

Evaluator Calibration: Best Practices for Inclusion

June 2016

- *Toolkit for districts aiming to improve evaluator calibration and support inclusive practices for general education classrooms*
- *Informed by calibration and inclusion work from the Everett and Milton public schools*



Closing the Equity Gap

Excellent teachers and leaders are critical to **improving student achievement**.

Our **moral imperative** is that all students should have **equitable access** to great educators.

Context

In the 2015-16 school year, the Department worked with nine districts who grappled with the question, “Which students have experiences with the most effective educators?”

Everett and Milton public schools undertook work to ensure equitable access to educators through the [Educational Equity Professional Learning Network](#) (PLN). This network of nine diverse districts developed strategies addressing inequities in students’ access to experienced, prepared and effective educators. The PLN is one of the key strategies in the 2015 [MA Plan for Equitable Access to Excellent Educators](#).

Milton and Everett selected strategies to calibrate their vision of effective teachers – especially around inclusive practice.

Both districts selected strategies to better calibrate

Laying the groundwork

In Fall 2015, PLN districts held self-assessment sessions to analyze data and root causes, and to select strategies to address equity gaps. Districts assembled teams of stakeholders including data specialists, school leaders, SPED directors, curriculum directors, teachers’ union president, superintendent, and assistant superintendent.

Including a data specialist on the team was useful for PLN districts.

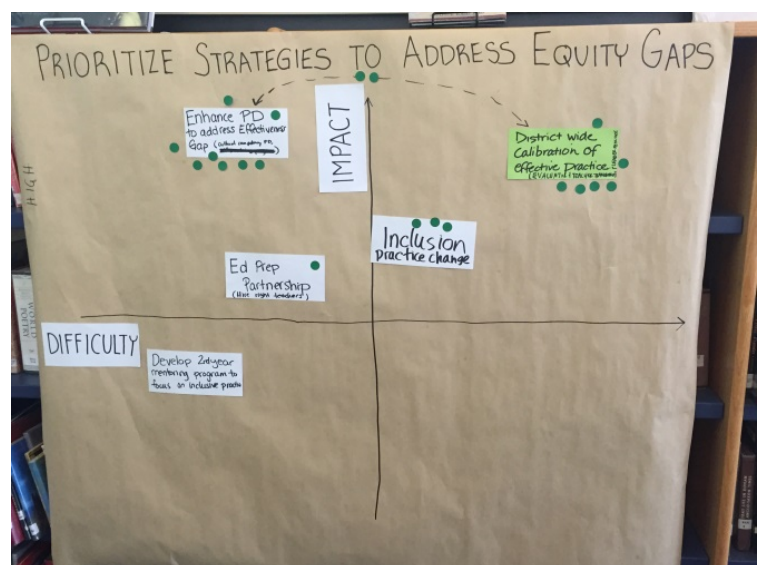
Data that informed Milton and Everett’s strategies included:

- ✓ **Achievement gap data from state and local assessments**
- ✓ **The Edwin Student Learning Experience (SLE) Report, which the PLN piloted**
- ✓ **[District Analysis Review Tools \(DARTs\)](#)**
- ✓ **Educator Evaluation ratings**

Everett examined Student Learning Experience Report data on different student subgroups’ rates of assignment to teachers with lower and higher evaluation ratings. They saw no clear trend in these assignment rates, but the data did fuel a conversation

Inclusive practice refers to the instructional and behavioral strategies that improve academic and social-emotional outcomes for all students, with and without disabilities, in general education settings. To support inclusive practice, the tools of the [Educator Effectiveness Guidebook for Inclusive Practice](#) are based on the frameworks of [Universal Design for Learning](#), [Positive Behavioral Interventions and Supports](#), and [Social and Emotional Learning](#).

Calibration is the result of ongoing, frequent collaboration of groups of educators to (1) come to a common, shared understanding of what practice looks like at different performance levels, and (2) establish and maintain consistency in aspects of the evaluation process, including: analyzing evidence; providing feedback; and using professional judgment to determine ratings.



PLN districts selected equity strategies based on data, root causes, and consideration of difficulty vs. impact.

about the district's evaluation ratings. Thus, Everett chose to work on evaluator calibration, and focused on best practices for inclusion due to persistent achievement gaps for Students with Disabilities (SWDs). Thus, Everett focused on calibrating their expectations of best practices for inclusion. The team selected the [Educator Evaluation Rubric](#) Standard II: Teaching All Students, Indicator A (Instruction), Meeting Diverse Needs as a focus indicator to guide this work.

Calibration around expectations of instruction was a also logical next step for Milton's evaluation implementation work. It also took into consideration teachers' concerns that different evaluators could rate the same practice differently. Evaluation ratings and [learning walks](#) have supported this concern in some cases. Dr. Karen Spaulding, a principal and the strategy lead in Milton, also noted the diverse backgrounds of evaluators and the range of teaching disciplines for which an evaluator might be responsible. She explained that "these varied backgrounds result in folks entering the conversation about instruction at very different places."

Choosing metrics

How do you measure closing the gap while you are trying to close the gap? We spent a lot of time discussing what data to collect before we would actually see changes in practice.

— Dr. Karen Spaulding, Milton Public Schools

In a December meeting, PLN members met to identify how to measure the success of their strategies, and plot trajectories for these metrics. Karen Spaulding from Milton said this was the most difficult step: "How do you measure calibration, as shifts in practice take time? How do you measure closing the gap while you are trying to close the gap? We spent a lot of time discussing what data to collect before we would actually see changes in practice." Milton selected as a final outcome metric the percentage of students in all subgroups assigned to teachers rated Proficient/Exemplary in Standards II and III.

Plotting trajectories

Everett set a trajectory for SWDs' SGP results, triangulating this data with the district's quarterly assessment benchmarks, as well as formative assessments the district uses in SPED. They found it helpful to compare their SWD assessment data with that of similar districts, identified in DART.

Districts plotted trajectories over multiple years. "We knew that to do the project with fidelity, it would have to be at least a one- or two-year process," said Pat Massa, Everett's curriculum analyst and strategy lead. "Change takes at least three years to see progress and it is more meaningful to analyze data to inform policy changes using trends over time." Milton agreed that it would take time to see outcomes, and also to include all stakeholders.

Bringing in stakeholders

Everett introduced their strategy at a monthly administrators' meeting, prior to the district's first evaluator training on inclusive practice. An early success of their implementation was gaining buy-in from the assistant superintendent, who is in charge of evaluation; his participation in activities reinforced their importance for evaluators. When introducing the strategy and focus on inclusive practices, principals shared with teachers [a tool from the Educator Effectiveness Guidebook for Inclusive Practice](#) to communicate what evaluators would expect to see teachers and students doing.



[“What to Look For - Observations”](#) and other tools in the Guidebook were developed and piloted by educators across Massachusetts. You can explore all the tools [online](#).

Both the Everett and Milton teams expressed the need to **shift teachers' views of educator evaluation** from one of anxiety to one of accepting support, or as Pat Massa, Everett's curriculum analyst, put it, “from a mentality of ‘they're out to get you’ to one of ‘they're out to help you.’” Fran Connolly, the Title I director in Everett, explained that “the more we talk with teachers, the more inroads we can make and frame [evaluation] as ‘we're in this together.’”

“When you bring yet another initiative in, you need to **give a good rationale** – how it's going to benefit the students. Sometimes people will be more accepting, any time you can bring it back to the students and **back it up with data**.”

– Pat Massa, Everett Public Schools



To gain evaluators' and teachers' support, Pat Massa advised being open about the reason for implementing the strategy: “When you bring yet another initiative in, you need to **give a good rationale** – how it's going to benefit the students. Sometimes people will be more accepting, any time you can bring it back to the students and **back it up with data**.”

Jodi Ronci, the Everett public schools' health coordinator and an evaluator, participated in a calibration activity for evaluators. She recommended that the strategy team provide more information for participants about the scope of the calibration work, and how they followed up on the calibration activity in admin meetings.

The Milton team communicated about their strategies both formally through administrator meetings and informally through updates with “key players,” who were comfortable in supporting teacher growth and represented a variety of subject areas and grade levels. **Recruiting “key players”** to participate in the work, such as planning PD, laid the foundation for gaining buy-in.

The bottom line: recommendations for gaining stakeholder support

- ✓ Be sensitive to time concerns and constraints
- ✓ Keep stakeholders informed
- ✓ Collect feedback through a streamlined process, rather than a series of lengthy meetings
- ✓ Provide PD that teachers and leaders need
- ✓ Strategically involve key roles; include all levels and roles (teachers, building leaders, central office)
- ✓ Learn about perspectives on calibration through collegial conversations

Implementation

The Milton and Everett teams **leveraged existing routines and structures** to implement the calibration of their vision for inclusive practice. As Karen Spaulding explained, “Initiatives are always competing for time and resources. You have to re-imagine your existing structures and make this work fit in.” Milton furthered their work through the district’s regular administrator meetings and PD sessions, such as PD on defining high-quality instruction.

Everett has used monthly administrator meetings to discuss and further calibration of expectations and quality feedback, as well as the assistant superintendent’s meetings that required all evaluators to attend. This approach also gained support from the superintendent, who had aimed to make the administrators’ meetings more collegial, and more focused on teaching and learning rather than district “housekeeping.” Meeting topics included improving teaching through conversations, and calibrating on SMART goals by bringing in example goals and providing feedback to one another. Participation increased noticeably in

Sample implementation timeline: Everett

January 2016:

→ SPED Director conducted a workshop for evaluators about the elements of inclusive classrooms, using the [Educator Effectiveness Guidebook for Inclusive Practice](#), what practices to look for in classrooms, and how to give effective feedback

→ Principals rolled out the strategy to their staff, to explain what practices evaluators would be focusing on; used a [Guidebook “What to look for” tool](#) to communicate what evaluators would expect to see teachers and students doing

January and mid-February:

→ Observations conducted, including feedback on inclusive practice

February through end of year:

→ Administrator team began professional reading (see “Tools & Resources” below) focused on evaluation, such as an article on quality feedback, which helped develop similar thinking about components of evaluation

→ Administrator team held monthly training on evaluation protocols with assistant superintendent

March and April:

→ Observations conducted, including feedback on inclusive practice

June:

→ Gained baseline data by surveying teachers’ perspectives of evaluation and feedback, adapted from the [Educator Evaluation Implementation Survey](#)

Summer:

→ PARCC & MCAS data analyzed to inform strategy

these meetings.

accomplish, such as setting the action plan. Meanwhile, the team implementing the calibration strategy met in person when there was specific work to accomplish, such as setting the action plan or deciding on metrics, but otherwise much of the communication was done via email to avoid becoming bogged down in scheduling obstacles.

Pat Massa recommended that “the team leader can help to alleviate some stress with the time factor by doing things in advance and then asking for input and edits from the rest of the team.” For example, for a survey on teacher perspectives (see “Tracking Progress” below), she made a draft and shared it with the team for feedback, saving time but still

building joint ownership of the work.

Group practice in developing and calibrating feedback

drove home the importance of calibration in Everett and Milton. Both districts used the pilot [Online Calibration Tool](#), with lessons from actual Massachusetts classrooms found in ESE’s [classroom instruction video library](#). In this activity, evaluators watch the same lesson, independently rate



Lessons such as this one from [the classroom instruction video library](#) can spark discussion about what teaching looks like at different levels of effectiveness.



When using the Online Calibration Tool, you can be **strategic**, selecting a video tagged to elements of particular importance to your school, or selecting a video based on content area or grade level.

the teacher, and write the feedback they would provide. The tool then shows the group the distribution of participants’ ratings, and how these compare to ratings from other users of the tool statewide. Evaluators then discuss both ratings and feedback to gain more clarity and concord about what different levels of instruction look like. Because Everett was working on calibration around inclusive practices, they focused the activity on Standard II: Teaching All Students, and used [videos tagged to that standard](#).

Jodi Ronci attended Everett’s calibration video session as an evaluator and found it eye-opening, particularly as colleagues gave the same teacher a spectrum of ratings. She said it would be helpful to continue such calibration activities, noting that it didn’t solve the entire calibration issue, but did start an important conversation.

Both Everett’s and Milton’s work included a **focus on calibrating evaluator**

feedback. For a meeting on professional feedback, Everett had success using a meeting structure that encouraged full participation in a comfortable, nonjudgmental way.

In a meeting of the Milton and Everett teams, Milton Math Curriculum Coordinator Michelle Kreuzer noted the need to calibrate around feedback: “Is it actionable? It’s helpful to talk it out so you can ensure that the feedback leads to something teachers can actually do.”

Milton worked to target feedback on priority areas through weekly scheduled meetings



Evaluators in Everett viewed this [poetry lesson from the instruction video library](#) as part of a calibration activity.

with regular discussion about evaluation. Evaluators came to a common understanding of high-quality feedback by surfacing existing ideas and comparing it to research literature.

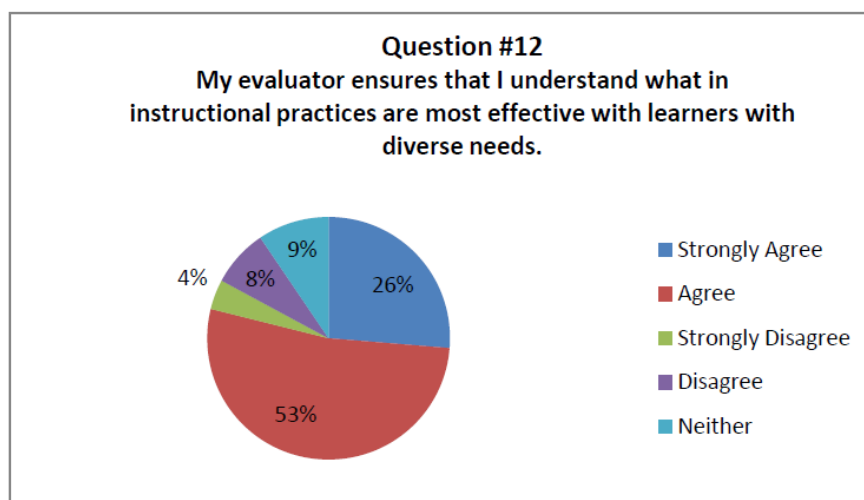
Using the teacher evaluation software TeachPoint was also helpful because it allowed evaluators to read one another’s feedback for the same evaluated educator and “bring impressions together

The bottom line: overcoming implementation obstacles	
Time constraints	<ul style="list-style-type: none"> ✓ Strategy lead develops work and then shares it with others for feedback ✓ Communicate during existing meeting times and via email
Competing priorities & leadership buy-in	<ul style="list-style-type: none"> ✓ Build off of existing structures and priorities ✓ Integrate into overall district evaluation work
Evaluators’ and teachers’ “initiative fatigue”	<ul style="list-style-type: none"> ✓ Openly explain the process from the beginning ✓ Back it up with student performance data ✓ Strategy lead takes on more work so others don’t have to ✓ Train evaluators on calibrated, effective feedback → teachers can view evaluation as a support, not a concern
Finding appropriate metrics	<ul style="list-style-type: none"> ✓ Return to the initial metrics that revealed an equity gap ✓ Establish new metrics for a baseline measure (i.e. survey of teacher perspectives) ✓ Consider outcomes at the evaluator, educator and student levels

Tracking progress and shifts in perspective

PLN districts determined how they would measure strategies' success over time. Everett used a survey for feedback on their initial evaluator training, which yielded positive responses. The district knew they wanted to track changes in teacher attitudes toward evaluation (particularly of inclusive practices) over time. To gather baseline data, Pat Massa created a teacher survey on views of evaluation and feedback, adapted from the [Educator Evaluation Implementation Survey](#). It yielded a 60 percent response rate. The Everett team will analyze PARCC and end-of-year assessment data, specifically for students with disabilities. This and the survey data will inform the direction of calibration work, and additional PD for teachers and evaluators.

In Milton, Karen Spaulding observed that calibration work affected educators differently depending on how they had previously viewed evaluation: "Those who were already committed to using the tool to improve practice of their teachers (as opposed to compliance) were able to dig more deeply. I think those who were in compliance mode shifted



Everett surveyed teachers to collect baseline data on perspectives of educator evaluation.

their thinking a bit about the intent and power of the tool – i.e. to improve practice and ultimately learning for all students." Participation also increased understanding of the rubric.

Other next steps

Everett will continue to work on quality feedback. In 2016-17, evaluators will conduct evaluations through the lens of inclusive practice, attend PD sessions, and continue to meet regularly to review evaluations. Admin meetings will maintain a focus on educational readings and collegial discussions that promote the effective implementation of evaluation to improve student achievement. Some school leaders will be conducting the video calibration activity with teachers during PD, to develop a shared idea of best practices. Everett may choose to apply a new "lens" to their calibration work in the future, but at least for the next year will focus on inclusion.

As part of the multi-year strategy, Milton will continue work on calibration through administrator PD and increasing shared learning walks. Karen Spaulding explained that the continued work "provides a framework for ongoing conversations about supporting teachers to improve their practice... It answers the question, 'Where do we go from here?'"

Tools & resources

Suggested for calibration & inclusion work



- ✓ The 2015 [MA Plan for Equitable Access to Excellent Educators](#)
- ✓ [Educator Evaluation Rubric](#) – Everett focused on Standard II: Teaching All Students, Indicator A (Instruction), Meeting Diverse Needs
- ✓ Educator Evaluation Implementation Surveys for [teachers](#) and [administrators](#), to measure educators' perspectives on evaluation
- ✓ ESE's Educator Evaluation [calibration](#) page includes:
 - A library of [classroom instruction videos](#)
 - [Sample calibration protocols and activities](#)
 - [Online calibration training tool](#)
 - [Research and case studies](#) about educator evaluation in MA
- ✓ Referenced/recommended tools from the [Educator Effectiveness Guidebook for Inclusive Practice](#):
 - 2A: [MA Classroom Teacher Rubric Resource](#)
 - [Power Elements for Inclusive Practice - MA Model System Classroom Teacher Rubric](#)
 - 3C: [Case Study Goals for Discussion - Accessible Instruction](#)
 - 3D: [Case Study Goals for Discussion - Positive Behavior Supports](#)
 - 4A: [WHAT TO LOOK FOR: Observations](#)
 - 4B: [WHAT TO LOOK FOR: Observation Feedback](#)
- ✓ Articles the Everett team used for professional reading and discussion:
 - Tomlinson, C.A., "[Pondering Good vs. Great.](#)" *Educational Leadership*: 72 (5).
 - Hattie, J. "High-Impact Leadership." *Educational Leadership*: 72 (5).
 - Arneson, S. "Improving Teaching One Conversation at a Time." *Educational Leadership*: 72 (7).
 - Novick, B., "10 Tips for Tackling Tough Conversations." *Educational Leadership*: 72 (7).
 - Brookhart, S. & Moss, C. "How to Give Professional Feedback." *Educational Leadership*: 72 (7).