a one-year plan developed by the educator and evaluator.

Rewards and Recognition: The educator evaluation regulations also describe Summative Performance and Student Impact Rating combinations that result in opportunities for educator recognition and reward.

- Educators with PTS who earn Summative Performance Rating of *Exemplary* and Student Impact Ratings of *moderate* or *high* will be recognized and rewarded. The regulations reference leadership roles, promotions, additional compensation, public commendation, as possible acknowledgements for these high performers.⁶
- Similarly, educators who earn Summative Performance Ratings of *Proficient* and Student Impact Ratings of *moderate* or *high* may be eligible for additional roles, responsibilities, and compensation.⁷

Investigating Discrepancies: One of the greatest benefits of the Massachusetts educator evaluation system is the ability to investigate educator practice independent from educator impact and vice versa. A discrepancy between an educator’s Summative Performance and Student Impact ratings serves as a signal for further professional conversation between the evaluator and educator, which may include looking at additional evidence related to student outcomes and/or practice. If there is such discrepancy between the Summative Performance and Student Impact ratings, the educator and evaluator are expected to analyze the discrepancy and seek to determine the cause. In fact, the educator’s Educator Plan might include one or more goals directly related to examining elements of practice that might be

⁵ [603 CMR 35.02](#)

⁶ [603 CMR 35.08\(7\)](#)

⁷ [603 CMR 35.06\(7\)\(a\)\(1\)\(c\)](#)

contributing to low impact. At the district-level, investigating discrepancies could identify the need to provide additional evaluator training or the need to refine certain DDMs.

In addition, the regulations require that the evaluator's supervisor review the performance rating with the evaluator when a notable discrepancy occurs.⁸ When there are significant discrepancies between impact ratings and summative performance ratings, the evaluator's supervisor may note these discrepancies as a factor in the evaluator's own evaluation.⁹ The supervisor is expected to look for consistency in application of Standards and Indicators as detailed in the district's rubrics, and evaluator's professional judgment.

These checks and balances made possible by investigating discrepancies between Summative Performance and Student Impact ratings are designed to promote deeper inquiry and help foster continuous improvement through close attention to all three elements that determine evaluation ratings: educator practice, educator impact on student learning, and evaluator consistency and skill.

⁸ [603 CMR 35.06\(7\)\(a\)\(2\)\(a\)](#)

⁹ [603 CMR 35.09\(4\)](#)

Appendix A

Student Impact Rating Examples

The following examples are intended to present scenarios that illustrate how professional judgment may be used to determine a Student Impact Rating.¹⁰ The scenarios are fictitious. Each example illustrates a distinct, but plausible, set of circumstances. However, based on additional contextual information not provided in these examples, an evaluator might reasonably arrive at different conclusions than those provided.

Example A: Robert Brookings (High School Math Teacher)

Year	Measure	Designation of Impact
1	Geometry Pre-Post Test DDM	Moderate
1	Geometry Problem Solving DDM	Moderate
2	Algebra Growth Rubric DDM	High
2	AP Calculus Exams	Moderate

Robert Brookings is a high school math teacher. In Year 1, Mr. Brookings taught 10th grade Geometry. His students demonstrated moderate growth on the two geometry DDMs he administered during the year, resulting in two “moderate” designations of impact. In Year 2, Mr. Brookings taught 9th grade Algebra and Advanced Placement (AP) Calculus. He demonstrated high impact on the Algebra DDM his students completed and moderate impact using AP Calculus exam scores. Even though the measures used in Year 1 and Year 2 are not the identical, a pattern can be established because multiple measures were used each year and trend can be established because there are results from measures administered in multiple years. Mr. Brookings and his evaluator discuss the composition of his classes and the learning context and determine that the designations of impact are an accurate reflection of student growth. Mr. Brookings’ evaluator concludes that his Student Impact Rating is **moderate**, a decision consistent with one of ESE’s recommended rules of thumb (if more than half of the designations of impact point to the same level, then that level should be the educator’s overall Student Impact Rating).

Example B: Lily Martin (4th Grade ELA Teacher)

Year	Measure	Designation of Impact
1	4 th grade MCAS Math SGP	Moderate
1	Writing to Text Portfolio DDM	Moderate
2	4 th grade MCAS Math SGP	Low
2	Writing to Text Portfolio DDM	Low

Lily Martin is a 4th grade classroom teacher. Ms. Martin administered the same DDMs in Year 1 and Year 2. In Year 1, both DDMs resulted in designations of moderate. In Year 2, both DDMs resulted in designations of low. The evaluator meets with Ms. Martin to understand more about the circumstances that led to designations in Year 2 that were lower than those in Year 1. They discussed differences in her students’ learning needs and identified new instructional strategies for Ms. Martin to try out the following year. Applying professional judgment to all of the evidence, the evaluator concludes that Ms. Martin’s Student Impact Rating is **moderate**. This is consistent with one of ESE’s recommended rules of thumb (when there is no clear conclusion to be drawn from patterns and trends, moderate impact should be the default rating unless there is compelling evidence pointing to a different conclusion).

¹⁰ All examples provided in Appendix A are fictional. Any connection to real educators is entirely coincidental.

Note: The evaluator’s determination would not have been different if the order of designations had been reversed. In other words, Ms. Martin’s Student Impact Rating would have been **moderate** even if she had earned designations of low in year 1 and moderate in year 2.

Example C: Mark Anderson (7th Grade ELA Teacher)

Year	Measure	Designation of Impact
1	Reading Improvement Inventory DDM	Low
1	Academic Vocabulary Pre-Post DDM	Moderate
2	Reading Improvement Inventory DDM	Low
2	Academic Vocabulary Pre-Post DDM	Moderate

Mark Anderson teaches 7th grade ELA. Mr. Anderson administered the same DDMs in Year 1 and Year 2. Mr. Anderson’s evaluator must consider two designations of moderate impact and two designations of low impact to determine a Student Impact Rating. Mr. Anderson and his evaluator meet to discuss the measures and agree that they both have considerable confidence in the results from the Reading Improvement Inventory DDM. The assessment has been in use for several years to measure student growth in reading skills. With multiple years of results, the educators in the district are very confident in the determinations of high, moderate, and low growth for students. In contrast, the academic vocabulary pre-post DDM was new to the district in Year 1 and the district’s parameters for determining growth resulted in nearly all students falling into the moderate growth band. As additional context, the evaluator notes that Mr. Anderson’s students’ scores on the academic vocabulary DDM were considerably lower on average than typical student results, despite the composition of his classes being typical for the district. As a result the evaluator determines that Mr. Anderson’s Student Impact Rating is **low**.

Note: In this example, the evaluator could have employed one of ESE’s recommended rules of thumb (when there is no clear conclusion to be drawn from patterns and trends, moderate impact should be the default rating unless there is compelling evidence pointing to a different conclusion) to arrive at a rating of moderate. However, the evaluator considered the circumstances and applied professional judgment to arrive at a rating of low. ESE’s recommended rules of thumb are meant to guide the decision-making process, but are not requirements.

Note: An evaluator could reach a different conclusion based on the same pattern of impact designations for the DDMs. For example, if an evaluator had similar confidence in the academic vocabulary DDM, Mr. Anderson’s Student Impact Rating would likely have been determined to be **moderate**.

Example D: Karen Jackson (High School Assistant Principal)

Year	Measure	Designation of Impact
1	10 th Grade ELA median SGP (58)	Moderate
1	10 th Grade Math median SGP (85)	High
1	9 th and 10 th Grade Attendance Rate (98%)	High
2	10 th Grade ELA median SGP (57)	Moderate
2	10 th Grade Math median SGP (83)	High
2	9 th and 10 th Grade Attendance Rate (94%)	Moderate

Karen Jackson is the assistant principal for a small high school. She is responsible for supervising instruction of 9th and 10th grade students, so the district is required to use median 10th grade SGPs to inform Ms. Jackson’s Student Impact Rating. Ms. Jackson’s evaluator consults the [Implementation Brief on the Use of SGPs](#) to learn more about required use of SGPs and parameters for high, moderate, and low impact. The district opts to use both math and ELA SGPs, in order to ensure that the Student Impact Ratings for Ms. Jackson and her peers reflect a broad range of

instructional content. The district also selects an indirect measure of student learning: attendance rate. The designation of impact for attendance is based on historical attendance data. When meeting with her evaluator, Ms. Jackson points out that the designations of moderate from the SGP data were very close to the cut scores for high, and the designations of high were well above the threshold. Considering all of the evidence, Ms. Jackson's evaluator concludes that her Student Impact Rating is **high**.

Example E: Liza Chase (Middle School Music Teacher)

Year	Measure	Designation of Impact
1	Music Terminology DDM	Low
1	Band Performance DDM	High
2	Music Terminology DDM	Low
2	Band Performance DDM	High
2	Music Composition DDM	High

Liza Chase teaches Music in a middle school. Ms. Chase administers two DDMs in Year 1 and three in Year 2. The district adds a Music Composition DDM in Year 2; her other DDMs remain constant in Years 1 and 2. Ms. Chase's evaluator is confronted with contradictory evidence of her impact. The music terminology (the ability to sing from written music without practice) measure Ms. Chase administered in both years resulted in designations of low impact each time, but her other DDMs suggest high impact.

The evaluator understands that the results of each DDM are just one piece of evidence of how one group of students grew on one assessment. As a result, this is not the first or the last time the evaluator expects to have to resolve this type of inconsistency. Using one of ESE's recommended rules of thumb (when there is no clear conclusion to be drawn from patterns and trends, moderate impact should be the default rating unless there is compelling evidence pointing to a different conclusion), Ms. Chase's evaluator could conclude that her Student Impact Rating is **moderate**. However, the other rule of thumb (if more than half of the designations of impact point to the same level, then that level should be the educator's overall Student Impact Rating) could apply to this scenario as well and result in a Student Impact Rating of **high**.

After discussing the results with Ms. Chase, the evaluator applies professional judgment and determines that the circumstances warrant a Student Impact Rating of **high**. They agree that the DDMs based on band performance and music composition are better indicators of her impact than the music terminology DDM. The evaluator and Ms. Chase also discuss why students are not demonstrating expected growth on the music terminology DDM. They discuss both the quality of the assessment to determine whether refinement is needed as well as Ms. Chase's instructional strategies. The evaluator flags this conversation as important information for Ms. Chase to consider during her next self-assessment.