

This report is the review of the educator evaluation system as implemented by Michigan's K-12 educators. The report was prepared for the Michigan Department of Education and reviews best practices, barriers to implementation and use of VAM and alternative assessments in educator evaluation.

# Educator Evaluation Research and Evaluation Activities Final Report – by Ray.Taylor and Associates, LLC

September 30, 2017

---

September 30, 2017

September 30, 2017

This evaluation review was conducted by the evaluation team of Ray.Taylor and Associates, LLC and was funded by the Michigan Department of Education through Gogebic-Ontonagon Intermediate School District. The findings and opinions expressed are those of the project team at Ray.Taylor and Associates, LLC.

The Ray.Taylor and Associates evaluation team:

Rossi Ray-Taylor, PhD	Paula Allen-Meares, PhD	Mathis Taylor
John Norwood, PhD	Ira Washington, PhD	Amber Tselios
Patrick (Rick) O'Neill	Jessica Thomas	Marshall Taylor



September 30, 2017

September 30, 2017

# EDUCATOR EVALUATION RESEARCH AND EVALUATION ACTIVITIES FINAL REPORT

---

## Executive summary

The Michigan legislature adopted Public Act (PA) 173 in 2015. The legislation governs K-12 educator evaluation and is similar to legislation adopted in other states in response to No Child Left Behind (NCLB) and Race to the Top (RTTT) guidelines from the U.S. Department of Education. PA 173 builds upon and clarifies PA 102 of 2011. The Michigan Department of Education (MDE) issued a request for proposal to review the implementation of the educator evaluation system in Michigan schools and school districts. This comprehensive review was conducted by Ray.Taylor and Associates, LLC and is designed to address the three questions posed by MDE.

**1: Identify best practices in educator evaluation implementation to inform ongoing and future MDE supports to, and programming by, districts.**

**2: Identify barriers to implementation in local educator evaluation systems and make recommendations for ways to mitigate barriers and inform ongoing and future supports to, and programming by, districts.**

**3: Evaluate the measurement of student growth using alternative methods/tools and processes specified in PA 173 of 2015.**

The report provides information and recommendations that can be used by MDE and Michigan K-12 educators as they implement educator evaluation.

The project methodology combines quantitative and qualitative methods from multiple data sources to provide information from educators evaluated by the system and those responsible for conducting evaluation. The review implemented by Ray.Taylor and Associates includes the following data collection sources:

- a) Literature review
- b) Document review
- c) Researcher white papers
- d) Practitioner white papers
- e) Focus groups

September 30, 2017

- f) Statewide survey of K-12 educators
- g) Interviews
- h) Identification of schools and districts implementing best practices

Information collected from these data sources is triangulated to provide deep understanding of the implementation of the educator evaluation system in Michigan and to identify opportunities for improvement.

Throughout this review we note the dual purposes that educator evaluations are asked to serve: to provide data needed for human resources decisions (usually noted in rankings) and to provide information and feedback to improve educator practice. Similar to this are the two functions that the educator evaluation review can serve: one purpose would be to judge the effectiveness of implementation of the educator evaluation, the second purpose to provide information that can be used by MDE, policy makers, K-12 educators and their supporters to improve the system. The focus of this review is this second purpose.

*“Evaluation is to help projects become even better than they planned to be...First and foremost evaluation should support the project...”*

*W.K. Kellogg Foundation  
Evaluation Approach, 1997*

## Summary – What We have Learned

*“Overall the evaluation system from the state has made improvements to teaching and learning with the rubrics of what good teaching should look like, and sound like, and feel like. But the punitive nature of student growth and other parts have set back the growth of teachers – where would we be if not for this setback?” – Focus group participant*

### What is working

An important charge to this evaluation initiative is to identify what should change or improve to support the educator evaluation system. But before launching into change one must be clear about what currently works and why. The Michigan Council for Educator Effectiveness (MCEE) report and pilots have given direction to the educator evaluation system that is present today. When we look back at the MCEE principles and recommendations we see evidence of their impact on the system. We learned through focus group participants, survey respondents, practitioner white papers and interviews that educators are astutely aware of the requirements to evaluate, and are making efforts to meet requirements within the constraints of their work demands. MDE provides a number of support resources through its website, e-newsletters, webinar, surveys, and meetings with school leaders throughout the state, along with other strategies to both share information and to listen.

### Educator Evaluation – Impact on the System of Education

Educator evaluation is best understood when contextualized within systems of school level, district, region, state and national stakeholders that define clientele, taxpayers, policymakers, administrators,

September 30, 2017

students and parents. Districts are often simultaneously addressing student achievement, parent outreach, competition for students, fiscal distress and budget challenges, a winnowing workforce, and high turnover in students and staff. In the environment of K-12 education, therefore, the evaluation system must be situated within the real world facing schools.

This review has assembled information to address the three questions posed by the MDE in its request for proposal regarding – 1) best practices, 2) barriers, and 3) value added measures (VAM) and alternate assessments. We find in our data strong overlap in the three questions and their impact on the implementation of the educator evaluation models on the system of education in schools and districts.

Several major points surface in this review of the educator evaluation system as designed and implemented in Michigan.

1) Although at different stages in implementation, overall educators in Michigan are in early stages of implementation of the evaluation system as described in PA 173. As with any systemic reform effort, educators are adapting the models and practices to fit their local circumstances. Much of the attention at this point seems to be compliance focused.

2) There is tension between the dual roles of the educator evaluation system to serve the high-stakes human resource function of rating, screening and documentation of workforce actions and the role of the evaluation system to provide feedback for professional growth and improvement.

3) Overall educators report little useful feedback to improve their professional practice and support their professional growth. They describe time and lack of skills in evaluation among barriers to implementation. These, too, may be barriers to providing useful feedback – creating a negative feedback loop.

4) Similarly, educators cite loose connections between evaluation models used as a reflection of professional standards and expectations, student learning outcomes, and information to increase student learning and educators' professional growth. This disconnect creates a cycle that drags down the utility of the educator evaluation and can result in compliance rather than substantive implementation –continuing the cycle.

5) Teachers report not being consistently engaged in goal setting with their evaluator or colleagues, and not consistently adopting improvement strategies recommended by their evaluator – factors that may be related.

6) There are technical questions surrounding the efficacy of the system to serve human resource functions, most notably the psychometric and technical issues related to student growth measures and the fidelity with which the evaluation system is implemented. To the extent that the educator evaluation system is used for high-stakes decision making the process would benefit from modeling to determine the reliability and validity of the VAM and alternate assessment system as implemented in Michigan schools.

September 30, 2017

8) Educators seem to be struggling to balance issues of inter-rater reliability and fidelity of implementation with practical issues of limited time and training.

9) There was little to no evidence of a link between the educator evaluation results and district, building or individual professional development plans or school improvement plans. If the educator evaluation system holds any value whatsoever this is an opportunity missed.

### Recommendations and next steps

*“Every system is perfectly designed to get the results it gets.” W. Edwards Deming*

We noted in this report the intent to provide information useful to MDE, policy makers, K-12 educators and their supporters to improve educator evaluation in Michigan. With that purpose in mind we offer the recommendations below. We have grouped the recommendations by the entity with primary responsibility for implementation.

### Recommendations for MDE

We begin by making recommendations directed toward MDE officials.

- I. **Stabilize M-Step, gather multiple year data with one version of the test** – throughout our review we heard concern raised regarding stability of M-Step and its resultant appropriateness for high-stakes decisions. Two factors are pertinent. First, reliable decision making takes multiple years of trend data. Second, school factors are adapted to revisions of the state assessment and the results produced. However, it takes time and resources to adopt revisions in local curriculum, instructional practices, instructional materials, teacher training and formative assessments to serve changes in the statewide test. For schools, teachers and students to experience the full value of the educator evaluation system the measures of student growth, and the state assessment specifically must be stable for multiple years.
- II. **Communicate standards for instruction** – Among best practices is standards that are well known and supported. One type of standards that impacts teaching and learning is the standards for instruction. These standards are communicated through the evaluation model selected by the local district and appear to have wide variation in rigor and fidelity in implementation across the state and within districts. Among best practices in Michigan we see evidence of school districts and ISD/RESA adopting professional learning communities to build common understanding of standards of instruction. MDE can advocate for these communities and support their work through professional development support.
- III. **Add to/continue to provide face-to-face opportunities for educators to give input to MDE policy and practice, especially teachers and principals in partnership** - participants in the focus groups thanked MDE for valuing their input and for staging the conversations among teachers and administrators that the focus groups provided. They indicated willingness to participate in future focus groups and asked that MDE continue this strategy for learning from those directly involved on the frontline.
- IV. **Enhance the mindset and focus of schools toward improvement of practice over ranking** – the work of instructional improvement strengthened through collaboration is being undermined by the



September 30, 2017

competitive culture of ranking educators. Research has shown that the vast majority of educators are capable of improvements in practice when focused feedback and support are given. Instructional improvement is a learning and organizational issue not a matter of sanctions. MDE can be a catalyst to a mindset that supports improvement of professional practice throughout the professional life of educators.

- V. **Along with ISD and RESA convene and support statewide professional learning communities (communities of practice) for those leading improvement of instruction and professional development and use of student learning objectives (SLO)** (see Dearborn model) – Educators from across the state could benefit from a convening of a statewide professional learning community to research and explore strategies for implementing educator evaluation within a unified system of improvement of teaching and learning.
- VI. **Assist districts and ISDs in creating access to training video of best instructional practices “What effective instruction looks like”** – MDE can work with local educators to develop and distribute video and online depictions of what effective instruction and rigorous standards look like. MDE can help assuage the notion *that students like the ones in my classroom cannot perform in those ways*
- VII. **Develop a calendar of required reports and document preparation for building principals at the end of the school year and work with district leaders to consolidate reporting and document preparation.** The end of a school year is an intense time that is flooded with end of year reports, next year grant applications, transparency reporting, testing, annual plan development, staffing, budget planning, and year end activities, both formal and informal. Each activity and report requires principal and teacher involvement. MDE and local district leaders can look for ways to consolidate, reschedule, or omit these requirements so educators can focus on teaching and learning.
- VIII. **Encourage and support learning strategies and resource sharing with other states and nations successfully tackling educator evaluation.** Literature shows evidence of innovative strategies across the nation. Michigan may be able to learn from what works and what has not in other states and nations.
- IX. **Identify factors that have resulted in state and national teacher shortages and identify means to improve the numbers and skills in the workforce** – with economists, researchers, university education schools and local districts, identify the factors that lead to the decline in workforce in K-12 education in Michigan and nationwide, and identify strategies for workforce development. Responsibility for workforce development is shared with local educators and MDE.
- X. **Consider leveraging Michigan’s voice to advocate for improvements in the state-approved educator evaluation models** – Establish an educator taskforce made up of teachers, principals /assistant principals and human resource professionals to develop feedback and recommendations for the developers of the approved evaluation models. Invite the models developers to Michigan to receive the recommendations and share their response.

#### *Recommendations for district implementation (school and district levels or ISD /RESA)*

- I. **Train evaluators and those evaluated** – to implement evaluation models, observe, record and provide effective feedback, understand the meaning and use of student growth measures to

September 30, 2017

improve teaching and learning, and to schedule and manage time to conduct all required observations

- II. **Convene and facilitate professional learning communities of member districts** – for an example see Newago ISD, Dearborn and Wayne RESA models
- III. **Evaluate the quality of feedback** – with emphasis on improvement /professional growth
- IV. **Strengthen the ties of evaluation to PD** – comprehensive professional development plan /model that is tied to individual needs and professional growth model (full system)
- V. **Assist local districts by convening work sessions to develop SLO** - allowing districts to share the burden and cost of the work and technical support
- VI. **Provide access to high quality national and state trainers to enhance training available at local levels** – smaller and resource strapped districts often do not have access to high quality trainers, by providing access to high quality trainers on a regional basis MDE and ISDs can enhance the quality of training provided to educators
- VII. **Enhance the quality of calibration /inter-rater reliability** – provide training and technical assistance to improve inter-rater reliability and calibration.
- VIII. **Enhance training regarding understanding student outcomes measures and results** – educators need deeper understanding of student outcome measures and results to make more informed decisions about their links to instructional planning, school improvement, and educator training and appraisal.
- IX. **Continue to emphasize and build educator awareness /knowledge of what quality teaching and learning look like** – without boundaries of class, income, culture, race, ethnicity, -- not reserved for the privileged few
- X. **Address matters of cultural relevance and competency regarding curriculum, instructional methods and observations** – provide leadership with educators throughout the state to develop understanding of culturally responsive teaching.

#### *Recommendations regarding policy and governance issues (state and national)*

- I. **Get the balance between improvement and sanctions right** - untangle the two functions educators are being asked to serve
- II. **Identify factors that have resulted in national teacher shortages and identify means to improve the numbers and skills in the workforce** – responsibility for workforce development is shared with local educators and MDE, but is also impacted by state and national policy and funding
- III. **Identify and address underfunding and resource disparities among schools and districts**
- IV. **Support MDE efforts to stabilize statewide assessment** – educators need a well-known and understood assessment system that remains in place long enough to provide multiple year trend data allowing educators to adapt and improve their practice based on results. The assessment system must meet psychometric standards sufficient for the high-stakes decisions it impacts.

September 30, 2017

## Contents

Executive summary .....	4
Summary – What We have Learned .....	5
What is working .....	5
Educator Evaluation – Impact on the System of Education .....	5
Recommendations and next steps .....	7
Acknowledgements and Thank you .....	12
Introduction .....	13
The charge – the evaluation question – What MDE Wants to Learn .....	13
Context – Legislation From Race to the Top, to ESSA and MCEE report .....	15
Our methodology .....	16
Data Sources .....	17
Literature Review – Summary paper .....	17
A Summary of Documents Reviewed .....	19
Researcher White Papers – methods and outcome summary .....	19
Voices from the Field: Practitioner White Papers – methods and outcome summary .....	23
Focus groups – methods and outcome summary .....	25
Focus Group Snapshot - Who Participated .....	27
Evaluation Models Used by Participants’ Schools .....	27
Statewide Educator Survey– methods and outcome summary .....	28
Interviews .....	31
Identification of Best Practices in Action .....	31
Evaluation Findings and Answers to the Three Questions Asked by MDE .....	33
Activity Question 1 – Best Practices .....	33
The Process – How Educator Evaluations are Conducted .....	36
Classroom Observation Models and Tools .....	38
Training .....	40
Type and frequency of classroom observations visits .....	77
Activity Question 2 – Barriers .....	119
Activity Question 3 – Measures of Student Growth using alternate methods/ tools and processes .....	124
Guiding Principles for Educator Evaluation System .....	134

September 30, 2017

Rating Systems .....	143
Time .....	148
Trust .....	148
Quality of Feedback .....	149
Links to professional development and professional growth.....	149
Using what we know about how adults learn, change and improve their practice .....	150
Summary – What We have Learned .....	151
What is working .....	151
Educator Evaluation – Impact on the System of Education.....	152
Additional questions – what haven’t we learned .....	153
Recommendations and next steps.....	155
Next steps .....	158

September 30, 2017

## Acknowledgements and Thank you

The Ray.Taylor and Associates team thanks the thousands of Michigan educators who took the time and to offer their thoughts and input through participation in focus groups, surveys, interviews, document review and white papers. The voices of these educators are essential to understanding the implementation of the educator evaluation system in Michigan. Their candid words, thoughts, experiences and recommendations are the foundation of this report.

The Ray.Taylor and Associates team is grateful to the following for their thoughtful insight and professional knowledge:

Research white paper authors Robert Floden, PhD and Suzanne Wilson, PhD

Practitioner white paper authors Crystal Wise, Ann Blais and Rosiland Brathwaite

Reviewers of draft survey Gary Court, Ellen Daniel, Lametria Eaddy, Monta Stevenson, Crystal Wise, Ann Blais, Rosiland Brathwaite, and Denise Eaddy-Richardson

Interview participants Robin Jacob, PhD, Brian Rowan, PhD, Gary Court, and Stephanie Burrage

### *Focus group participants:*

Scott Smith	Amy Rogers	Jadyn Bescnen
Mike Wilson	Tori Wentela	Lori Roy
Trent Bellinger	Ashley Stadel	Bill Turner
Luke Siebert	Julie Block	Tom Martin
Alicia Thompson	Nicole Gasper	Jessica Garcia
Dave Beauchamp	Jason McVoy	Shanna Beal
Scott Moore	Amanda Brescia	Tricia Tervo
Lisa McFarland	Dean McGuire	Brian Reattoir
Bruce Cook	Chery Taylor	Chris Marana
Kris Vydareny	Carrie Smith	Nikki Lishinski
Kyle Talicska	Robert Smith	Karie Belanger
Beth Marcus	Elizabeth Walker	Alan Tulppo
Jim Ply	David Treder	Joyel Hyvarinen
Terry Breen	Mark Pogliano	Tom Wagner
Jim Jensen	Stan Masters	Kevin Hooper
Adam Bayne	Mike Hubert	Deb Asano

---

We also thank the staff at the Michigan Department of Education – Joshua Roesner, department analyst for the Office of Educator Talent, and Jared Robinson, with the leadership and support of Abbie Groff-Blaszak, Director of the Office of Educator Talent. Gogebic-Ontonagon Intermediate School District served as fiscal agent for this project under the leadership of Mark Rainey and Debbie Vargas, grant officers Ashley Nevins and Jared Stenson, and Bruce Mayle, GOISD superintendent – we owe all our utmost thanks.

September 30, 2017

# EDUCATOR EVALUATION RESEARCH AND EVALUATION ACTIVITIES FINAL REPORT

## Introduction

In 2015 the Michigan legislature adopted Public Act 173 which governs educator evaluations. The legislation follows the practice of most other states that adopted educator evaluation legislation in response to No Child Left Behind (NCLB) and Race To The Top (RTTT) guidelines from the US Department of Education. Michigan’s 2015 legislation clarifies and builds upon PA 102 of 2011.<sup>1</sup> The evaluation process for Michigan is linked to the Michigan Department of Education’s Strategic Goal #3, to “develop, support, and sustain a high-quality, prepared, and collaborative education workforce.”

In January 2017 the Michigan Department of Education (MDE) awarded three contracts to evaluate the implementation of the Educator Evaluation System. Ray.Taylor and Associates was awarded one of the contracts. This is the final evaluation report by Ray.Taylor and Associates.

## The charge – the evaluation question – What MDE Wants to Learn<sup>2</sup>

In the request for proposal three activity areas were outlined by MDE. These were accompanied by a series of related questions.

Figure 1: Questions Posed by MDE

**Activity 1: Identify best practices in educator evaluation implementation to inform ongoing and future MDE supports to, and programming by, districts.**

- What are evidence-based best practices for:
  - provision of professional development and mentoring for teachers and administrators which is aligned with their individual educator evaluation areas/result
  - integration of cultural competency into evaluation models and professional development for teachers and administrators?
  - provision of quality feedback to teachers and administrators throughout the school year as part of the educator evaluation process?
  - training of educators (teachers and administrators) on educator evaluation systems and the multiple components within the educator evaluation system and tools?
  - administrator evaluation in general, and specifically for school-based

<sup>1</sup> Link to PA 173 of 2015 <https://www.legislature.mi.gov/documents/2015-2016/publicact/pdf/2015-PA-0173.pdf>

<sup>2</sup> From MDE Request for Proposal, November 7, 2017

September 30, 2017

administrators as compared to district-level administrators? using multiple measures of student growth in educator (teacher and administrator) evaluation, including the aggregation of multiple measures of growth and the combination of aggregated growth measures with the professional practice component to produce an overall effectiveness rating?

- How has implementation of these best practices impacted learning outcomes for educators and students?

**Activity 2: Identify barriers to implementation in local educator evaluation systems and make recommendations for ways to mitigate barriers and inform ongoing and future supports to, and programming by, districts.**

- What is the impact of educator evaluation systems and practices on:
  - innovative, personalized instruction and learning?
  - the role of the administrator?
  - student learning and outcomes?
- What are the most common barriers in Michigan districts to implementing high quality educator evaluation systems?
- What are recommendations for ways these barriers can be mitigated or eliminated by the district and/or supports provided by MDE?
- How does implementation of quality educator evaluation systems differ (if at all) in schools and districts that educate large numbers of historically underserved student populations?
- What barriers exist for equitable implementation of the educator evaluation system within the school and/or district?

**Activity 3: Evaluate the measurement of student growth using alternative methods/tools and processes specified in PA 173 of 2015.**

- What are evidence-based best practices/processes for measurement and implementation of student growth for educator evaluation using:
  - student learning objectives (SLOs)?
  - achievement of individualized program goals?
  - nationally normed or locally developed assessments that are aligned to state standards?
  - alternative assessments that are rigorous and comparable across schools within the school district, intermediate school district, or public school academy?
- What are critical components in the measurement and implementation of alternative methods/measures/tools for student growth?
  - Are there different considerations for alternative growth methods/measures/tools for student growth based on content area and/or grade levels?

**For each question, which Michigan districts are implementing these best practices/processes**

September 30, 2017

The evaluation design by Ray.Taylor and Associates and final report address each of the questions raised by MDE in its request for proposal. It is the intent of this review to also provide actionable steps that can be taken by Michigan Department of Education officials and Michigan K-12 educators.

This report will describe the evaluation questions pursued, the methodology, the data gathering tools, findings, recommendations and next steps. Attached to this report is an appendix that includes documents prepared for this report including the full literature review and document review reports, survey instrument and results, white papers commissioned for the evaluation, and full descriptions of best practices found in districts throughout the state.

### **Context – Legislation From Race to the Top, to ESSA and MCEE report<sup>3</sup>**

Michigan’s legislation regarding educator evaluation is a direct outgrowth of a series of school reform efforts that took place in the early 2000’s nationwide. Sparked by Race to the Top (RTTT) and No Child Left Behind (NCLB) reforms, states were incentivized to adopt accountability measures that included evidence-based educator evaluation systems among other components. Prior to the legislation individual school districts nationwide had been adopting reforms in teacher observation and feedback as seen in the adoption of the Charlotte Danielson’s Framework for Teaching and Marzano Teacher Evaluation Model among others. Districts and schools that received competitive federal School Improvement Grant (SIG) funds beginning in 2010 were required to adopt these rigorous evaluation models, as were schools in Michigan identified as priority schools. These reforms were described as accountability measures. As described in the 2009 report “The Widgets Effect”<sup>4</sup> accountability was increasingly defined as a system to enhance quality assurance in the profession through weeding out poor performing educators in general, and teachers more specifically<sup>5</sup>. Taking a backseat was the intent to provide feedback to guide the professional growth of the educator. This juxtaposition between professional support and growth, and human resource transactions (e.g., screening, selection, promotion and reduction in force), was alluded to in the researcher and practitioner white papers, seen in the discourse of focus group participants, and found in the literature review. The quest for teacher quality as a linchpin of school improvement is seen throughout the world as schools struggle to reform<sup>6</sup>. The tension between the purpose of professional evaluation for professional growth versus evaluation

---

<sup>3</sup> The Michigan Council for Educator Effectiveness (MCEE) was established in June 2011 as part of Michigan’s teacher tenure reform efforts (PA 102 of 2011). Council members were appointed in September 2011, and the legislature appropriated funding in mid-December 2011. The MCEE was a temporary commission with a life of no more than two years and officially disbanded in June 2013.

The council had five voting members, three of whom were appointed by Governor Rick Snyder, and one each by Senate Majority Leader Randy Richardville and Speaker of the House Jase Bolger. Governor Snyder appointed Deborah Loewenberg Ball, dean of the University of Michigan School of Education, as chair of the MCEE. In addition to Ball, the governor appointed Mark Reckase from Michigan State University’s College of Education and Nick Sheltrown from National Heritage Academies in Grand Rapids. Majority Leader Richardville appointed David Vensel, a principal from Jefferson High School in Monroe, and Speaker Bolger appointed Jennifer Hammond, a principal from Grand Blanc High School. Joseph Martineau served on the MCEE without a vote and was the designee of the Michigan Department of Education’s superintendent of public instruction. (See Appendix A for biographies of council members.) p.4

<sup>4</sup> “The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness” 2009, The New Teacher Project, Daniel Weisberg, et al

<sup>5</sup> Widgets Effect op cit; and Motoko Akiba Educational Researcher, May 2017 article

<sup>6</sup> Motoko Akiba, Educational Researcher, op cit, James Hiebert and James W. Stigler, Education Researcher May, 2017; Policy Brief: Teacher Appraisal and Feedback, National Center on Education and the Economy (NCEE) 2016; Country Brief: Canada: Diversity and Decentralization, National Center on Education and the Economy 2016]



September 30, 2017

for human resource decisions like reduction in force, merit pay, termination and promotion is similar to the discord seen in other professions and in the business community<sup>7</sup>.

*Education Researcher* produced a featured series of articles contrasting educator evaluation systems in high performing nations. We find in the *Education Researcher* review and a policy brief by the National Center on Education and the Economy (NCEE) that compares educator appraisal practices in high performing nations a marked emphasis on educational evaluation practices aimed toward improving teachers' instructional practices.

To prepare Michigan for the adoption of enhancements to the educator evaluation system the Michigan Council for Educator Effectiveness (MCEE) was established in 2011 with five members appointed by Governor Rick Snyder. In 2012- 2013 a Pilot of Educator Effectiveness Tools was conducted by the MCEE. A report regarding pilot implementation was prepared for MCEE by the U of M Institute for Social Research (ISR)<sup>8</sup>. The MCEE taskforce produced a series of recommendations in 2013. As part of the evaluation the Ray.Taylor and Associates evaluation team has revisited the MCEE principles to assess their status in 2017.

## Our methodology

This evaluation was designed by Ray.Taylor and Associates to address the questions posed by MDE shown above and to determine factors for MDE and educator improvements in the process of educator evaluation. Our methodology combines quantitative and qualitative methods, and borrows from social innovation methodology (as described in Social Innovation Review of Stanford University) and *developmental evaluation*<sup>9</sup>. The evaluation design was crafted to compile information from the educators

*“When state legislatures drafted legislation to amend the evaluation systems under which nearly every teacher earned the same satisfactory rating, legislators and advocacy groups primarily focused of introducing additional measures beyond a few classroom observations, increasing the frequency of evaluations, and, mainly, including measures of student learning in ways that were objective and fair to all teachers.”*

**From, *Running in Place: How New Teacher Evaluations Fail to Live Up to Promises*; page 3, National Council on Teacher Quality, January 2017**

<sup>7</sup>The Performance Management Revolution, Peter Cappelli and Anna Tavis, Harvard Business Review, October 2016; and The Push Against Performance Reviews, Vauhini Vara, The New Yorker, July 24, 2015

<sup>8</sup> Promoting High Quality Teacher Evaluations in Michigan: Lessons from a Pilot of Educator Effectiveness Tools, Brian Rowan et al , December 2013

<sup>9</sup> “Developmental evaluation (DE) is grounded in systems thinking and supports innovation by collecting and analyzing real-time data in ways that lead to informed and ongoing decision making as part of the design, development, and implementation process.” As such, DE is particularly well suited for innovations in which the path to success is not clear. By focusing on understanding what’s happening as a new approach is implemented, DE can help answer questions such as:

- What is emerging as the innovation takes shape?
- What do initial results reveal about expected progress?
- What variations in effects are we seeing?
- How have different values, perspectives, and relationships influenced the innovation and its outcomes?
- How is the larger system or environment responding to the innovation?”

From The Case for Developmental Evaluation, March 1, 2016. FSG

September 30, 2017

evaluated by the system and those responsible for conducting evaluation. The evaluation design was also built to allow triangulation of data from multiple sources to explore themes, patterns, and incongruences. Data was compiled from literature and resource review, project commissioned white papers from researchers and practitioners, focus groups, a statewide survey, analysis of public data sources and targeted interviews. The design also identified schools and districts that implement best practices for educator evaluation. The design was built to yield a rich source of information that reflects both the educator evaluator and those evaluated, gathered from educators across the state. Relationships across the data sources are identified and explored. The review was conducted by a team of evaluators from Ray.Taylor and Associates. The evaluation team reached out to others to test the survey, and to explore background information useful to the evaluation. [See the appendix for the full list and description of Ray.Taylor and Associates' team of evaluators and support.] Each section of this report includes a description of related methodology. Although the evaluation team spent hundreds of hours analyzing the data compiled for this review, much more insight remains to be mined from the data with further analysis.

## Data Sources

We turn next to descriptions of the sources of data assembled for this evaluation report. This section will provide an overview of the data source, how collected and synopsis of findings. The multiple data sources of our evaluation include a literature review, document review, researcher and practitioner white papers, focus groups, statewide survey, interviews and identification of best practices. To achieve triangulation the overall findings from all data sources are interwoven in each section of this report to address the full set of evaluation questions.

## Literature Review – Summary paper

We began our work with a comprehensive review of the literature surrounding educator evaluation<sup>10</sup>.

The cannon of literature included peer reviewed journals, books and reports regarding professional evaluation is rapidly expanding – with more published each week.

Our review draws on an in-depth review of literature on the topic to summarize best practices in teacher evaluation. It is organized under the three main evaluation questions posed by MDE:

- What are the best practices in education evaluation?
- What are the barriers to implementation?
- How do we evaluate the measurement of student growth?

We address the first question by reviewing established best practices in several areas of educator performance evaluation. The first, developing teacher evaluation systems, identifies evidence-based components of a rigorous local evaluation system. The second, selecting tools for conducting a teacher evaluation, surveys a range of tools currently in use, including Charlotte Danielson's Framework for

---

<sup>10</sup> The full literature review can be found in the appendix to this report

---

September 30, 2017

Teaching, the Robert Marzano Approach, the Thoughtful Classroom method, and the 5 Dimensions of Teaching and Learning method. The third, best practices in principal leadership style and student learning, draws on Leithwood's insights into the following questions: How do teachers perceive the leadership of their principal, and what school principal behaviors do teachers perceive as influential in student achievement? The review then analyzes the advantages and disadvantages of using each as a tool for teacher evaluation. In best practices for implementing teacher-performance evaluation measures, we discuss findings regarding the structure of evaluation implementation and the training required to ensure an accurate snapshot of teacher performance. Our review of best practices in teacher evaluation methods examines inspection and demonstration frameworks, reviewing specific examples of each framework and the advantages and disadvantages of deploying each. Best practices in data use and feedback centers on findings that reinforce the importance of translating data gathered during observation/evaluation into feedback educators can use to grow in specific and measurable ways.

We address the second question by examining findings from the literature pointing to several categories of challenge, arising both within and outside of schools, which hinder the implementation of teacher evaluations. Within schools, organizational and personnel issues—among them lack of institutional support, high leadership turnover, inadequate skills, lack of appropriate training, unions, inadequate human resource supports, and inadequate financial support for training and merit raises—are coupled with cultural, technical, and political challenges to inhibit effective implementation of evaluation. Similar barriers that are located outside the school include impediments at the community, state, or national levels. We conclude with a review of three challenges to evaluation design that affect effective implementation of educator evaluation tools: bias, validity, and reliability.

We address the third question, how to evaluate the measurement of student growth, by first clarifying the meaning of student growth in an evaluative context and surveying possible approaches to measuring it. We report research-based observations of two primary approaches to assessing student growth, formative and summative. The former, a generally low-stakes approach designed to monitor student learning and provide regular feedback, enables educators and students alike to iteratively improve their classroom performance. The latter, a higher-stakes assessment designed to evaluate student learning at the end of an instructional unit, compares student performance on the assessment to a normative standard or benchmark. We evaluate two broad approaches to measuring student growth, which roughly parallel the formative-summative distinction: value-added models (and the assumptions underlying them), and alternative methods. Value-added assessment, a statistical method, is one process for isolating the effect instruction has on student learning. Among alternative approaches, student learning objectives (SLO) are most common, since administrators and evaluators may use them in a teacher evaluation/observation regardless of the grade level or subject being taught. A number of other alternative methods have been proposed, but they remain less studied. We focus primarily on a survey of value-added methods and on a summary of the key elements of rigorous, high-quality SLO.

September 30, 2017

### ***Introduction to the Literature Review***

Meaningful teacher evaluation in schools can be an important catalyst for organizational learning and school improvement when it is linked to broader conceptions of leadership in schools (Davis, Elliott, & Annunziata, 2002). For example, a recent report on the Boston Public Schools found that only half of all tenured teachers had been evaluated in the past two years (National Council on Teacher Quality, 2010). Many of the evaluations that do occur consist only of so-called “drive-by” observations, in which a principal stops into a teacher’s classroom for a brief visit and uses a basic checklist of practices to indicate whether the teacher is “satisfactory” or “unsatisfactory” (Toch & Rothman, 2008).

Research suggests that a rigorous evaluation program does boost teacher effectiveness and student achievement (Taylor & Tyler, 2011), and new conceptual and methodological developments in teacher evaluation and effectiveness have emerged in recent years. These stem in part from the changing focus of classroom-based evaluation systems from teaching to learning, and to the work of the National Board for Professional Teaching Standards (NBPTS) to develop assessments for national certification of teachers (Elliott, 2003).

This shift in focus has practical implications. It has been recommended that a comprehensive teacher assessment and evaluation system should have two distinct components, which should remain distinctly separate from one another (Popham, 2013; National Education Association, 2010), implemented as follows: 1) Ongoing, consistent, formative assessments of performance for the sole purpose of fostering professional growth and improved practice and 2) periodic summative evaluations of teacher performance for use in making decisions regarding reappointment, tenure, promotion, etc.

### **A Summary of Documents Reviewed<sup>11</sup>**

MDE relies on newsletters, electronic posting and webinar, among other means, to communicate the requirements and guidelines, and provide training regarding educator evaluation. In addition MISchoolData hosts extensive benchmarking /dashboard data about Michigan districts and schools. Review of these documents gives insight to information available to schools. The data collection and methods section of the Proposal for Educator Evaluation Research and Evaluation Activities submitted by Ray.Taylor and Associates, LLC includes reviewing existing documents from the Michigan Department of Education (MDE) and the Center for Educational Performance and Information (CEPI), among other resources. Overall the documents reviewed provide useful information for educators.

### **Researcher White Papers – methods and outcome summary<sup>12</sup>**

To build context for the review the evaluation team commissioned white papers from two researchers active in the field of education reform and teacher preparation. Dr. Robert Floden, Dean of the College of Education at Michigan State University prepared one of the white papers. The second researcher, Dr.

---

<sup>11</sup> See appendix for the full Documents Review Summary

<sup>12</sup> See appendix for the full copies of the researcher white papers. The views and recommendations expressed in each paper are those of the individual author.

September 30, 2017

Suzanne Wilson also served as an expert consultant for the MCEE. Drs. Floden and Wilson were asked to respond to the prompts from MDE RFP in the context of their professional academic research experience.

The individuals selected met several important criteria: 1) each had deep knowledge of teacher preparation and performance evaluations; 2) each was familiar with Michigan's approach to the evaluation of teacher performance; 3) each was nationally recognized authorities on the topic; and 4) each was currently affiliated with a major research university.

Robert Floden is University Distinguished Professor and Dean of the College of Education, at Michigan State University. Floden received an Bachelor's degree with honors in philosophy from Princeton University, as well as an master of arts degree in statistics and PhD in philosophy of education from Stanford University. Floden is co-editor of the Journal of Teacher Education and chair of the Research Advisory Committee for the National Academy of Education. Floden's work has been published in the Handbook of Research on Teaching, the Handbook of Research on Teacher Education, the Handbook of Research on Mathematics Teaching and Learning, and in many other books and journals. He is currently working on approaches to the evaluation of teacher preparation.

Dr. Suzanne Wilson is an endowed professor and is a nationally renowned expert on teacher preparation and professional development. Currently she is on faculty at the University of Connecticut Neag School of Education. She served on the faculty of Michigan State for 26 years. Dr. Wilson obtained a Master's degree in statistics and a Ph.D. in education from Stanford University. She also served as the first Director of the Teacher Assessment Project (P.I., Lee Shulman), which developed prototype assessments for the National Board for Professional Teaching Standards.

Drs. Floden and Wilson's respective complete research white papers are located in the appendix. Below is a brief summary of major points of each researcher's white paper:

*Dr. Robert Floden's Research White Paper:*

Floden's research white paper focused on practices and issues that have driven and/or fostered attention to the need for a system of teacher performance evaluation. He addresses 1) contextual issues; 2) best practices; 3) barriers to implementation; and 4) the measurement of student growth.

1) Contextual Issues.

- In terms of contextual issues, Floden addresses the need for accountability and the need to continue to improve the educational system and the professional development of teachers.
- Educators should be given detailed constructive feedback on classroom observation— strengths and areas warranting improvement.
- Educators must be a part of the process— engaged and fully informed.
- Student learning and data are crude indicators of performance; these data need to be examined within the context of other factors.

2) Best Practices.

- The system should be transparent, rigorous, and fair.

September 30, 2017

- It should reward those teachers who are the best performing educators. The outcome of the evaluations should inform the school's professional development plans for its teachers.
- Observational tools should be evidence-based.
- The observers need to be highly trained in the observation tools used.
- The teacher being evaluated should understand the tools and system used as well as the outcomes.

### 3) Barriers to Implementation.

- Time and effort required for observation and feedback can be considerable.
- The establishment of student growth/measures for teachers where state assessment data are not available is a must.
- The computation of adjustments to student growth measures to take account of student characteristics in a teacher's class can be challenging.
- The public may not support the adopted evaluation system.
- There may be ambiguity related to the performance standards.
- A lack of buy-in from those being evaluated can undermine the evaluation system.
- Teachers may lack motivation to embrace professional development opportunities. Teachers must be motivated and see these opportunities as adding value.

### 4) Measurement of Student Growth: Tools and Processes.

- Decisions about tests to be used should be made with great care.
- One should look at multiple years of data whenever possible.
- Measures of student growth should be consistent across grade levels, and over time, so that proper consideration can be given to the characteristic of students at the beginning of the school year.
- Educators should be involved in setting the standards.

### *Dr. Suzanne Wilson's Research White Paper:*

Dr. Suzanne Wilson's research white paper, a synthesis of the last 10 years of both policy implementation and education, is divided into the following sections: 1) why do we need educator performance evaluation; 2) a vision for effective teaching; 3) best practices in education evaluation system implementation; 4) barriers/challenges to implementations; and 5) conclusion.

#### 1) Why do we need educator performance evaluation?

The increased emphasis on the evaluation of educator performance can be understood and is driven both by concern to identify teachers who aren't performing and need remediation, and by an interest in identifying areas of growth for all teachers. This emphasis is about being accountable and improving the learning outcomes of students.

#### 2) A Vision for effective teaching.

There are two anchoring visions: a vision of what students learn in school and a vision of what teachers do to enable learning. An ambitious and comprehensive view of effective teaching anchors a high

September 30, 2017

quality evaluation system. Michigan has a set of professional standards but the assumptions undergirding the four observation tools for teachers that were state approved (Danielson's Framework, the Marzano Teaching Evaluation Model, the Thoughtful Classroom and the 5 Dimensions of Teaching and Learning) are each based on different assumptions/conceptions of effective teaching.

3) Best practices in education evaluation system implementation

- An elective state system allows for flexibility, but makes it difficult to compare progress across districts.
- Multiple measures provide a more accurate estimate of teacher effectiveness; no research sheds light on exactly how many measures are optimal.
- Research demonstrates that educator evaluation systems need to be consistently implemented and resourced.
- The quality of the data procured from observation of the teacher in the classroom depends on the quality of the system itself. Observers need to be trained with explicit guidelines.
- The use of evaluation results to improve teaching needs additional study. However, high powered incentives linked to multiple indicators of teacher performance did substantially improve measured performance.

4) Barriers/Challenges to Implementation.

- The school district must have “buy-in” from broad-based stakeholders. The absence of “buy-in” is one of several barriers.
- One must limit the sentiment that the evaluation system is a top down mandate. Instead, develop a system whereby teachers are engaged and full participants. The system must be relevant and meaningful and have consequences.
- The system must provide useful feedback that enhances professional development.
- There are human and financial costs in conducting meaningful teacher evaluations. The process is resource intensive, e.g., trained observers, observation members, setting aside time to meet with the teacher to discuss observations, and creating a data/information management system.
- The school must adopt a culture of continuous improvement that is shared by all. A mandate to adopt a system of education will not work.

5) Conclusion.

Based on an analysis of states' experiences in creating systems, and in implementing educator evaluation systems, nine critical steps are offered (see research white papers). The author concludes that state systems that are implemented need to be regularly evaluated for bias (measures, subject matter, etc.) and for their effectiveness.



September 30, 2017

## **Voices from the Field: Practitioner White Papers – methods and outcome summary<sup>13</sup>**

A central theme of the educator evaluation design conducted by Ray.Taylor and Associates is to learn directly from those who are evaluated and who are responsible for conducting evaluations. These practitioners are given voice in this evaluation in four ways. In addition to surveys, interviews and focus groups the evaluation team asked three practitioners to develop white papers based on their experience and to share their observations and recommendations. We gave few parameters for these papers. We did share the overall purpose of the evaluation and the three questions posed by MDE and the other data gathering components. Aside from this the practitioners were free to construct their papers.

The use of white papers from K-12 practitioners was one of the data gathering designs of Ray.Taylor and Associates. Three individuals who had served as a teacher, supervisory and administrator agreed to author papers from three to five pages in length. These three practitioners each have first-hand experience with being evaluated and conducting evaluation in school settings. Their perspective adds to the context for this evaluation. The papers were written in response to three questions of interest: Identify best practices in the educator evaluation implementation to inform ongoing and future MDE supports to, and programming by, districts; identify barriers to implementation in local educator evaluation systems and make recommendations for ways to mitigate barriers and inform ongoing and future supports to, and programming by, districts; evaluate the measurement of student growth using alternative methods/tools and processes in PA 173 of 2015.

The writings of the three practitioners generated the following common themes. Teacher evaluation models must be viewed as a vehicle for professional growth and not punitive in nature. Time must become a more valued resource during the evaluation process, including pre and post evaluation meetings. Trust was referenced in all of the white papers. The professional relationship between the individual doing the evaluating and the evaluatee was crucial to the meaningfulness of the evaluation. Establishing professional learning communities added value to the evaluation process, especially for new teachers. Understanding student growth measures required that teachers be trained in data for instructional decision making. Student performance on assessment measures should not be the sole device for determining teacher effectiveness.

### *Introduction to Practitioner White Paper Summary*

#### *Selection of Practitioners*

The selection of the K-12 practitioners for preparing white papers relied heavily on the previous educational experiences of the Senior Associates with Ray.Taylor and Associates. Team meetings set aside an agenda item designated for discussing potential practitioners, with priority given to individuals currently working as a teacher or administrator in public schools in the State of Michigan, including

---

<sup>13</sup> Practitioner white papers were commissioned by Ray.Taylor and Associates as a component of GOISD funded MDE Educator Evaluation Research Project. The views and recommendations expressed in each paper are those of the individual authors. See appendix for full copies of practitioner white papers



September 30, 2017

public school academies. Graduate students, central office personnel, or intermediate school district employees were also considered by the team.

Three individuals agreed to author a white paper. They were the principal of a public school academy, and two recent K-12 administrators and current doctoral students. All were presented the three questions of interest: Identify best practices in the educator evaluation; Identify barriers to implementation in local educator evaluation systems and make recommendations for ways to mitigate barriers and inform ongoing and future supports to, and programming by, districts; Evaluate the measurement of student growth using alternative methods/tools and processes in PA 173 of 2015.

### *Background of White Paper Practitioners*

The practitioners submitting white papers were Crystal Wise, Ann Blais and Rosiland Brathwaite.

Ms. Crystal Wise is a former elementary school teacher with a Bachelor of Arts in Elementary Education from Indiana University, a Master's in Educational Leadership from Indiana University, a Master's in Language and Literacy from Harvard Graduate School of Education. Ms. Wise is currently working on a Doctorate in Literacy, Language and Culture from the University of Michigan. She has also served as a graduate instructor and research assistant.

Ms. Ann Blais has been a high school English teacher, a department chair, a professional developer and a university lecturer. She has a Bachelor of Arts degree from Harvard and Radcliff, a Masters of Arts from Teachers College and is currently enrolled in a doctoral program in Teaching and Teacher Education at the University of Michigan. She has led and developed multiple, long-term innovative initiatives to close opportunity gap for students of color, particularly at the advanced placement / honors level.

Ms. Rosiland Brathwaite is a former elementary teacher, elementary principal, dean of students and is currently a middle school principal at a public school academy. She has a Bachelor of Arts in Elementary Education with a minor in Psychology from Chicago State University, and a Master of Arts in Supervision and Administration from the University of Phoenix.

### *Findings From Practitioner White Papers<sup>14</sup>*

**Best Practices:** The practitioners seemed to have agreement on the use of professional development as a best practice. Agreement was also apparent when professional learning communities existed for teachers to support each other with positive feedback for professional growth as part of a peer review process or mentor relationship.

Trust between evaluators and evaluatees seemed to be a common theme from the practitioners, especially when trust was established early in the evaluation process.

Immediate feedback was referenced by the practitioners as a best practice that would benefit teachers immensely, both written and in person with the evaluator.

---

<sup>14</sup> See appendix for full copy of practitioner white papers

September 30, 2017

While the value of formal evaluations was the focus of the practitioners, they also valued the importance of informal feedback as a contributor to meaningful evaluation. The unannounced visit by the evaluator, especially the principal, allowed the teacher to feel supported. This support was strengthened when there was immediate feedback.

**Barriers to Effective Evaluations:** The practitioners mirrored the focus groups when pointing out barriers which could also be best practices. Using professional development as a tool for anything related to reprimanding or non-professional growth was viewed as a barrier. Effective evaluation must emphasize professional growth and improving student performance.

Practitioners pointed to the lack of administrative time and not implementing evaluation models with fidelity as barriers.

Immediate feedback for informal and formal evaluations was also seen by the practitioners as a barrier. Another barrier was the lack of trust in the evaluation process, which prevented meaningful evaluations from occurring.

The practitioners described the lack of training for teachers and administrators as a barrier when it came to using data for instruction.

**Student Growth Using Alternative Methods/Tools:** The practitioners appeared to agree that using data, especially data for determining student growth, required training through professional development. While growth models are necessary, the practitioners cautioned against their use as the sole source when evaluating teachers.

Other important areas cited by one of the practitioners but not necessarily mirrored by all three were the following: Having evaluators who were strong in pedagogical content knowledge and how to teach content as a best practice in evaluation. One formal evaluation was not sufficient or conducive to effective evaluation. Alternative forms of assessments such as pre and post must take into consideration validity and reliability. Curriculum and instruction that attempt to incorporate culturally relevant pedagogy are overtaken by past experiences and normal approaches in school districts.

### **Focus groups – methods and outcome summary**

In yet another strategy employed to learn directly from educators Ray.Taylor and Associates designed and conducted four focus groups to solicit perceptions and experiences related to the educator evaluation systems across the state. The process was designed to encourage frank and open discussion centering on best practices, barriers to implementation and student growth measurements used in educator evaluations. The half-day focus groups were designed to provide feedback from field practitioners on the strengths and issues of the current systems.

September 30, 2017

## *METHODS*

Ray.Taylor and Associates sent invitations through MDE to all ISD/RESA superintendents in the state, asking them to identify potential participants in focus groups. Focus groups were held in four locations to accommodate reasonable travel time and promote attendance. The request was for a diverse population of teachers and administrators in order to obtain a range of experiences, opinions, and reflections on the essential questions posed at the focus group sessions. Follow up requests were sent to encourage responses, and Ray.Taylor and Associates sent invitations to each person identified through the process. The invitations explained the purpose for the focus groups, and identified dates and locations of the sessions. Participating educators selected the location and date that best met their schedule. In total, 48 individuals attended a focus group session. Ray.Taylor and Associates designed the sessions to include approximately 6 to 20 participants to allow for free conversation and full participation of attendees.

## *THE AGENDA AND DISCUSSION QUESTIONS<sup>15</sup>*

A list of launch questions for discussions was created by Ray.Taylor and Associates and reviewed with MDE. These questions appeared on the session agenda. More detailed probing questions were also developed and reviewed with MDE. These were used by the facilitator and Ray.Taylor and Associates team members to deepen and extend the conversation as needed.

## *FACILITATION, NOTE TAKING AND RECORDING*

Each session was facilitated by a Ray.Taylor and Associates team member, with 2-3 team members serving as note takers/observers. Sessions were audio recorded for note taking verification, with assurances to participants that their individual comments and responses would not be attributed by name or district. Following each session, participants received a feedback sheet on which to add additional comments if they desired. Notes, comments, and possible quotes from each focus group session were compiled into a reference document. The Ray.Taylor and Associates team reviewed all session notes for consistency, omissions, misstatements or incomplete information. Audio tapes were available to validate notes and quotations as needed.

## *FINDINGS*

There were consistent themes and language reflected during the sessions. Many best practices were identified, often with qualifying statements regarding the limitations of faithful implementation of those practices. Common barriers included having inadequate time to implement the existing evaluation process as designed and expected, and dealing with the ambivalence of the dual purposes of evaluation – personnel decisions and improvement of instructional practice. Participants raised a variety of implementation issues including communication, feedback, and the reliability of ratings between evaluators. As we will see later in this review issues raised by the focus groups echo survey findings from educators across the state.

The three areas of interest – best practices, barriers, and student growth measurement – generated spirited discussion at all sessions. Here is one example.

---

<sup>15</sup> Focus group agenda and supporting documents are in the appendix.

September 30, 2017

*“I see highly effective as where you should be.”*

*“It is fluid and [they] should get rid of highly effective label – ridiculous label – if have label that is where everyone wants to be.”*

*“I communicate that with teachers but [it] doesn’t matter, everyone wants highly effective”.*

*“Comes from how evaluation used to be – we will get there.”*

*“But if you split hairs over how one student is doing it will be the same old **Got Ya** – need to look at why we are here – to make everyone better.”*

Following the focus group session participants completed a survey. In it participants highlighted points made during the focus group and offered new thoughts. The focus group participants appreciated that MDE was asking for and willing to listen to and act on their feedback. They also appreciated the opportunity to talk with and share ideas with colleagues. These are important statements given the distance traveled and the time commitment required on what were incredibly demanding schedules of these educators.

*“Thank you for allowing us the opportunity to have a voice in this process. I especially appreciate that it was teachers and administrators working together”.*

– Focus group participant feedback survey



Focus group participants were from 41 districts from throughout Michigan.

#### Focus Group Snapshot - Who Participated

##### *48 Educators participated*

8 at COOR ISD

19 at Kent ISD

9 at Wayne RESA

12 at Marquette-Alger RESA

##### *Of the participants there were:*

13 Teachers including elementary and secondary, special education, and career and technical education

21 Principals including elementary and secondary, and career center

4 Superintendents

4 Other administrators

(We are confirming role of the remaining 6 educators)

##### *Participants were from 41 school districts and ISD / RESA*

#### Evaluation Models Used by Participants' Schools

20 reported using Danielson

18 reported using 5 Dimensions

5 reported using Thoughtful Classroom

4 reported using Marzano

(1 participant did not report the model used by their schools)

September 30, 2017

### Statewide Educator Survey– methods and outcome summary <sup>16</sup>

Ray.Taylor and Associates designed and conducted a statewide survey of Michigan educators regarding educator evaluation to gather perceptions from the widest possible audience of K-12 educators. The survey design process included review of literature and former surveys conducted by MDE, review of focus group notes, and analysis of RFP guiding questions. The survey was developed and posted using the Survey Monkey application for electronic distribution. The draft survey was piloted with 11 educators for their input regarding length and clarity of the survey and reviewed by the Michigan Department of Education representatives. The survey was distributed electronically to all Michigan teachers, principals, assistant principals, superintendents and central office administrators using MDE email lists. Over 80,000 emails containing a link to the survey were distributed. The survey remained open from June 1 through July 16, 2017 and 9,000 survey responses were received.

Figure 2: How the survey was distributed

To whom	Date sent	Number
<b>Principals</b>	June 1, 2017 (by MDE) June 15 reminder	3,554
<b>Teachers</b>	June 5 – 7 June 16 reminder	78,960
<b>Superintendents</b>	June 8 (by MDE in MDE Official Weekly Communication) To Middle Cities Education Association(MCEA) for inclusion in meeting packet and posting in newsletter (sent to MCEA June 16) June 21-23 Sent to Michigan Association of School Administrators (MASA) members through newsletter June 23	
<b>Other</b>	Survey link emailed to Focus Group participants June 20	48
	Cory L. Micheel-Mays, Executive Director, Michigan Music Education Association (MMEA), Chair, Michigan Music Conference (MMC) and Chair, NAFME Council of State Executives Requested and received permission to distribute the survey link to music educators.	NA

### Survey Results

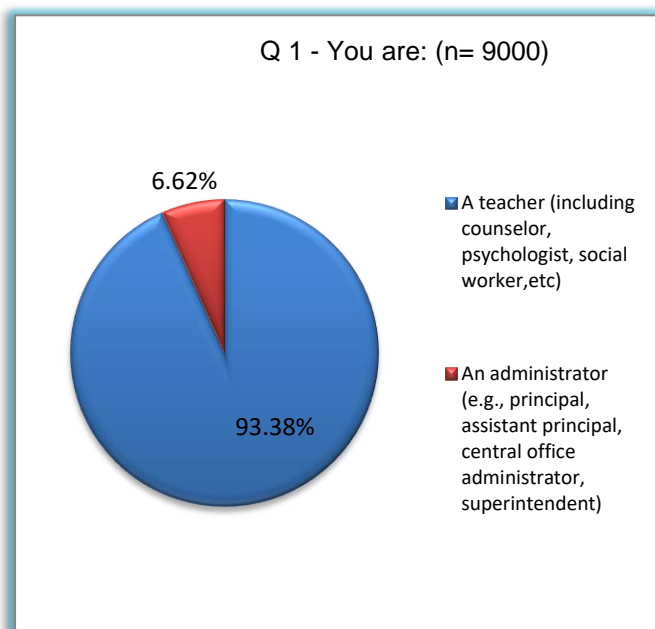
Charts of the results of our statewide survey will be included throughout this report. The title of each chart will include the number of the survey question and note if the question was directed toward teachers or administrators. For example Q1 was answered by all survey respondents, and Q5 – Teachers: was directed toward teachers. The first survey question (Q1) in which respondents identified whether they were a teacher or administrator was used to branch respondents to questions designed

<sup>16</sup> The survey distribution email, along with a copy of the full survey, is available in the appendix.

September 30, 2017

for teachers (questions 2-34) and questions designed for administrators (questions 35-81). The appendix to this report contains a copy of the full survey.

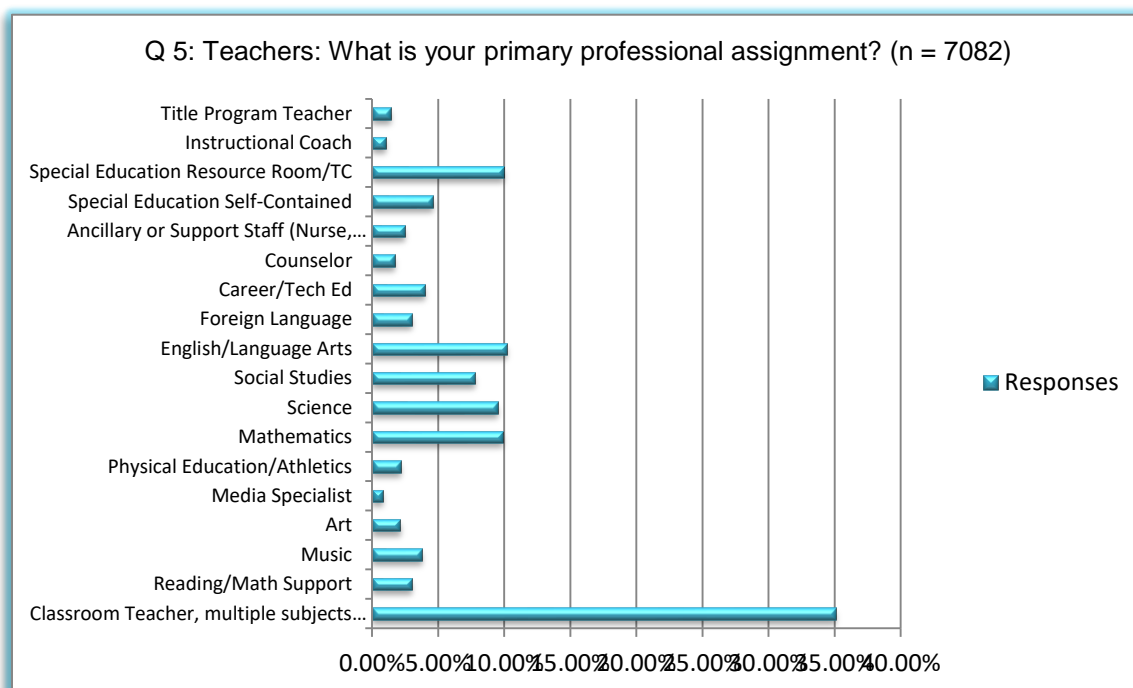
Figure 3: Description of survey respondents



Survey responses were received from 8,404 teachers (including counselors, psychologists, social workers, etc.) or 93.38 percent of the survey responses, and from 596 administrators (e.g., principals, assistant principals, central office administrators, superintendents), representing 6.62 percent of survey responses.

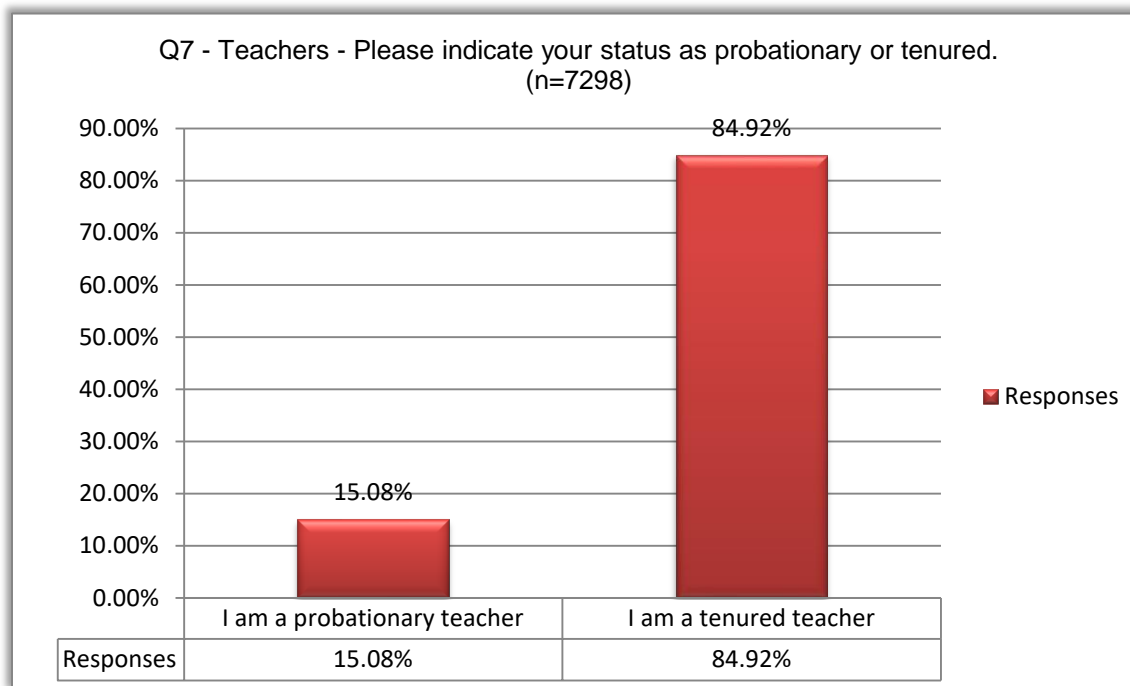
The vast majority of survey respondents identified themselves as classroom teachers of core subjects, and as tenured teachers.

Figure 4: Teachers: Professional assignment



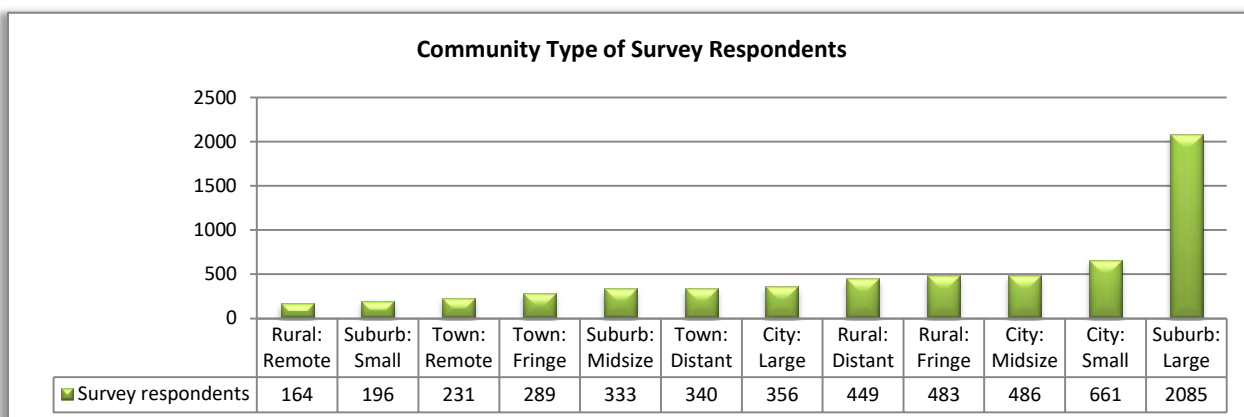
September 30, 2017

Figure 5: Teachers: Tenure status



Of the survey respondents who identified their school district (6,366 respondents) we are able to show the distribution among urban, suburban and rural locations using census designations. Although only a portion of the respondents, the data and map below give some information about the representation of survey respondents.

Figure 6: Distribution of survey respondents identified by U.S. census designation for community type.



September 30, 2017

**Survey analysis:** Survey responses were analyzed through frequency tools from the survey application and through Ray.Taylor and Associates’ analysis of cross tabulated data. The survey analysis provided a host of information including the differences among groups completing the survey (e.g. teachers responses versus administrators; principal and assistant principals versus other administrator). Far more combinations for analysis purposes are available in the data provided in the survey than are included in this report. The report presents survey data that addresses the questions raised by MDE as framework for this review. See the survey methodology section of this report contained in the appendix for further discussion of data analysis methods used.

## Interviews

The Ray.Taylor and Associates evaluation team conducted informal interviews with two researchers and two practitioners with deep knowledge of school reform, teacher professional development and educator evaluation. The purpose of these unscripted interviews was to gather deeper knowledge and nuance to the context and conduct of educator evaluation.

## Identification of Best Practices in Action

The request for proposal posted by MDE defined the questions that frame this evaluation study. For each of the sub-questions raised by MDE a parallel question was asked about where in Michigan is there evidence of best practices in action. To address this question the Ray.Taylor and Associates evaluation team decided to ask educators to tell us who is using best practices.

## Our Process

We identified sites using best practices in three ways. First, we developed an application form distributed electronically asking for self-nominations from schools, LEAs and ISD/RESA. The request for self-nominations was distributed through regional representatives meeting with MDE in June 2017. They passed the form along to local districts. A second way of identifying districts using best practices was to contact a particular focus group participant who had described his district’s use of practices during one of the focus group sessions. This administrator was completing a case study about his district’s implementation of the educator evaluation system for his dissertation. He was invited to submit a self-nomination. The third approach for identifying use of best practices was to review clusters of survey findings. Two specific survey questions explored the use of best practices by survey respondents. We looked at the cluster of responses to these questions to identify reported evidence of use of best practices in individual districts and schools.

Eight self-nominations were submitted. Following review by the Ray.Taylor and Associates team telephone interviews were conducted with district contact people. Eight interviews were conducted along with review of supporting documents provided by the districts and review of district website transparency links. Selected best practices are described in detail in the appendix and appear throughout the body of this report as sidebar titled *Best Practices in Action*. Evidence of best practices identified through survey cluster analysis is displayed in sidebar titled *Examples from the field*. More details are found in the appendix.



September 30, 2017

The districts and schools cited as using best practices are intended as a starting point for professional conversations regarding implementation of an educator evaluation model. In each best practice area identified you can see the attempt of the district or school to draw upon evidence-based practices to overcome barriers, meet needs and improve professional practice.

The next section of the report presents the findings of the review and discusses each of the three questions of interest to MDE.

The first MDE activity question is: *Identify best practices in educator evaluation implementation to inform ongoing and future MDE supports to, and programming by, districts. In this section of the report we will begin our focus on what models are used for evaluation, and the process and evidence of implementation.*

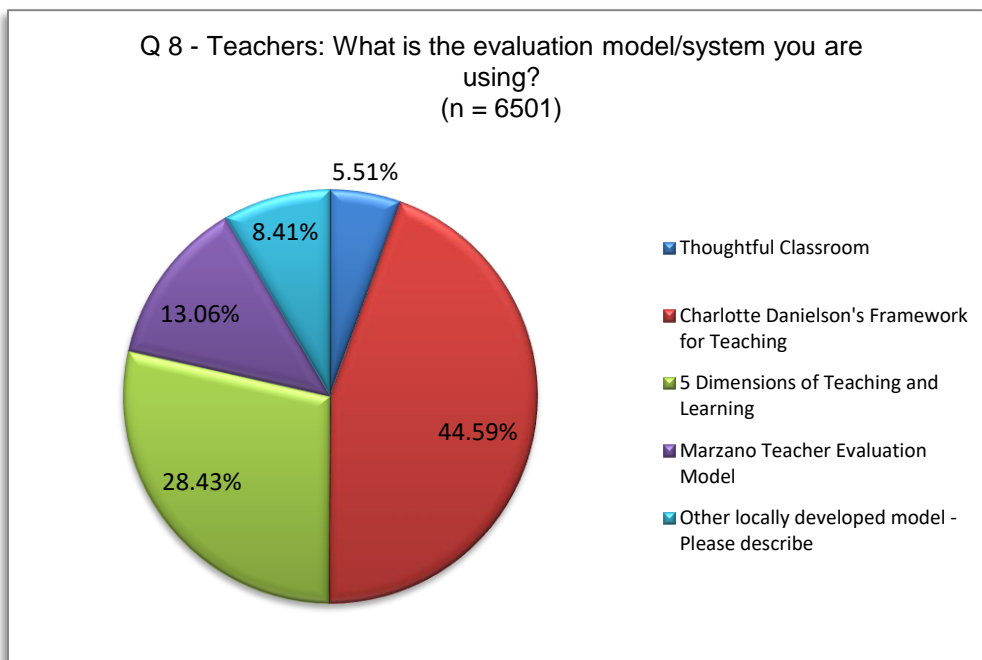
Deming Process Relationships  
Appropriate Evidence/Artifacts  
Structured feedback Honest  
Reflective Buy in  
Informative Growth Mindset  
Risk-Taking Transparent  
Realistic Safe Sharing  
Thorough Coaching Constructive Conversation  
Overdone  
Real Meaningful Open Conversation  
Accurate Frequent Consistency  
Individualized Mindset! Systematic  
Coherence Quality Feedback Conversation  
Be Relevant Create Change  
Engaged Dialogue  
Meaningful Feedback

The Michigan legislation requires that schools and districts select from an approved list of models for teacher and administrator evaluation. Our survey shows that 91.59 percent of respondents report using one of the four approved models for teacher evaluation, and 8.41 percent report use of another not approved model “other”. These figures are compared with findings of the 2013-2014 Educator Evaluations & Effectiveness in Michigan report by MDE<sup>17</sup>

33

September 30, 2017

Figure 7: Evaluation Models Used



The next table shows the distribution of models used as reported by teachers (Q 8), administrators (Q 42), a focus group participants, and as reported in the 2013-2014 MDE evaluation report.

September 30, 2017

Figure 8: Evaluation model used as reported by teachers and administrators in survey, during focus groups, and reported in 2013-2014 MDE Report

Teachers (Q 8)  n = 6501	%	Administrators (Q42)  n = 347	%	Focus group participants  n = 48	Actual number and %	MDE Educator Evaluations & Effectiveness in Michigan 2013-2014 n = 775	%
<b>Charlotte Danielson's Framework for Teaching</b>	44.59	Charlotte Danielson's Framework for Teaching	42.36	Charlotte Danielson's Framework for Teaching	20 41%	Charlotte Danielson's Framework for Teaching	61.4
<b>5 Dimensions of Teaching and Learning</b>	28.43	5 Dimensions of Teaching and Learning	30.84	5 Dimensions of Teaching and Learning	18 37.5%	Other locally developed or Other Tool	43.9
<b>Marzano Teacher Evaluation Model</b>	13.06	Marzano Teacher Evaluation Model	13.54	Thoughtful Classroom	5 10.4%	Marzano Teacher Evaluation Model	18.1
<b>Other locally developed model</b>	8.41	Thoughtful Classroom	10.66	Marzano Teacher Evaluation Model	4 8.3%	5 Dimensions of Teaching and Learning	11.1
<b>Thoughtful Classroom</b>	5.51	Other locally developed model	2.59	--		Thoughtful Classroom	2.7

Similar to the findings for teacher evaluation models, 65.88 % of administrators responding to our survey report using one of the 2 approved models for administrator evaluation and 34.12 percent report using “other”.

Figure 9: Administrator evaluation system used (Q 39 n= 340).

<b>MASA School ADvance Administrator Evaluation Instrument</b>	<b>58.53%</b>
<b>Other</b>	<b>34.12</b>
<b>The Multidimensional Leadership Performance System (formerly Reeves Leadership Performance Rubric)</b>	<b>7.35</b>

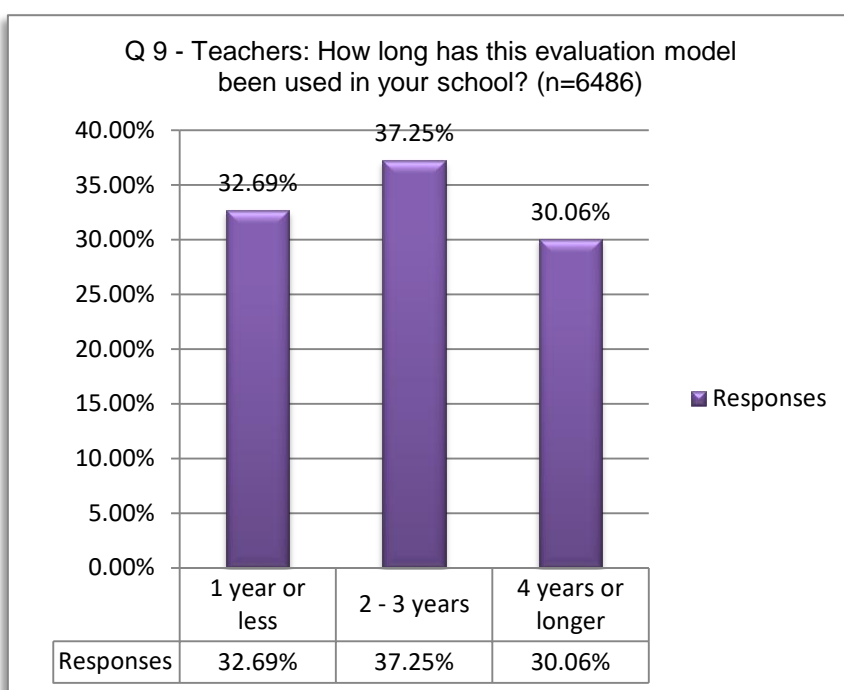
The 2013-2014 MDE report showed a number of evaluation tools used for administrator evaluation including some that are not currently approved for administrator evaluation.

September 30, 2017

Figure 10: 2013-2014 Frameworks Used in Local Evaluations of Administrator Professional Practice (n=611)<sup>18</sup>

Other	MASA School ADvance Administrator Evaluation Instrument	Marzano Leadership Evaluation Model	Reeves Leadership Performance Rubric (now The Multidimensional Leadership Performance System)	NHA- Developed Observation Protocol	Ron Edmonds Effective Schools Model
53.4%	45.8	26.5	8.5	7.7	2.1

Figure 11: Length of time evaluation model used



### The Process – How Educator Evaluations are Conducted

We asked the rating survey respondents generally expect to receive in their evaluation. Of the 5470 teachers who responded to this question only 54 expected a rating of *minimally effective* or *ineffective* (0.75 percent and 0.24 percent of respondents respectively). 99 percent of teachers expected to receive a rating of *effective* or *highly effective* (52.65 percent and

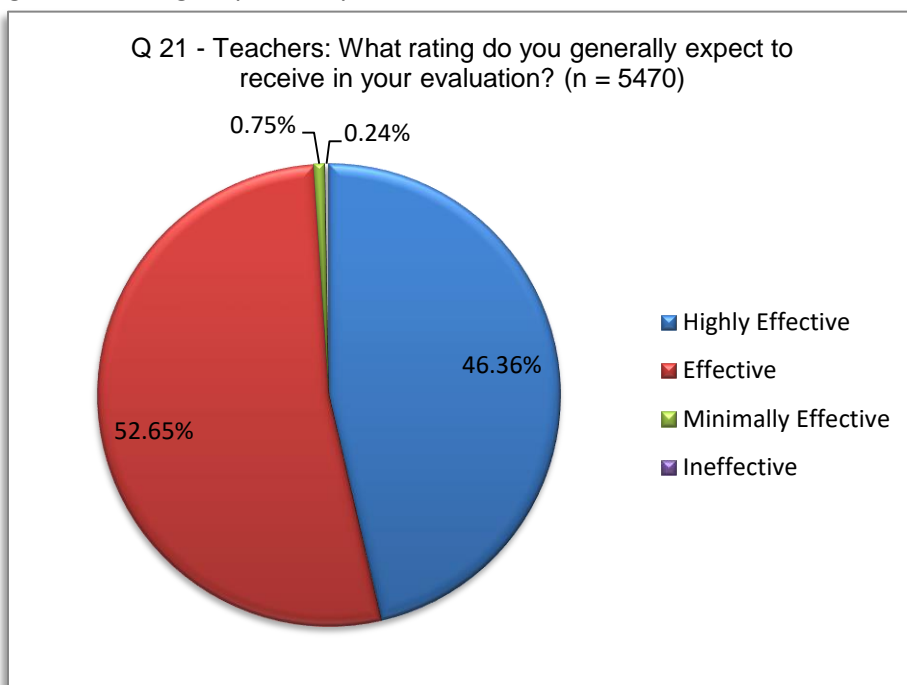
46.36 percent respectively). We asked a similar question of administrators and found similar responses. Of the 250 administrators who answered this question *none* of the administrators responding expected to be rated *ineffective*, and only 6 *minimally effective*. 97 percent of administrators expected to be rated *effective* or *highly effective* (76.40 percent and 21.20 percent respectively). There are many things that

<sup>18</sup> Ibid. page 9, figure 9

September 30, 2017

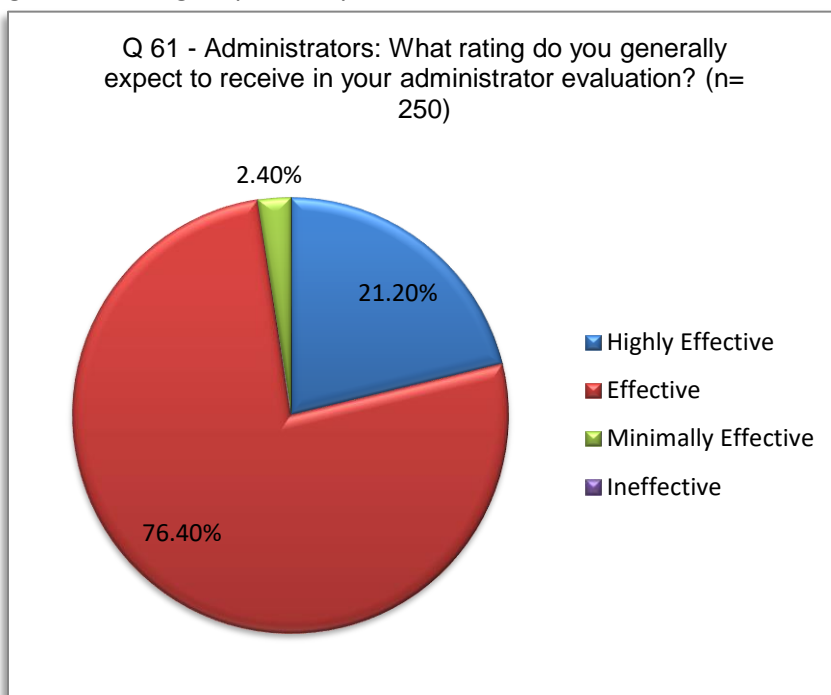
contribute to this finding that we discuss later in this report, not the least of which the potential for self-selection of who was willing to complete the survey or who was willing to answer the question frankly. Suffice it to say that nearly all of the respondents to this survey say that they expect to be rated *effective* or *highly effective*.

Figure 12: Ratings expected by teachers



September 30, 2017

Figure 13: Ratings expected by administrators



### Classroom Observation Models and Tools

Several indicators can describe the implementation process used by schools for educator evaluation. Differences in implementation can give information about fidelity, usefulness, effectiveness and effect of the total evaluation model. Focus group participants told us there is a considerable amount of variety in implementation processes seen across and within districts.

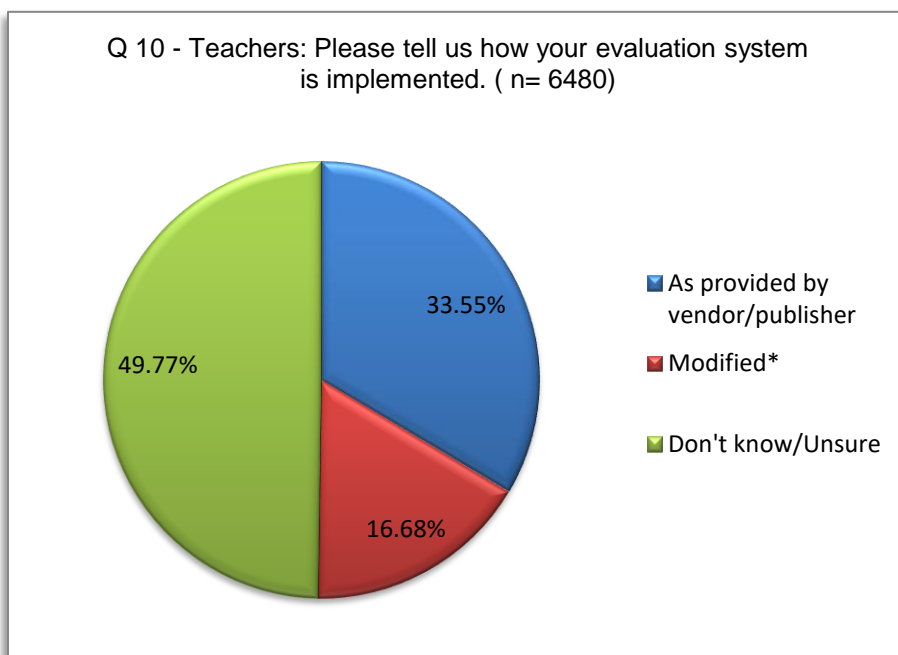
Implementation is important no matter if the evaluation model has tested for reliability and validity by vendors, variation in implementation can compromise both.

Consideration of the precision, reliability and validity of the observation /evaluation models and measures is most pertinent when making high-stakes personnel decisions.

Half of teachers responding did not know if the model used in their district was implemented as provided by the vendor (49.77 percent). And 16.68 percent of teacher survey respondents report that the model was modified.

Figure 14: Implementation

September 30, 2017



Focus group participants provide some detail. They described modifications made to evaluation models to simplify and shorten implementation, or to focus attention on most valued domains. For instance, the Danielson model is often cited as an example. The version of the Danielson model used by many districts has 72 indicators and four domains. Some focus group participants lamented that it was nearly impossible to effectively attend to all elements at once. Here the pragmatic conflicts with measurement fidelity and design choices.

However, no matter how well intended modifications may compromise reliability of the model (citations from Rowan, Roebber, etc.). This may be less important if the evaluation model is used as a tool for discussion and feedback leading to strategies for continuous improvement, but may be critical if the evaluation process is used to lead to high stakes human resources decisions like termination or pay decisions.

Another design consideration cited in the literature and reflected in focus group, survey and white papers is the extent that there is coherence in the evaluator ratings within and between schools and districts, and over time. Rowan and others question inter-rater reliability of the instruments and models to provide credible information for high stakes decisions.

The extent to which educators use processes, tools or strategies to calibrate observations to improve inter-rater reliability is unknown. Some of the training tools (e.g., TeachScape /Frontline<sup>19</sup>) provide calibration training at extra cost. Focus group participants expressed concern regarding differences in judgement and lack of consistency by evaluators. They describe concern for inconsistencies in ratings

<sup>19</sup> TeachScape is the training tool associated with the Danielson model. It has recently been purchased and renamed Frontline



September 30, 2017

within schools, within the school district, and across school districts. Some focus group participants reported teachers requesting specific evaluators that they expected to be more lenient. Others expressed concern for the impact of moving from a district with more rigorous evaluation rating to a district with more lenient rating. To paraphrase one focus group participant: *“How will my rating of effective compare in a district where everyone is rated highly effective?”*

### Training

Considerable differences are seen in the amount of training reported by survey participants. Training seems to be focused on those conducting evaluations. Administrators receive the most training in their role as conducting evaluations of teachers. More than 85 percent of administrators report 6 hours or more training for initial preparations, and another 40% report 6 hours or more for ongoing annual training. In contrast, 24 percent of teachers report 6 hours or more training for initial evaluation, with 66% in receipt of 1-3 hours training. Ten percent of teachers say they receive more than 6 hours annual training, and 53 percent say they receive just one to three hours annual training. Principals and assistant principals report more training time than teachers.

### *Best Practices in Action*

#### *Leverage Leadership – Newaygo County RESA*

**Administrators in Newaygo work collaboratively to improve their skills in classroom observation and feedback. Using the Leverage Leadership model, administrators get real time classroom observation training and practice with feedback from their colleagues. The strategy works to build a professional learning community among participating administrators and to improve inter-rater reliability in observation ratings.**

**Contact Person: Nicole Gasper, Chief Instructional Officer**

September 30, 2017

Figure 15: Training provided to teachers

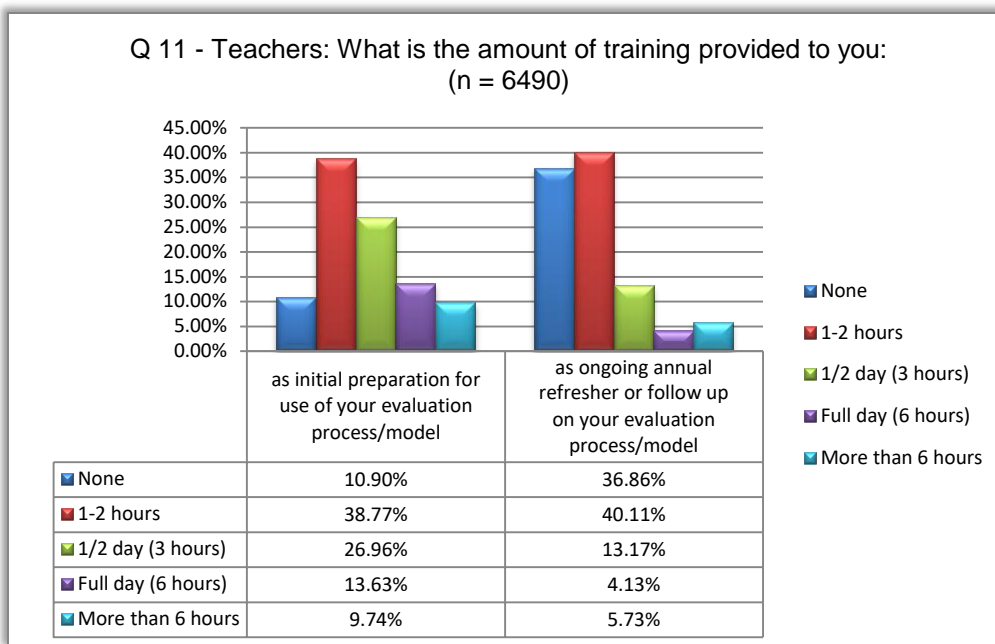
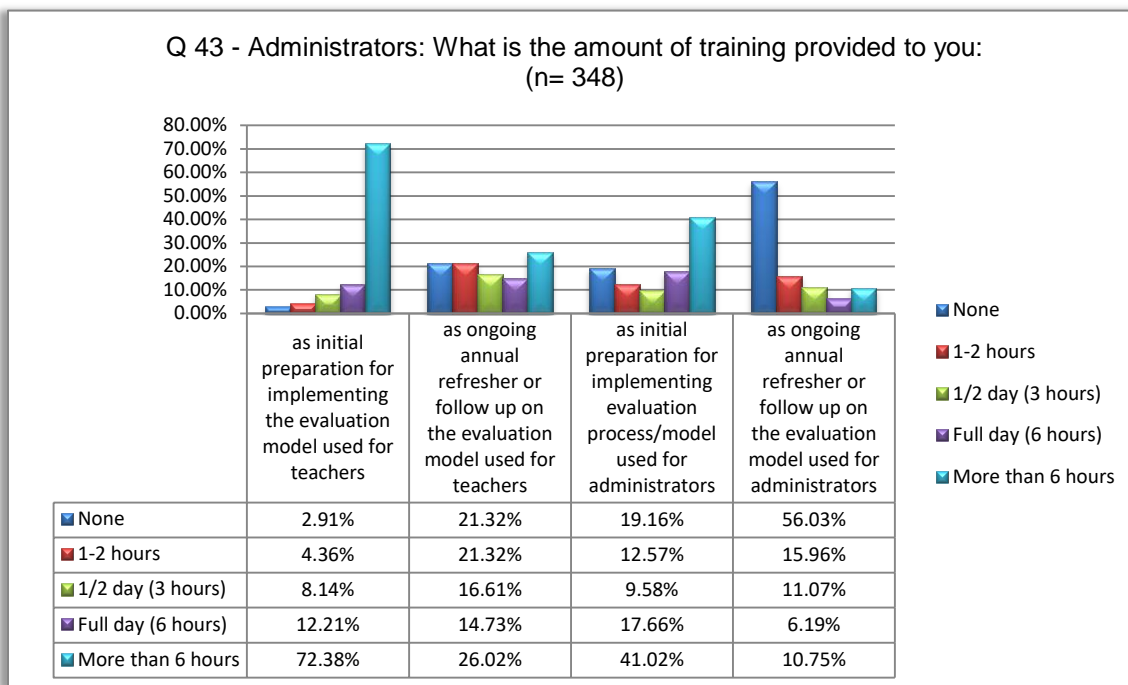


Figure 16: Training provided to Administrators

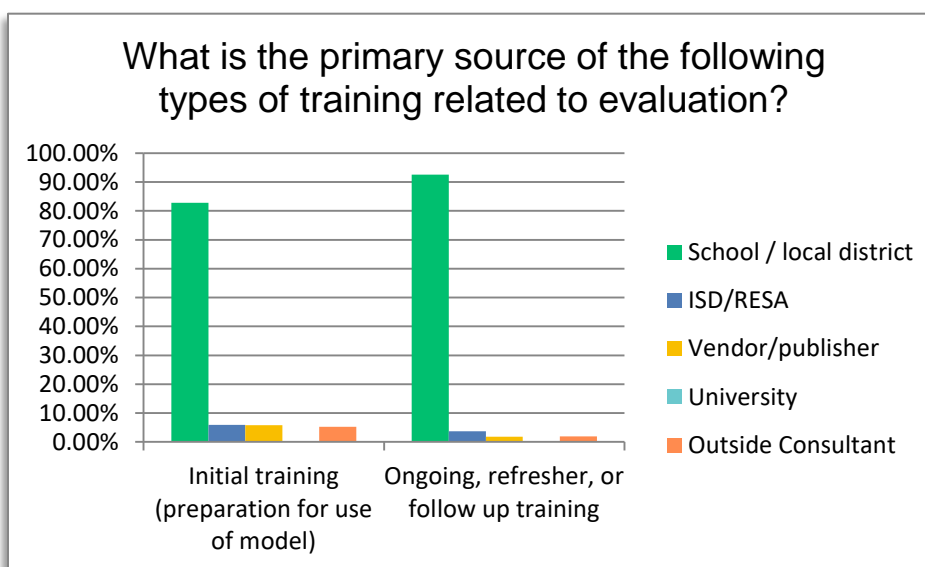


September 30, 2017

Focus group participants report similar disparities between the amount of training given principals/ assistant principals and teachers. Along with more training for administrators responsible for conducting evaluations, focus group participants requested more preparation of teachers as the recipients of evaluation. Some focus group participants also questioned the quality of training provided and lack of follow up. They observed that principal-provided training to teachers may be a watered-down version of the more substantial training provided by the vendors. The white papers from researchers and practitioners, and focus group participants echoed a need for more and improved quality of training. This request was contrasted with concern for the limited amounts of time that educators have and demands on time.

Nearly all (87 percent) of the training provided for teachers was from the local district, as opposed to ISD/ RESA (5.86%), vendors (5.83%), outside contractors (5.26 percent), or universities (0.26 percent). The numbers of teachers to train, time and cost likely influence the delivery system for teacher training. By contrast administrators were much more likely to be trained by vendors (33.72 percent), with others training providers evenly split among outside consultants (23.17 percent), local district (21.99 percent), and ISD/RESA (21.11 percent).

Figure 17: Q13 Teachers: Source of training n = 6214

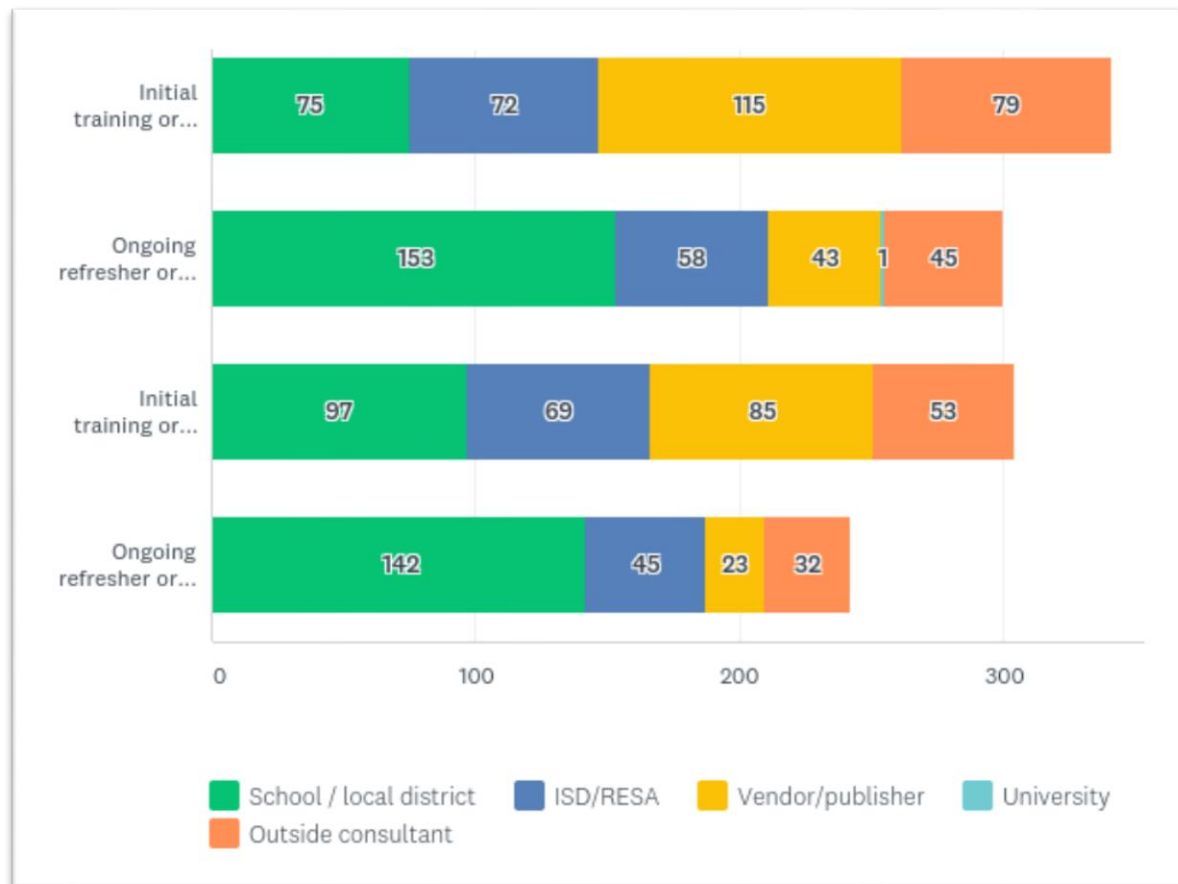


Administrators were also asked the primary source of training that they receive to use the teacher evaluation model and to use the administrator evaluation model. Administrators were much more likely to report receiving training provided by the vendor for their initial training in teacher evaluation models.

September 30, 2017

Figure 18: Q 54 – Administrators: Source of training n = 345

**Q45: What is the primary source of the following types of training related to evaluation?**



September 30, 2017

	SCHOOL / LOCAL DISTRICT	ISD/RESA	VENDOR/PUBLISHER	UNIVERSITY	OUTSIDE CONSULTANT	TOTAL
Initial training or preparation for use of teacher model	21.99% 75	21.11% 72	33.72% 115	0.00% 0	23.17% 79	341
Ongoing refresher or follow up training for teacher model	51.00% 153	19.33% 58	14.33% 43	0.33% 1	15.00% 45	300
Initial training or preparation for use of administrator model	31.91% 97	22.70% 69	27.96% 85	0.00% 0	17.43% 53	304
Ongoing refresher or follow up training for administrator model	58.68% 142	18.60% 45	9.50% 23	0.00% 0	13.22% 32	242

September 30, 2017

In addition to contrasting training provided teachers versus training provided administrators the design of our survey allows the opportunity to contrast the responses of school-based administrators (i.e., principals and assistant principals), and other administrators. Figures 19-27 show the contrasts between school-based administrators and other administrators.

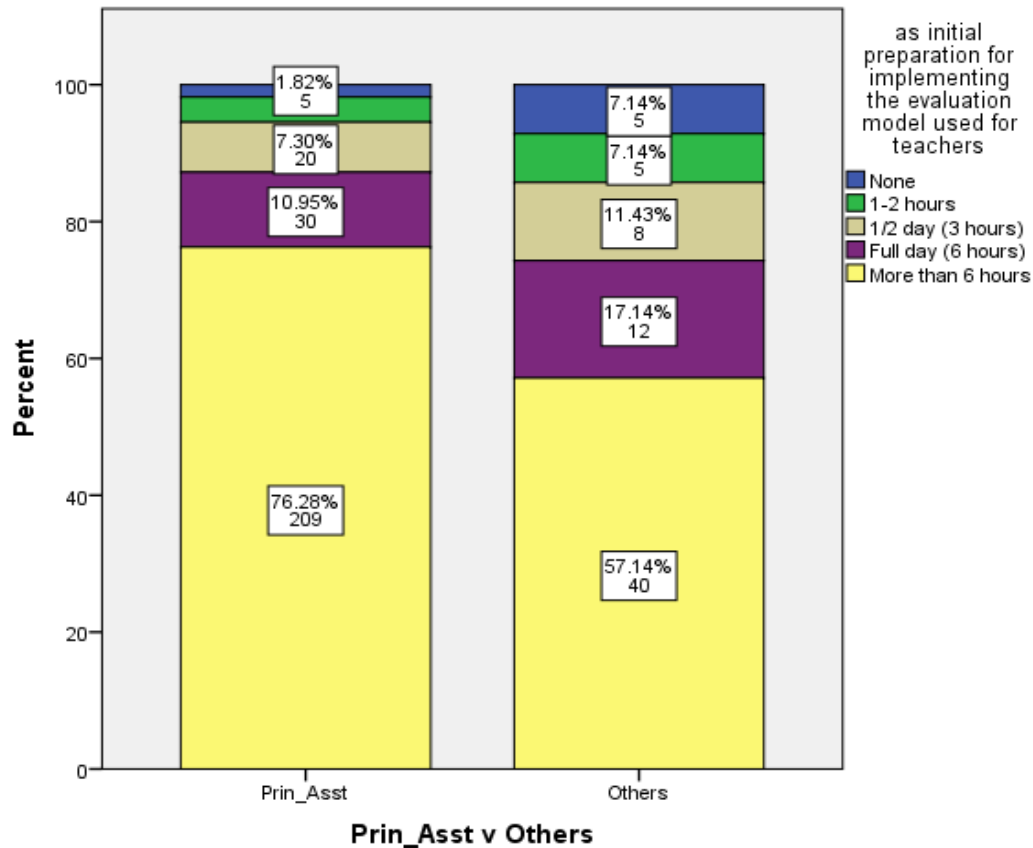


Figure 19: Principal / assistant principal contrasted with other administrator training – initial preparation

September 30, 2017

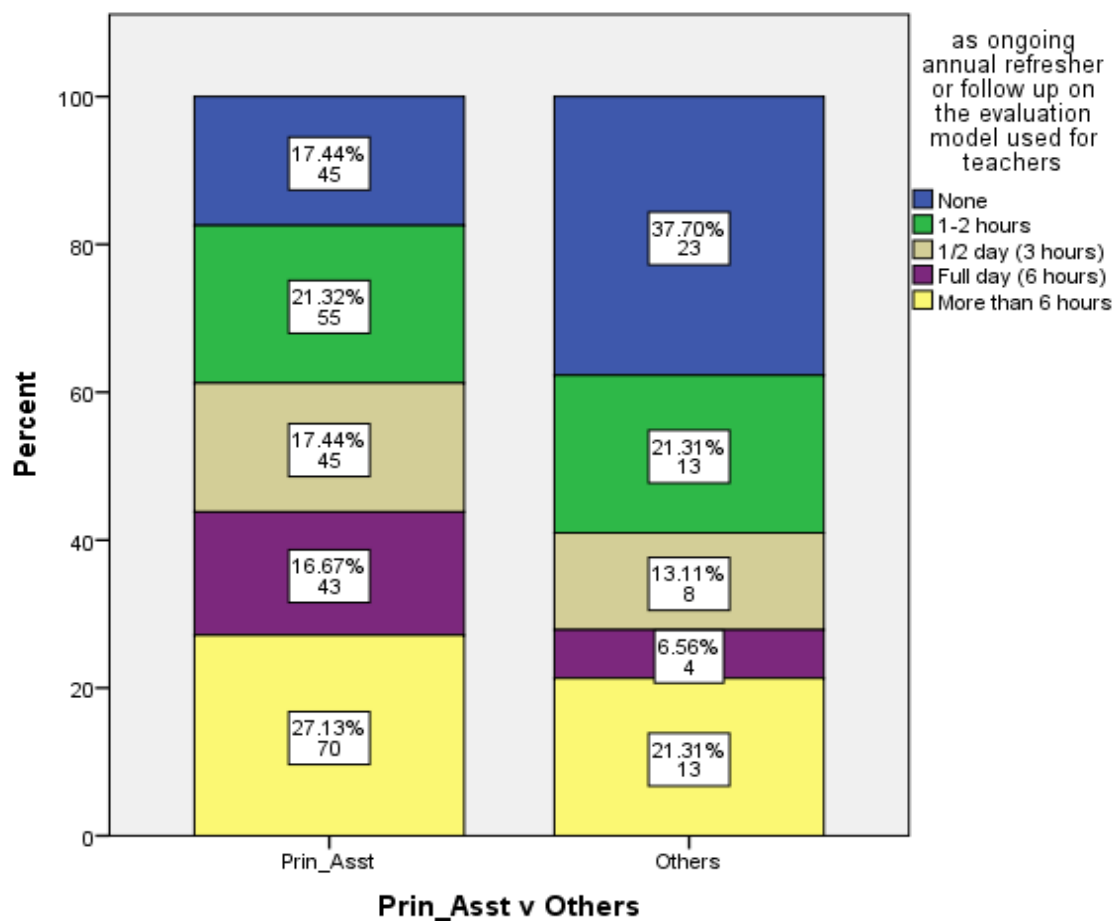


Figure 20: Principal / assistant principal contrasted with other administrator training – ongoing annual refresher training

September 30, 2017

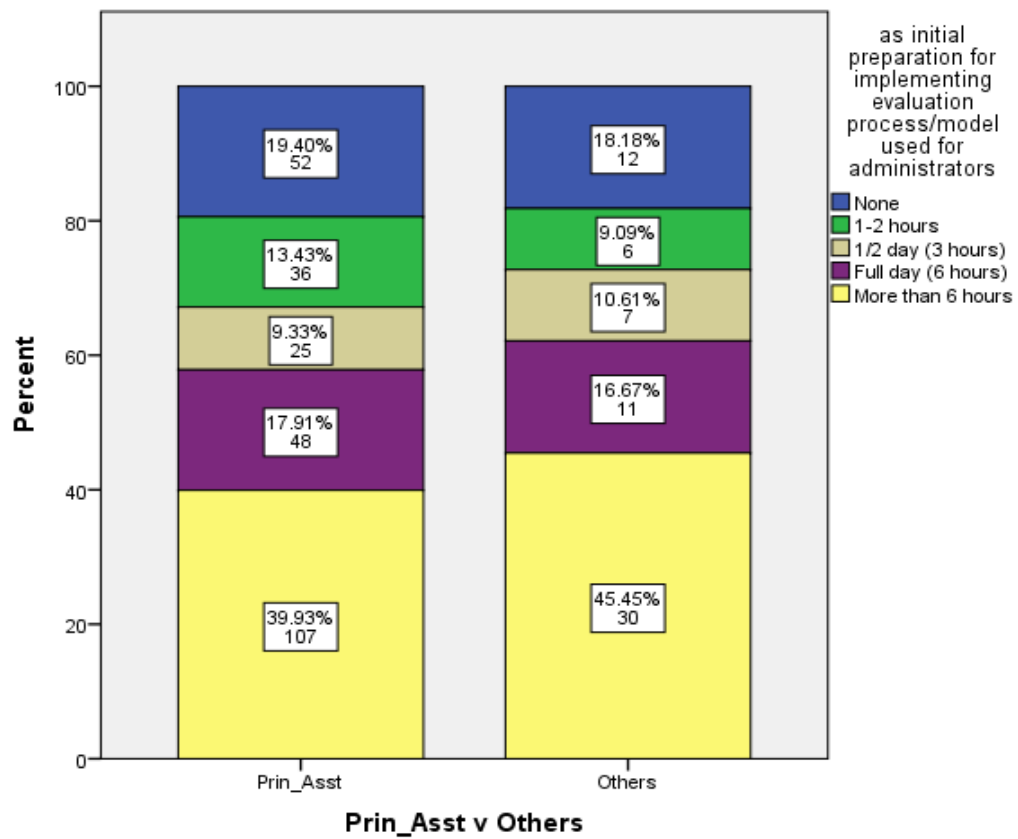


Figure 21: Principal / assistant principal contrasted with other administrator training – initial preparation in models used for administrators



September 30, 2017

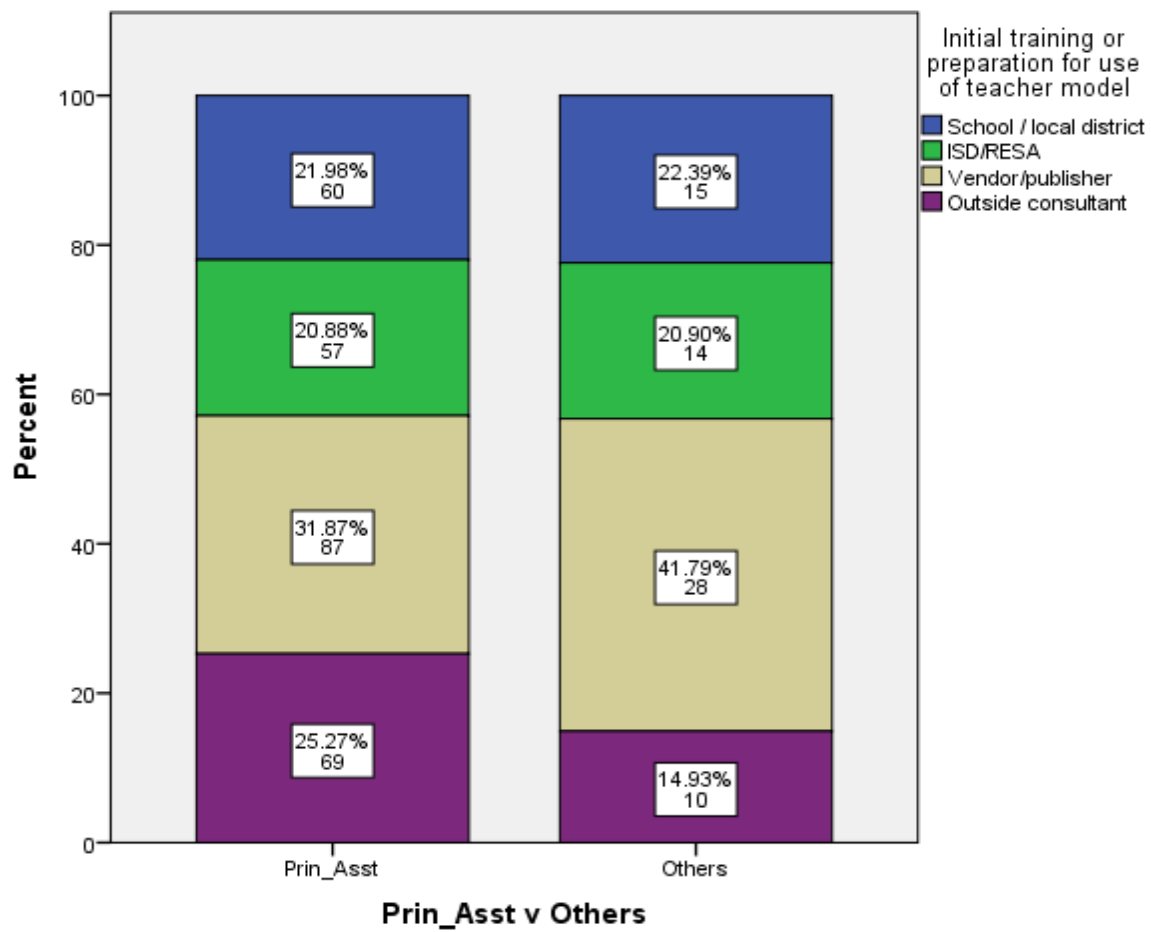


Figure 22: Principal /assistant principal contrasted with other administrator training – source of training for initial preparation for teacher model

September 30, 2017

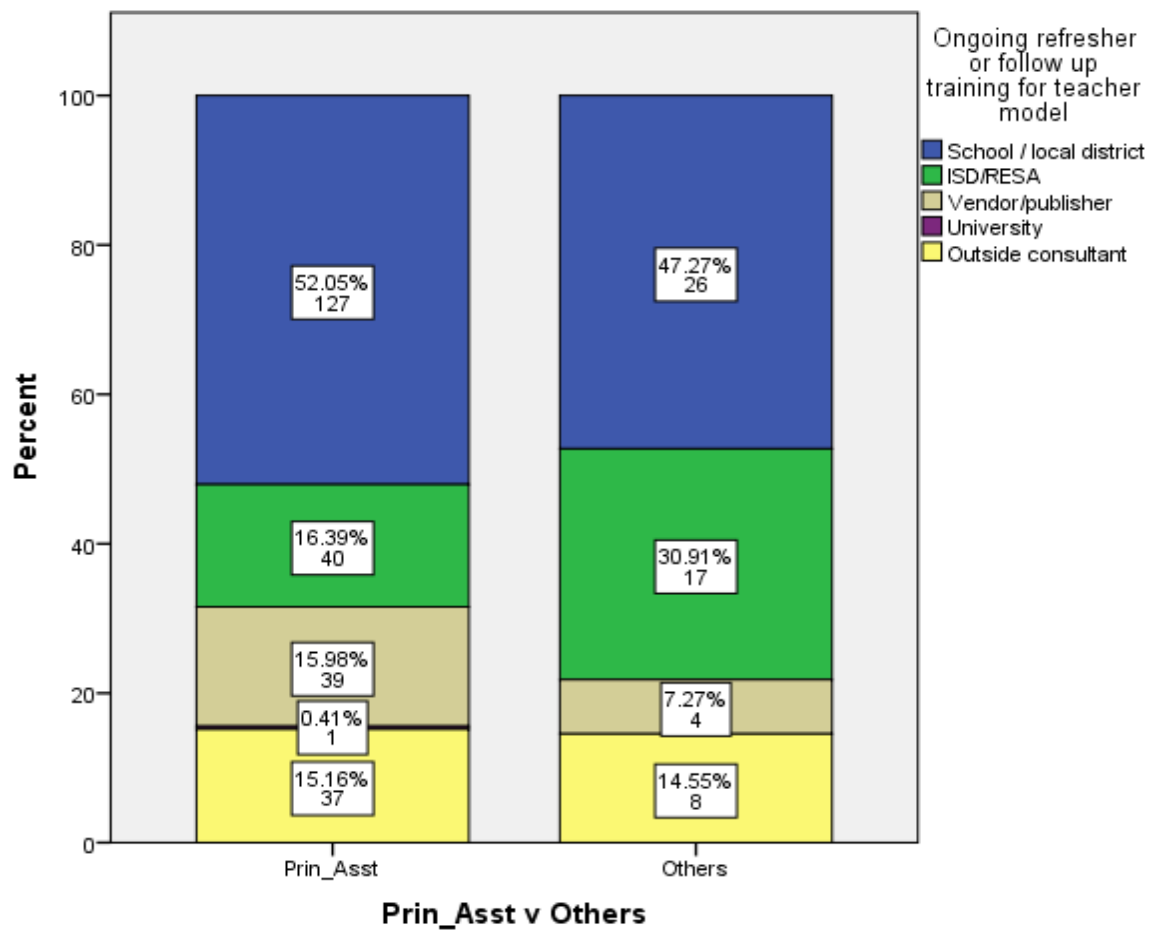


Figure 23: Principal /assistant principal contrasted with other administrator training – source of training for ongoing refresher training for teacher model

September 30, 2017

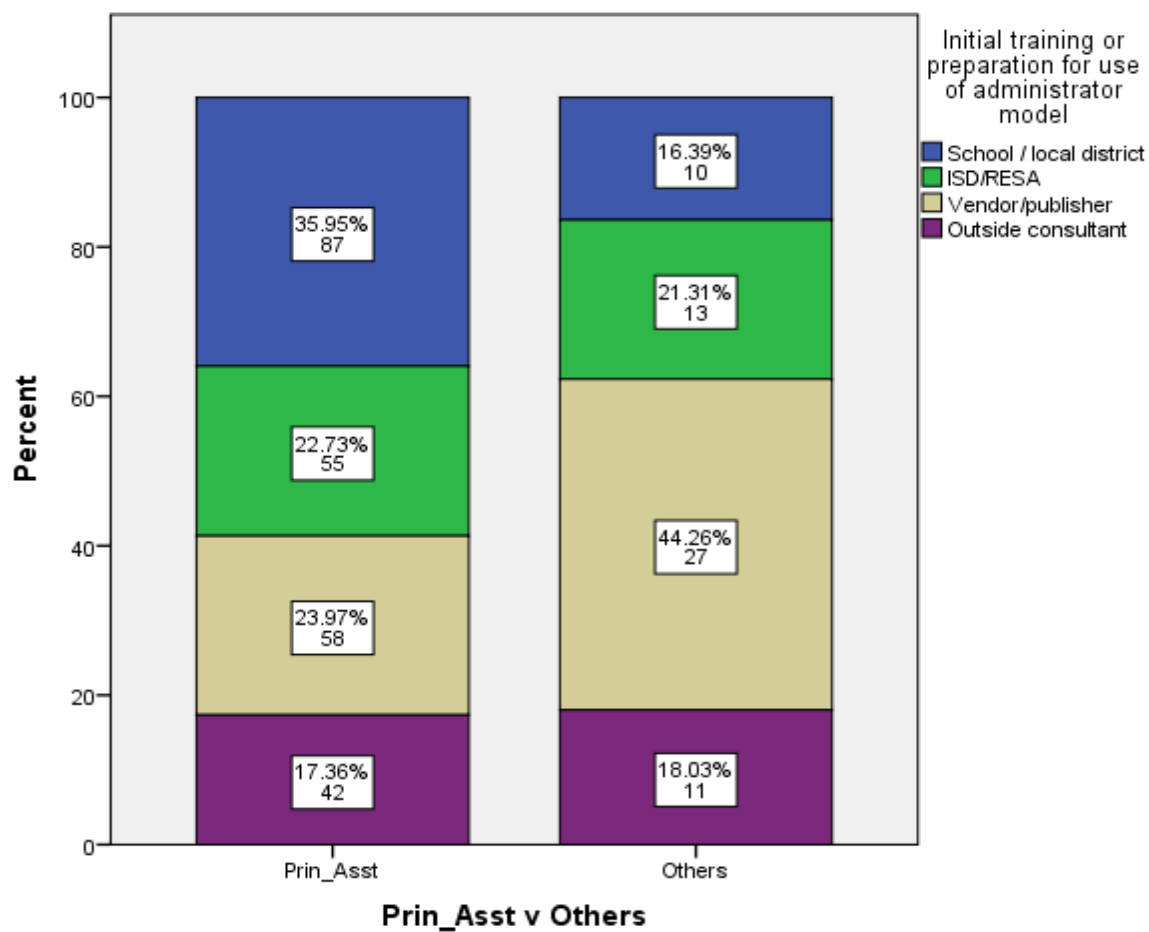


Figure 24: Principal /assistant principal contrasted with other administrator training – source of training for use of administrator model

September 30, 2017

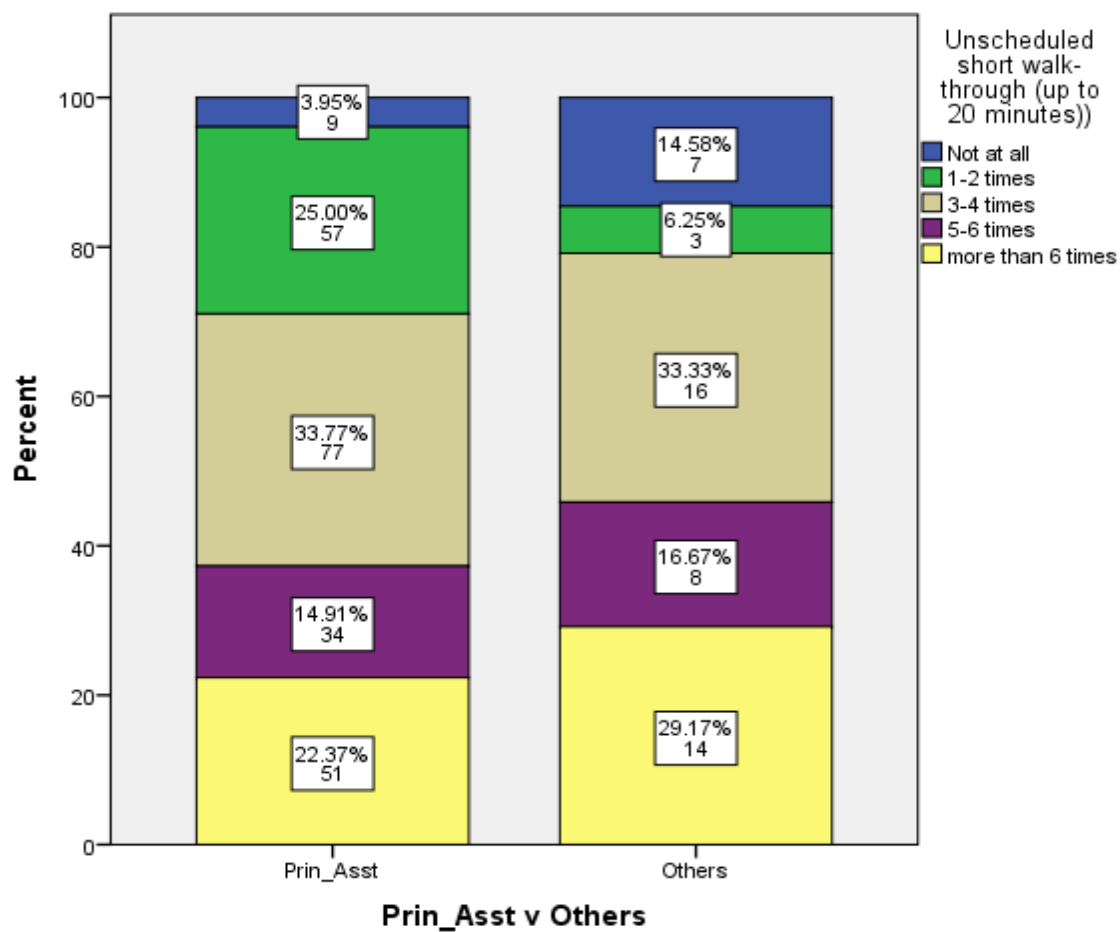


Figure 25: Principal /assistant principal contrasted with other administrator – unscheduled short walk-through

September 30, 2017

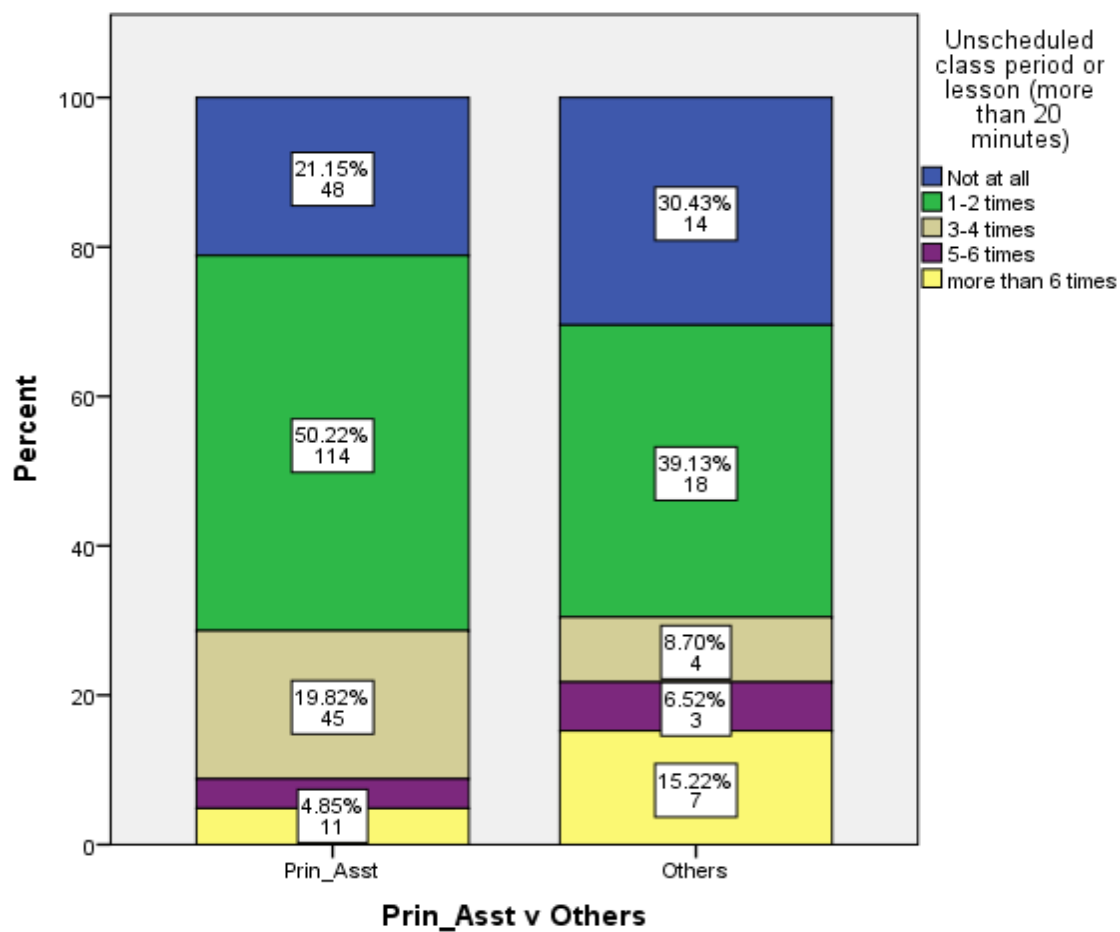


Figure 26: Principal /assistant principal contrasted with other administrator – unscheduled class period of lesson

September 30, 2017

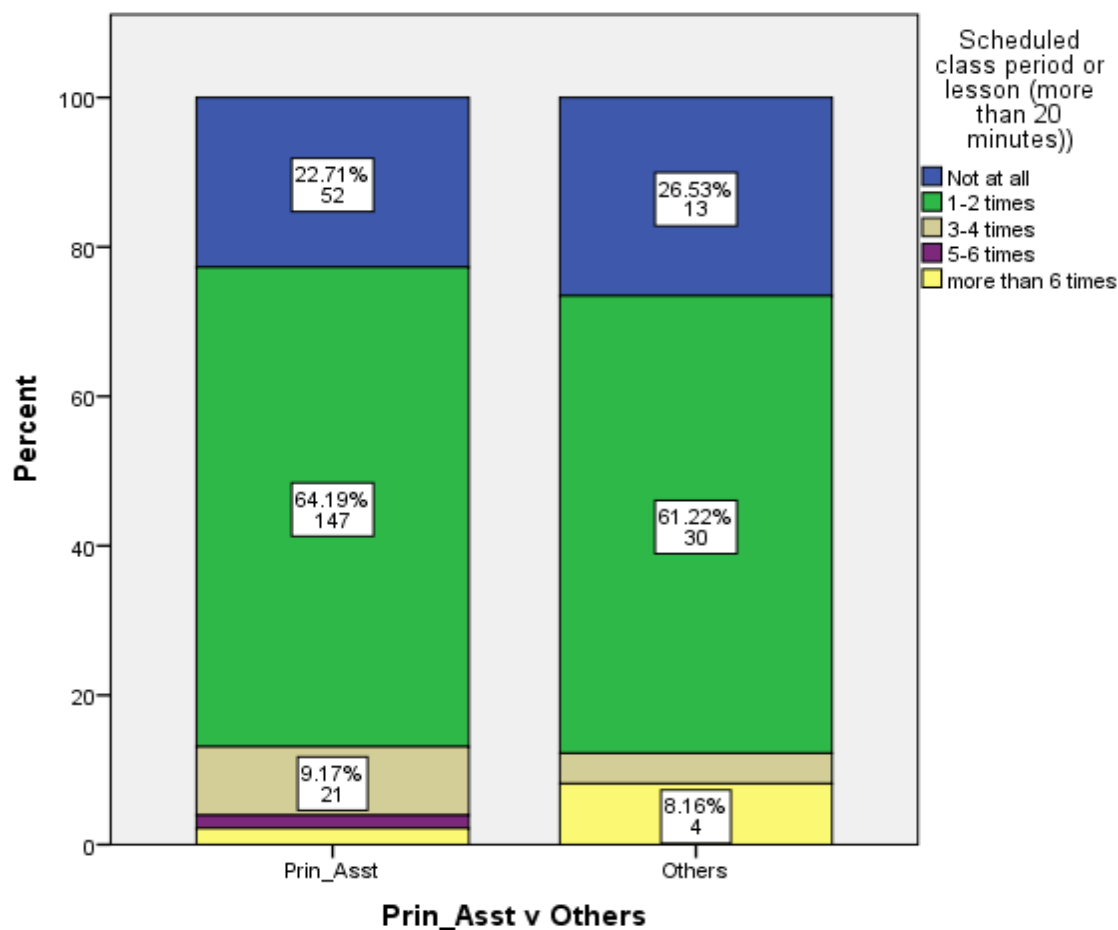
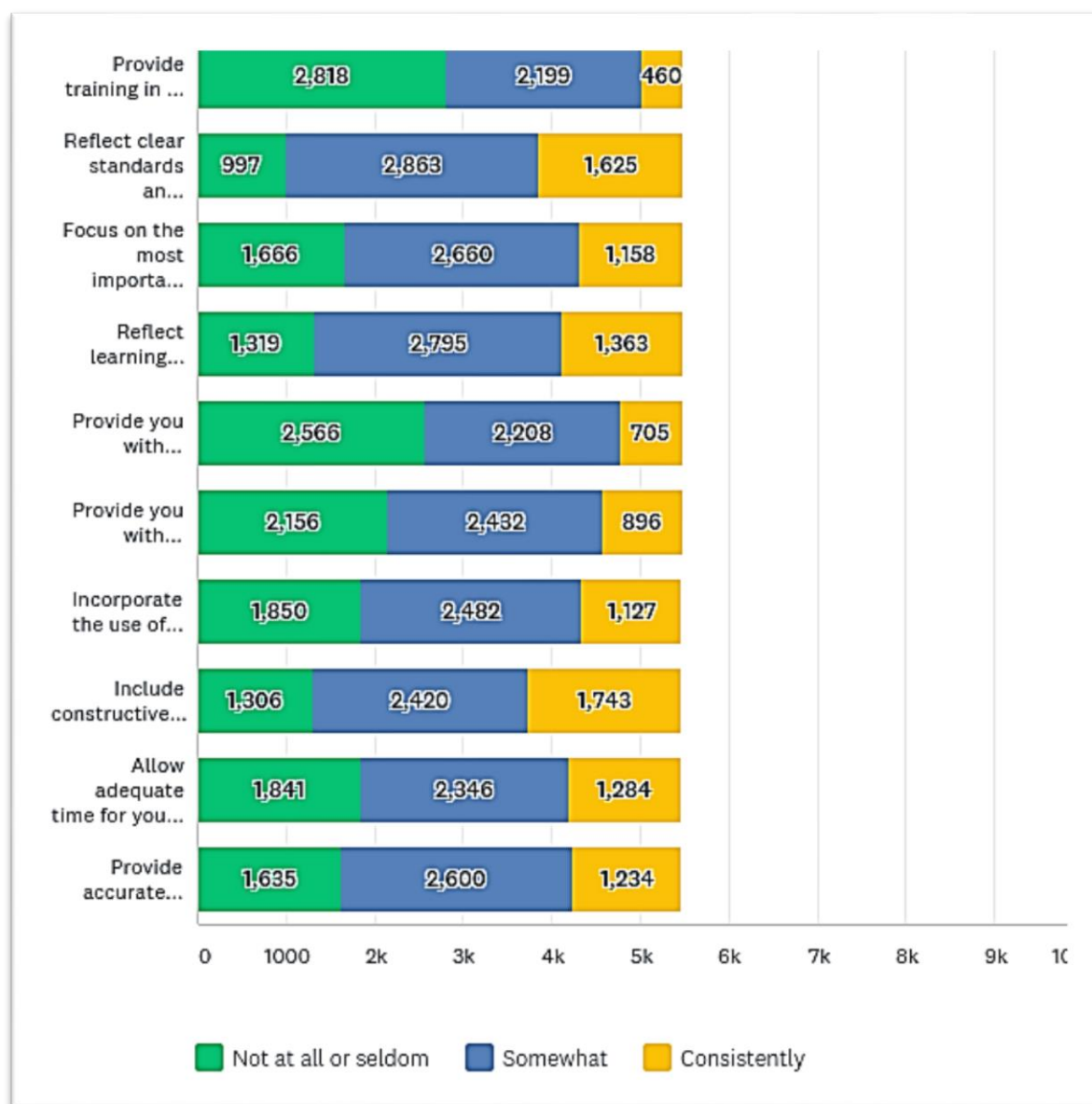


Figure 27: Principal /assistant principal contrasted with other administrator – scheduled class period or lesson

In survey questions 18 and 19 (teachers) and 50 and 51 (administrators) we asked questions about the use of best practices for educator evaluation. Survey respondents were asked to describe the extent 10 factors were used in the process of evaluation (Q 18 and 50) and to judge the factors either as *not at all* or *seldom*, *somewhat*, or *consistently* used. They were next asked the extent to which 9 factors occur in their evaluation process (Q 19 and 51). Charts summarizing the findings follow.

September 30, 2017

Figure 28: Teachers: Indications of Best Practices

**Q18: To what extent does your evaluation process (n = 5,505):**

September 30, 2017

	NOT AT ALL OR SELDOM	SOMEWHAT	CONSISTENTLY	TOTAL	WEIGHTED AVERAGE
Provide training in the protocols	51.45% 2,818	40.15% 2,199	8.40% 460	5,477	1.57
Reflect clear standards and expectations	18.18% 997	52.20% 2,863	29.63% 1,625	5,485	2.11
Focus on the most important areas of your professional practice	30.38% 1,666	48.50% 2,660	21.12% 1,158	5,484	1.91
Reflect learning outcomes of your students	24.08% 1,319	51.03% 2,795	24.89% 1,363	5,477	2.01
Provide you with information you need to increase your students' learning	46.83% 2,566	40.30% 2,208	12.87% 705	5,479	1.66
Provide you with information you need for your professional growth	39.31% 2,156	44.35% 2,432	16.34% 896	5,484	1.77
Incorporate the use of multiple data sources	33.89% 1,850	45.47% 2,482	20.64% 1,127	5,459	1.87
Include constructive feedback from your evaluator	23.88% 1,306	44.25% 2,420	31.87% 1,743	5,469	2.08
Allow adequate time for your evaluation process to be implemented as prescribed	33.65% 1,841	42.88% 2,346	23.47% 1,284	5,471	1.90
Provide accurate ratings of your practice and performance	29.90% 1,635	47.54% 2,600	22.56% 1,234	5,469	1.93

We were interested if there is an observable difference in responses to the questions about best practices depending upon the evaluation model that is used. The following tables, figures 29 – 38, show the responses to question 18 sorted by the evaluation model used by the respondent. For example, in



September 30, 2017

the chart below when asked the extent the evaluation process provides training in the protocol respondents who reported using Danielson more frequently responded *not at all or seldom*.

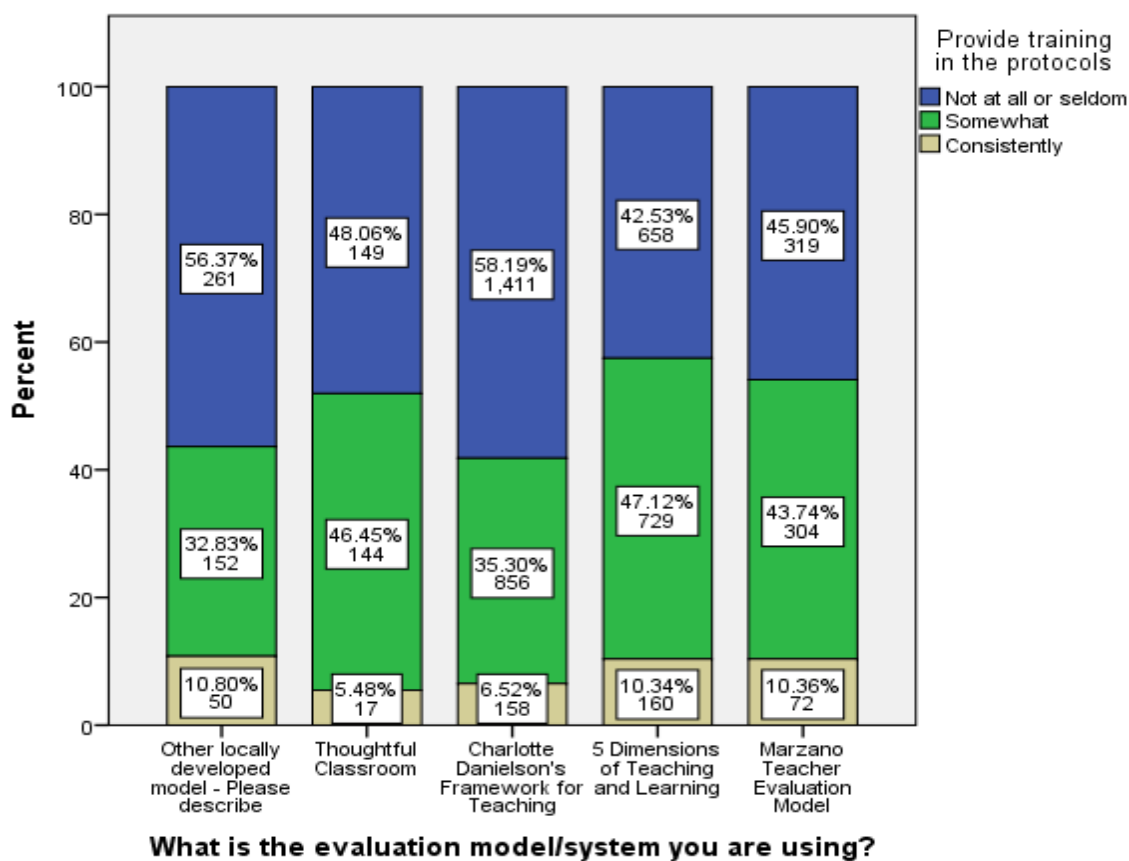


Figure 29: Q 8 x Q18 – Teachers: Model provides training in the protocol sorted by model used (n= 5440)

September 30, 2017

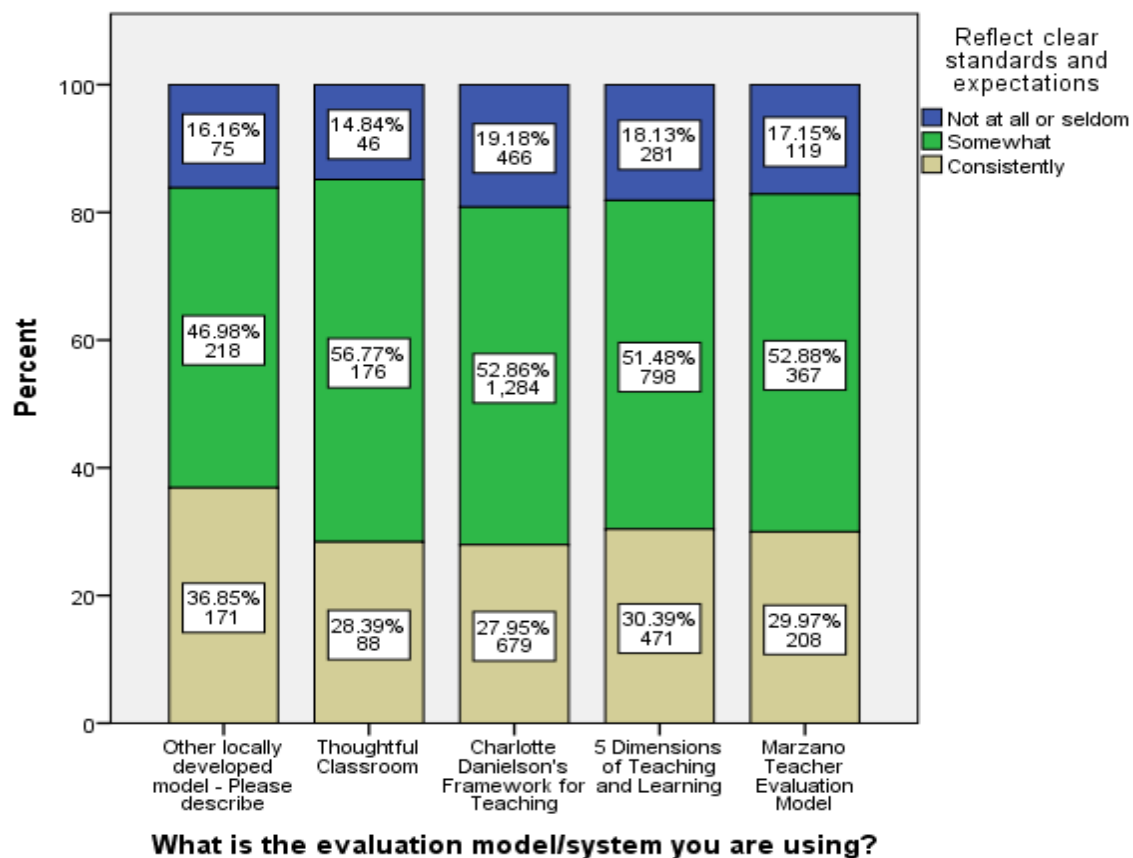


Figure 30: Q 8 x Q18 – Teachers: Reflection of clear standards sorted by model used (n= 5440)

September 30, 2017

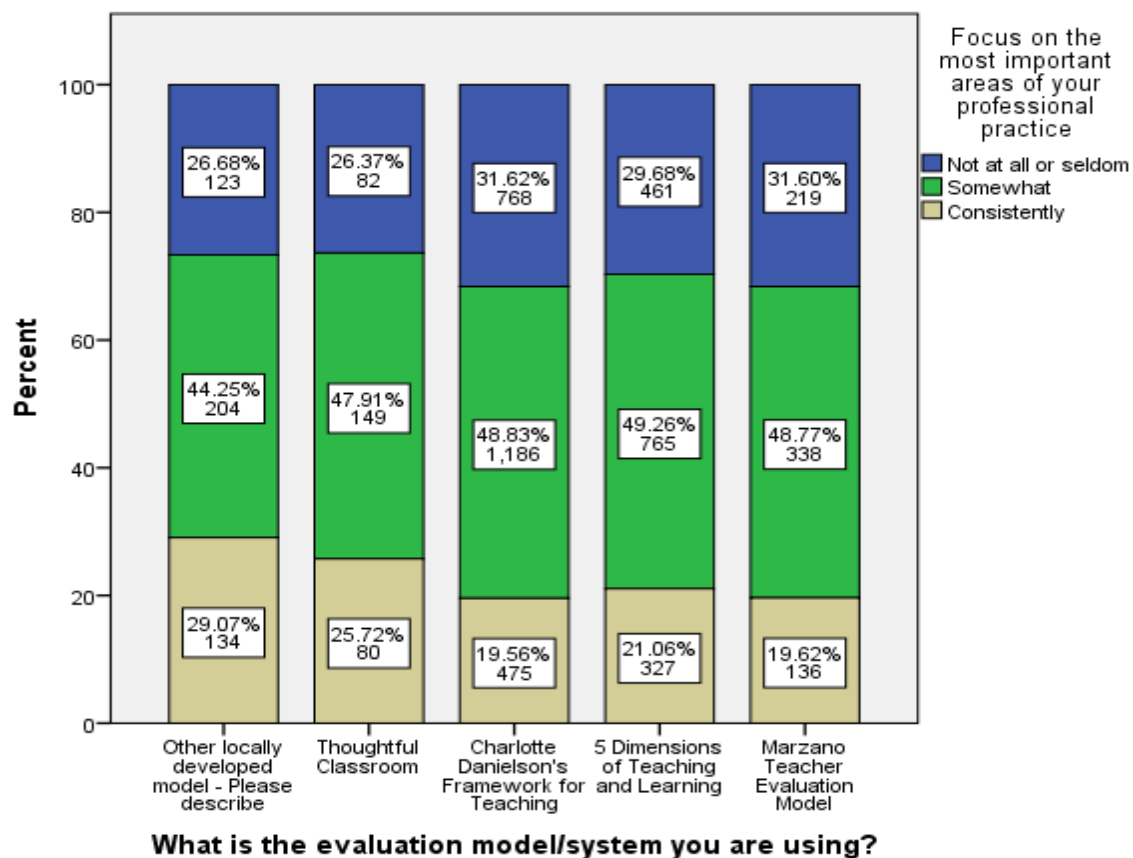


Figure 31: Q 8 x Q18 – Teachers: Focus on most important areas of professional practice sorted by model used (n= 5440)

September 30, 2017

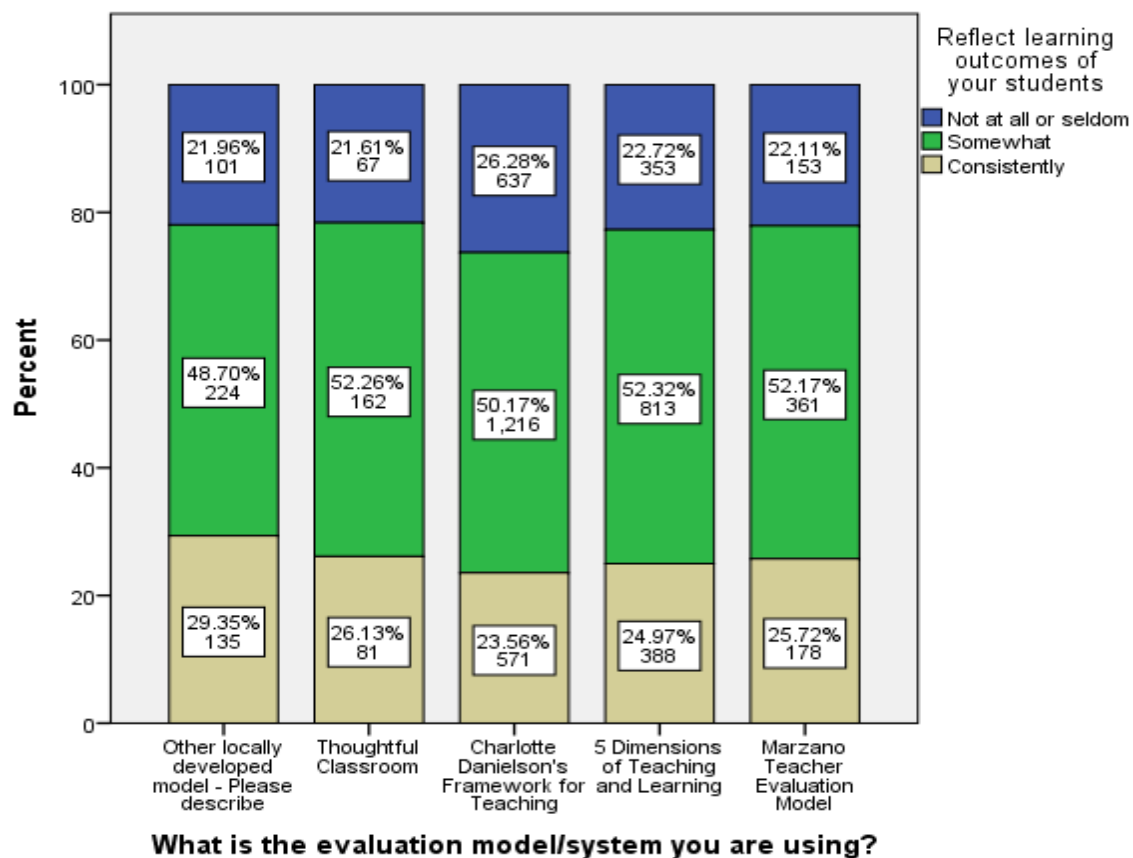


Figure 32: Q 8 x Q18 – Teachers: Reflect learning outcomes of your students sorted by model used (n= 5440)

September 30, 2017

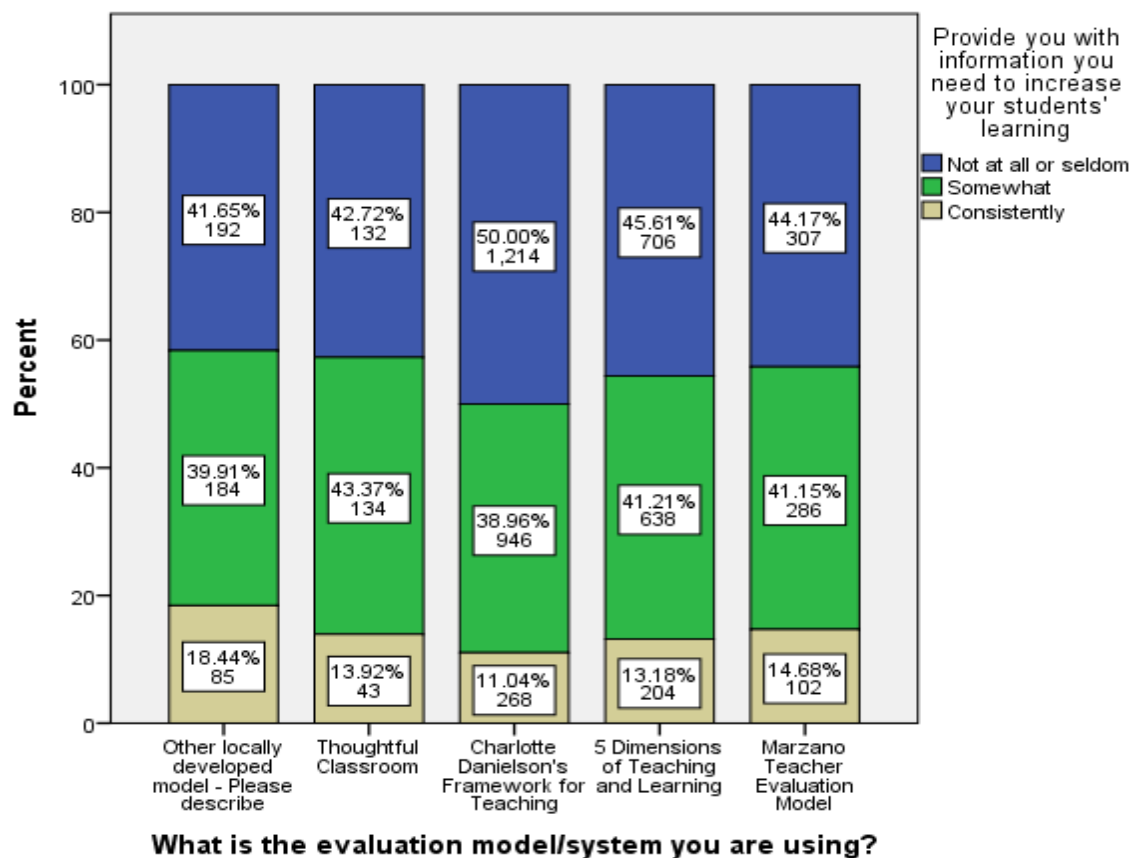


Figure 33: Q 8 x Q18 – Teachers: Provides you with information you need to increase your students' learning sorted by model used (n= 5440)

September 30, 2017

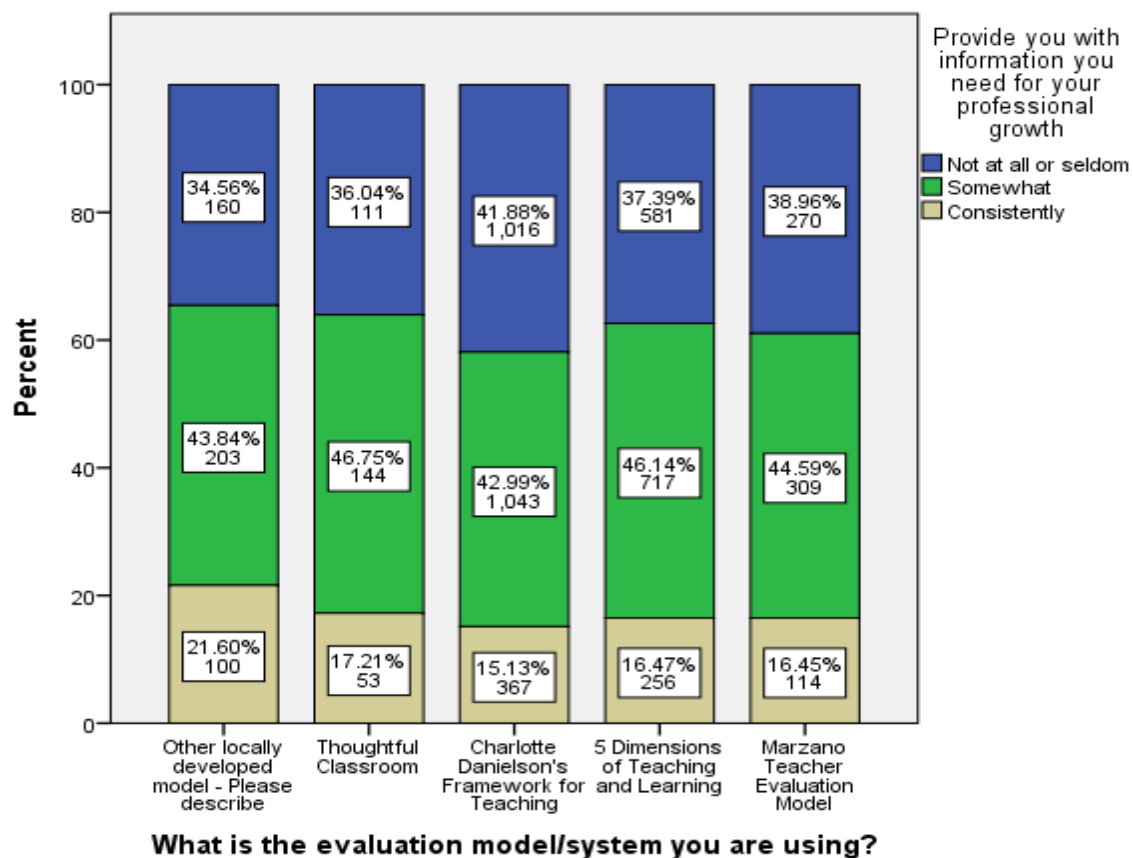


Figure 34: Q 8 x Q18 – Teachers: Provides you with information you need for your professional growth sorted by model used (n = 5440)

September 30, 2017

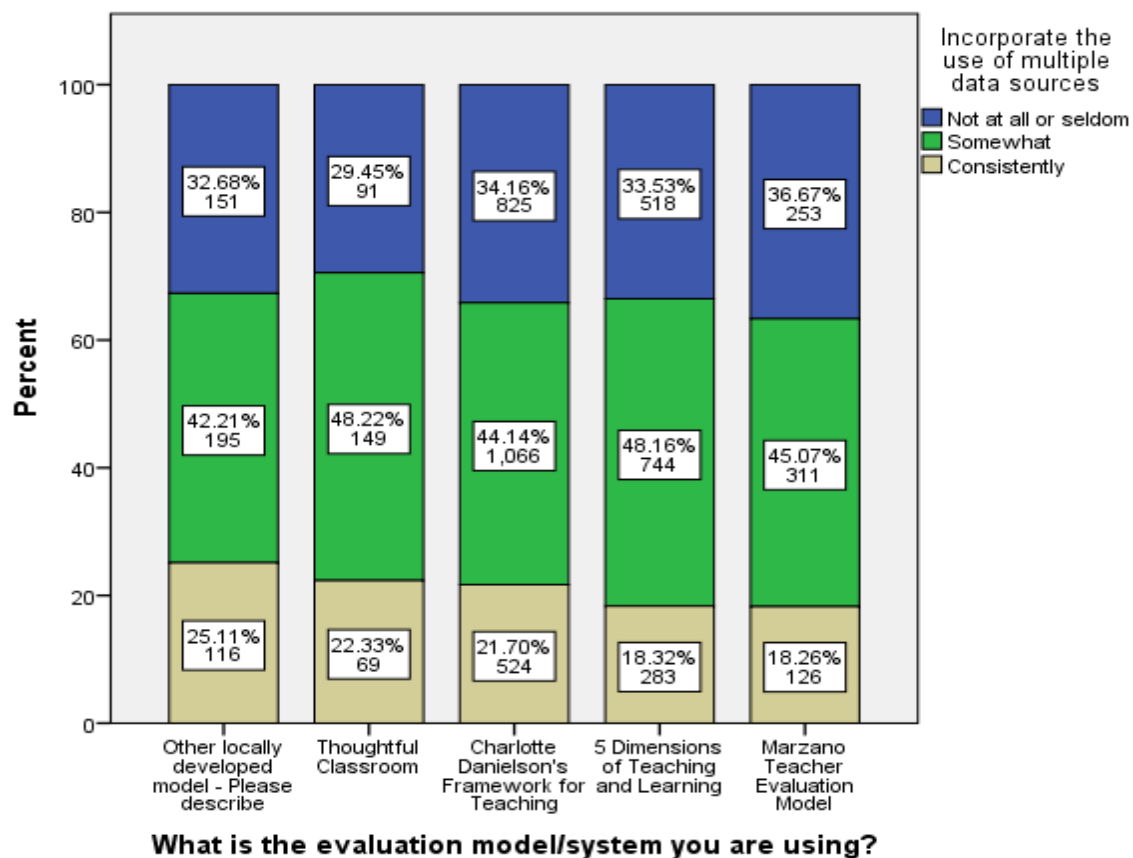


Figure 35: Q 8 x Q18 – Teachers: Incorporate the use of multiple data sources sorted by model used (n = 5440)

September 30, 2017

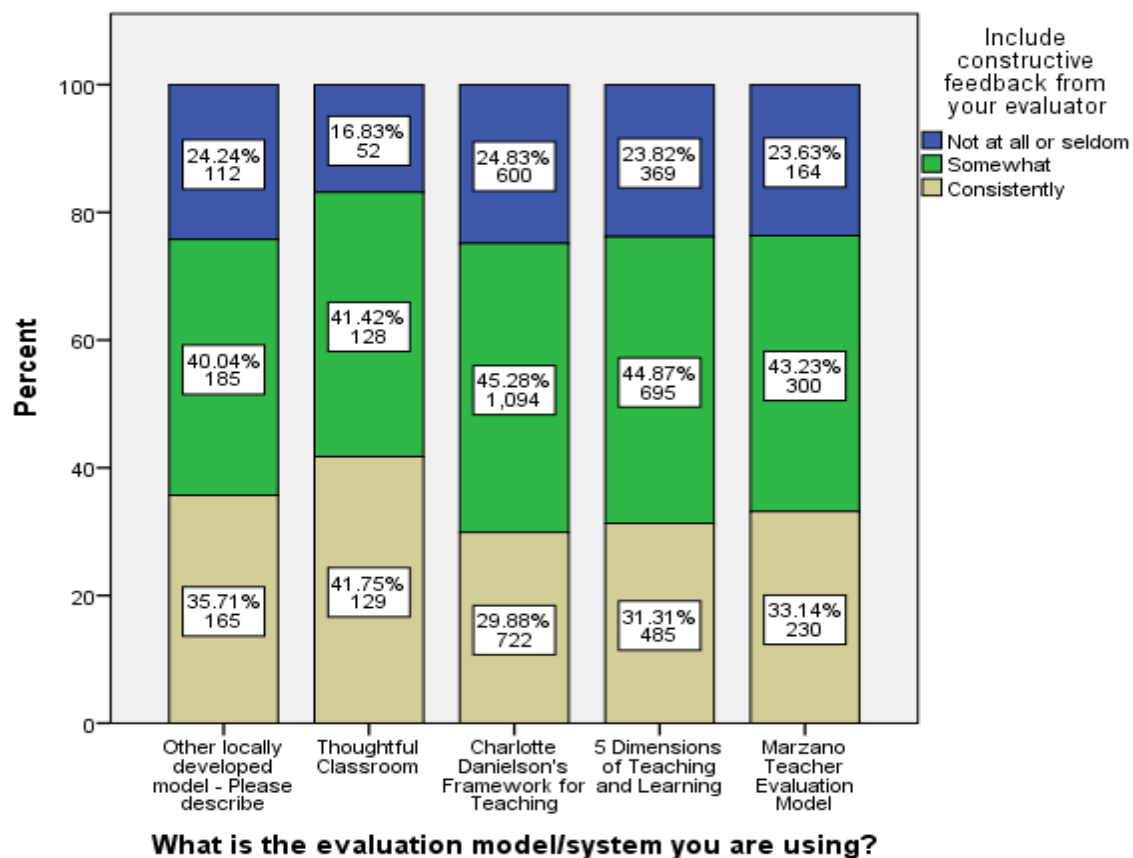


Figure 36: Q 8 x Q18 – Teachers: Include constructive feedback from your evaluator sorted by model used (n = 5440)



September 30, 2017

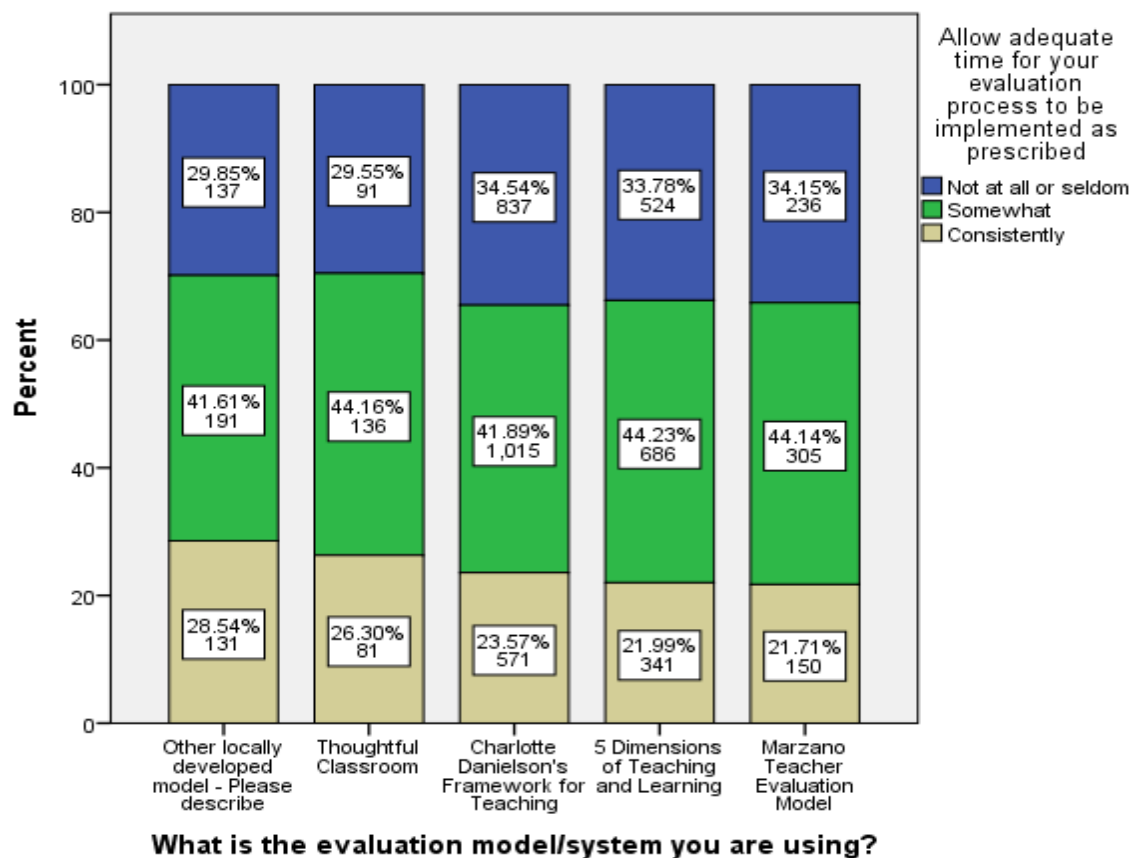


Figure 37: Q 8 x Q18 – Teachers: Allow for adequate time for your evaluation process to be implemented as prescribed sorted by model used (n = 5440)

September 30, 2017

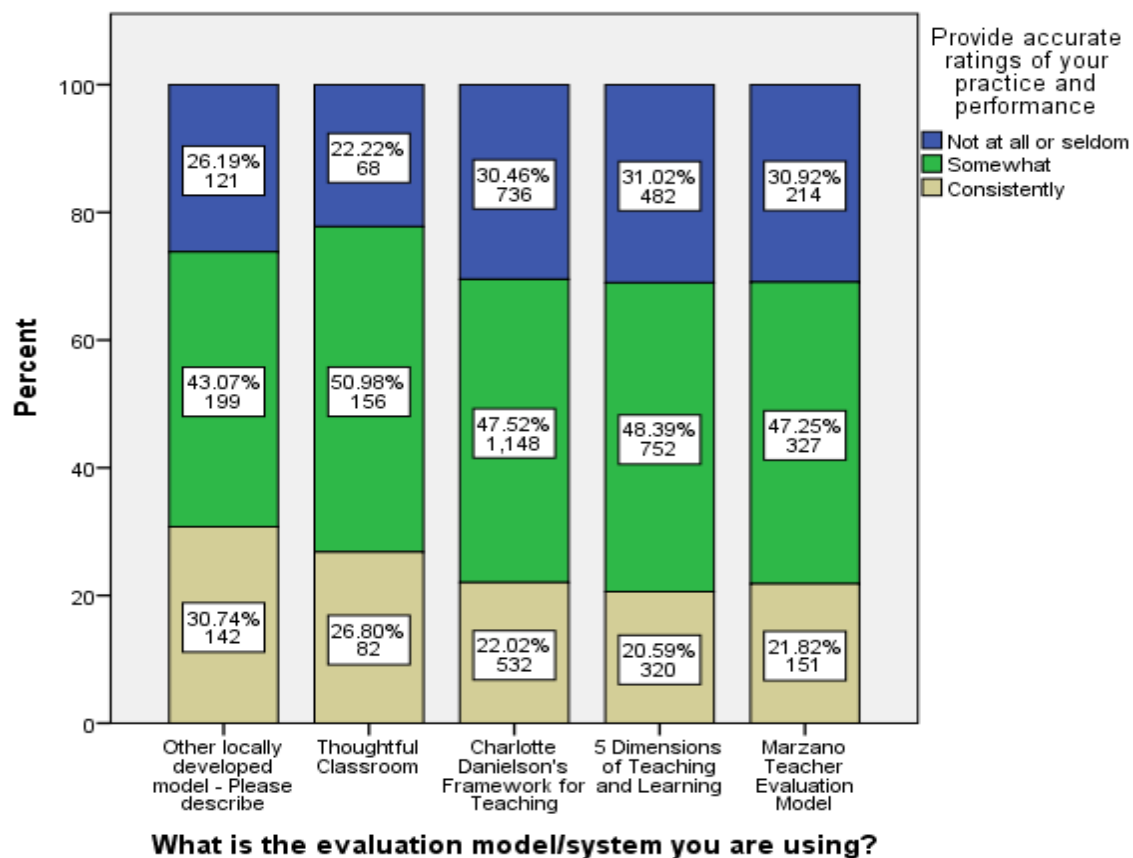


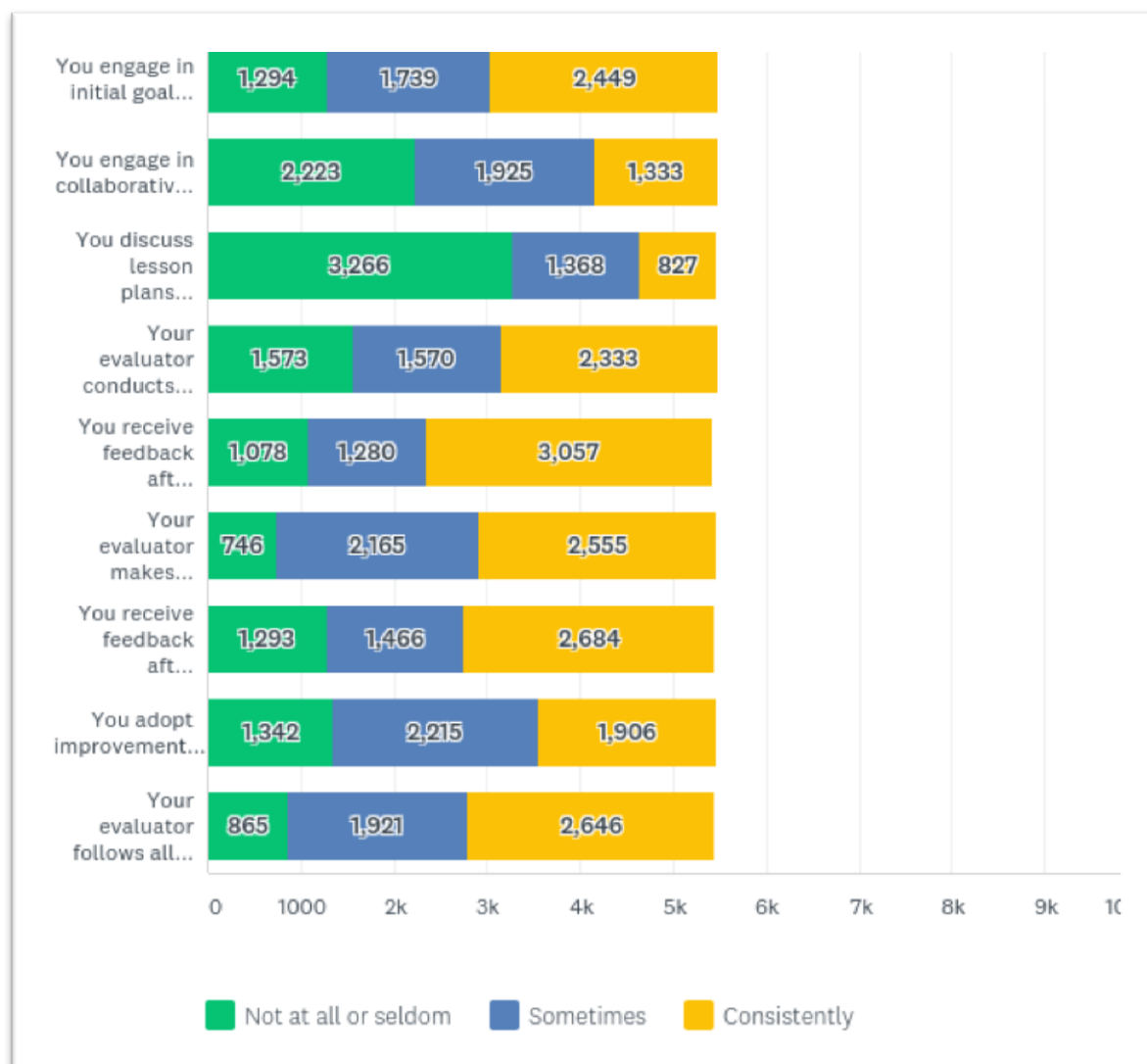
Figure 38: Q 8 x Q18 – Teachers: Provide accurate ratings of your practice and performance sorted by model used (n = 5440)

September 30, 2017

In survey question 19 we asked teachers, and later in question 51 we asked administrators, about processes associated with best practices for educator evaluation.

Figure 39: Teachers: Characteristics of the evaluation process

**Q19: To what extent do each of the following occur in your evaluation process? (n=5,488)**



September 30, 2017

	NOT AT ALL OR SELDOM	SOMETIMES	CONSISTENTLY	TOTAL	WEIGHTED AVERAGE
You engage in initial goal setting with your evaluator	23.60% 1,294	31.72% 1,739	44.67% 2,449	5,482	2.21
You engage in collaborative goal setting with your colleagues	40.56% 2,223	35.12% 1,925	24.32% 1,333	5,481	1.84
You discuss lesson plans with your evaluator before each scheduled observation	59.81% 3,266	25.05% 1,368	15.14% 827	5,461	1.55
Your evaluator conducts scheduled observations	28.73% 1,573	28.67% 1,570	42.60% 2,333	5,476	2.14
You receive feedback after each scheduled observation	19.91% 1,078	23.64% 1,280	56.45% 3,057	5,415	2.37
Your evaluator makes unscheduled observations	13.65% 746	39.61% 2,165	46.74% 2,555	5,466	2.33
You receive feedback after each unscheduled observation	23.76% 1,293	26.93% 1,466	49.31% 2,684	5,443	2.26
You adopt improvement strategies and recommendations as discussed with your evaluator	24.57% 1,342	40.55% 2,215	34.89% 1,906	5,463	2.10
Your evaluator follows all procedures and elements prescribed by the evaluation system	15.92% 865	35.36% 1,921	48.71% 2,646	5,432	2.33

September 30, 2017

The following series of charts (figures 40 -48) unpacks the best practices and associates responses to survey question 19 to the model used by the respondent.

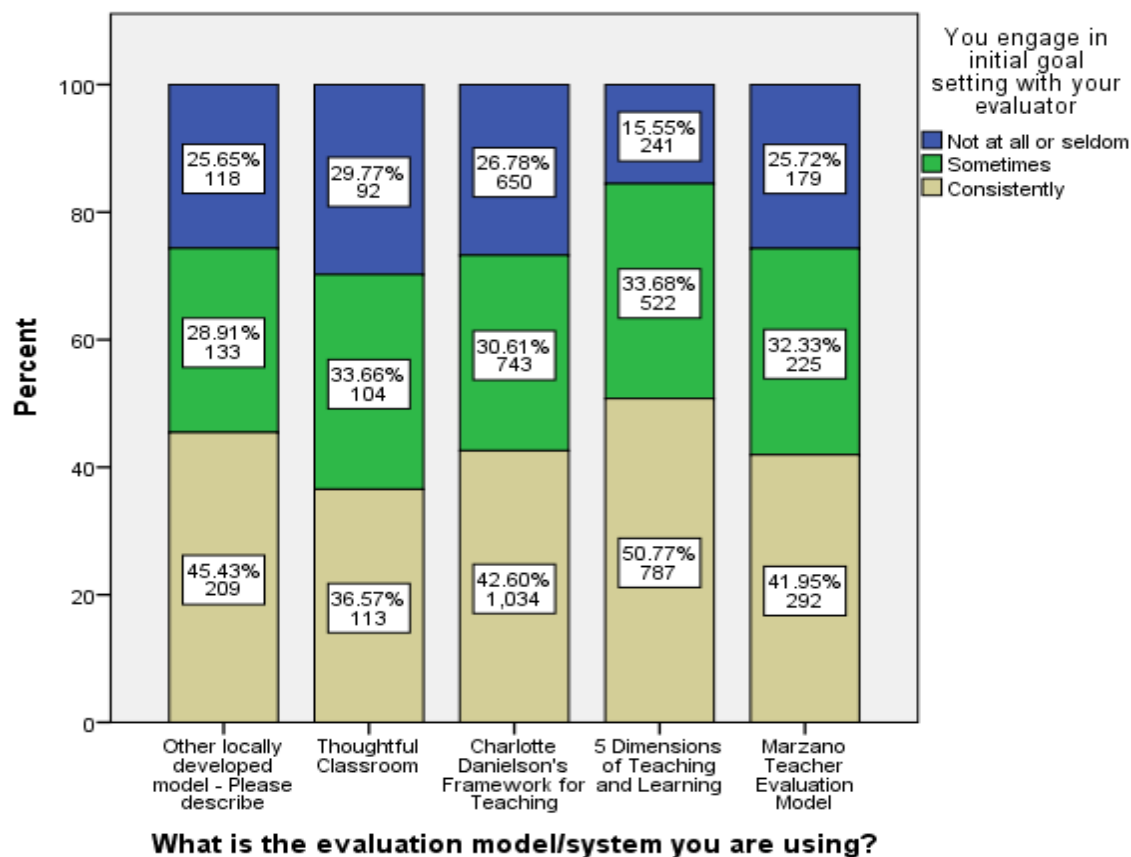


Figure 40: Q 8 x Q19 – Teachers: You engage in initial goal setting with your evaluator sorted by model used (n= 5430)

September 30, 2017

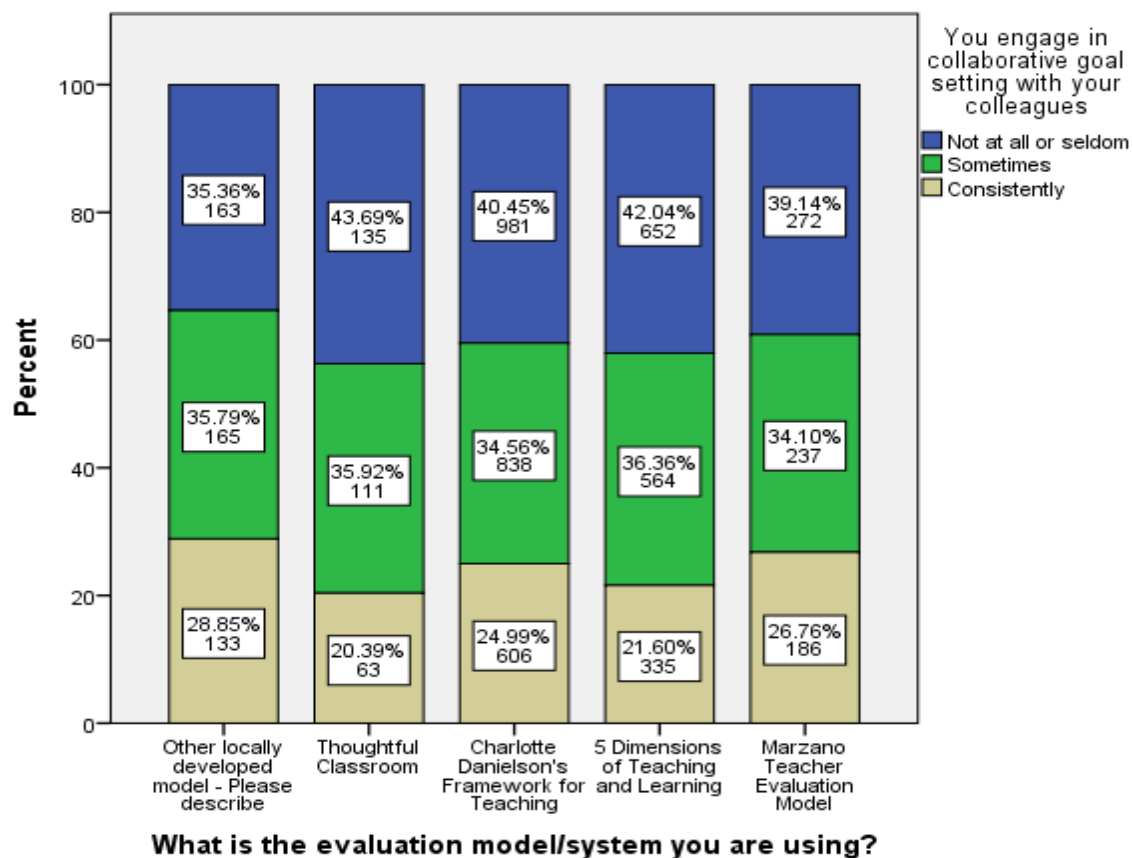


Figure 41: Q 8 x Q19 – Teachers: You engage in collaborative goal setting with your colleagues sorted by model used (n = 5430)

September 30, 2017

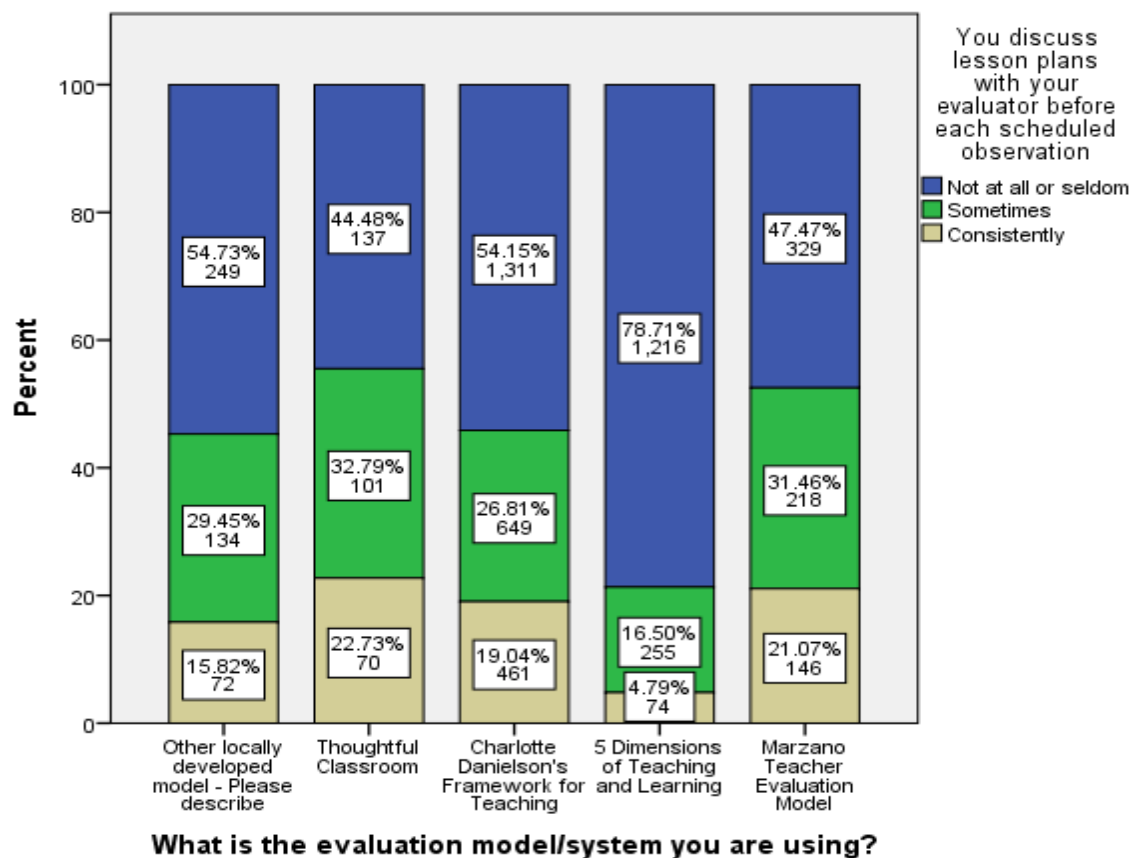


Figure 42: Q 8 x Q19 – Teachers: You discuss lesson plans with your evaluator before each scheduled observation sorted by model used (n = 5430)

September 30, 2017

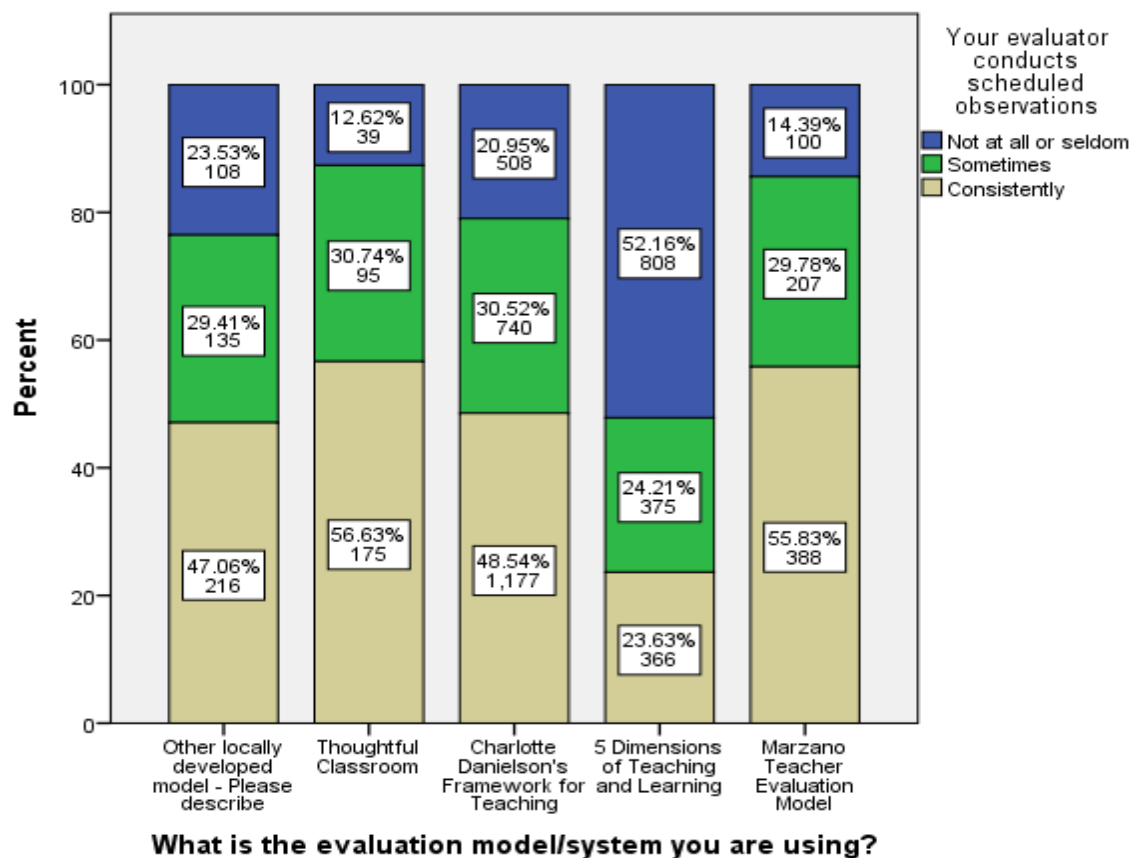


Figure 43: Q 8 x Q19 – Teachers: Your evaluator conducts scheduled observations sorted by model used (n = 5430)



September 30, 2017

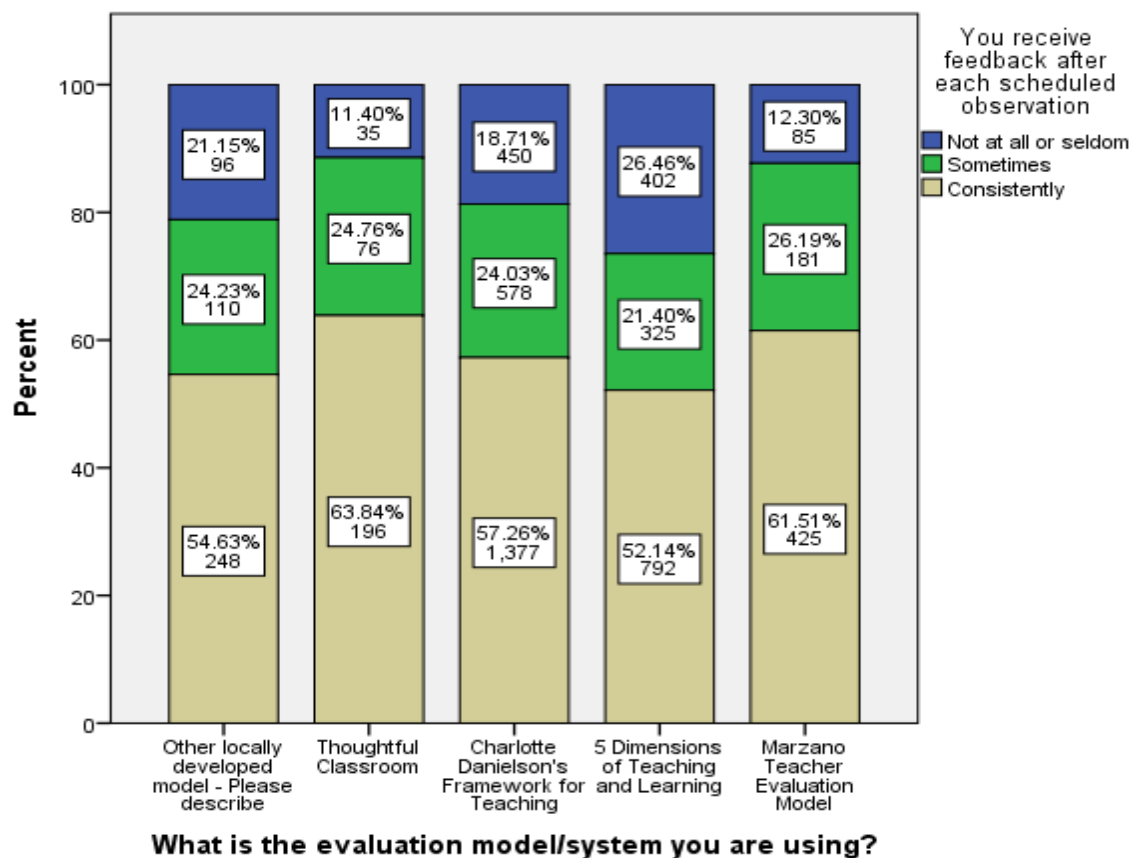


Figure 44: Q 8 x Q19 – Teachers: You receive feedback after scheduled observation sorted by model used (n = 5430)

September 30, 2017

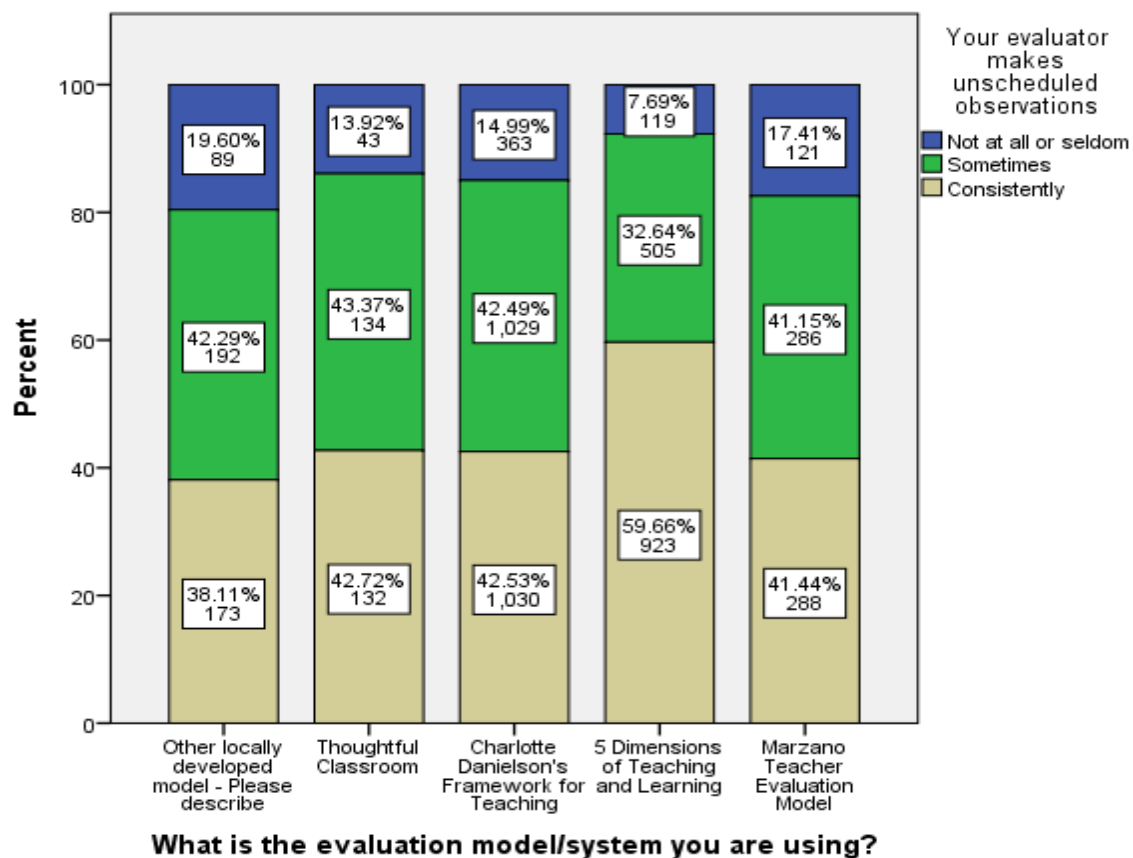


Figure 45: Q 8 x Q19 – Teachers: Your evaluator makes unscheduled observations sorted by model used (n = 5430)

September 30, 2017

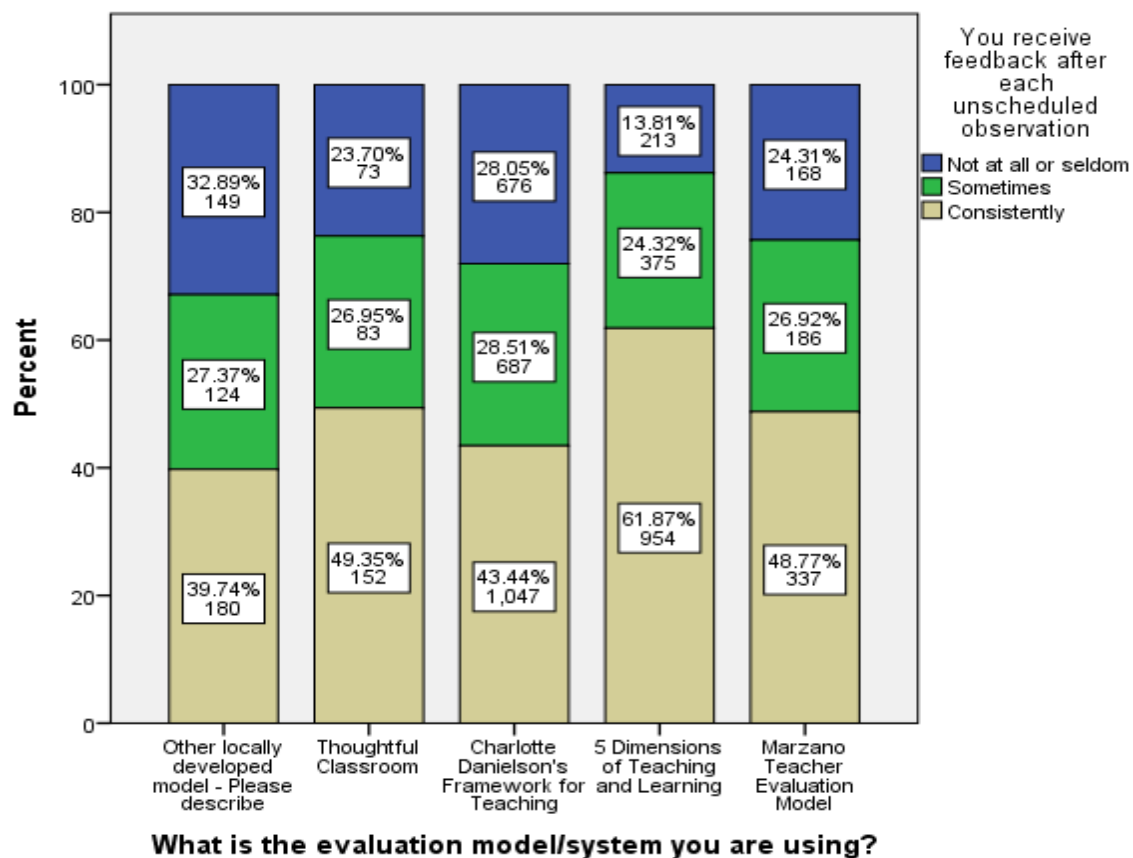


Figure 46: Q 8 x Q19 – Teachers: You receive feedback after each unscheduled observation sorted by model used (n = 5430)

September 30, 2017

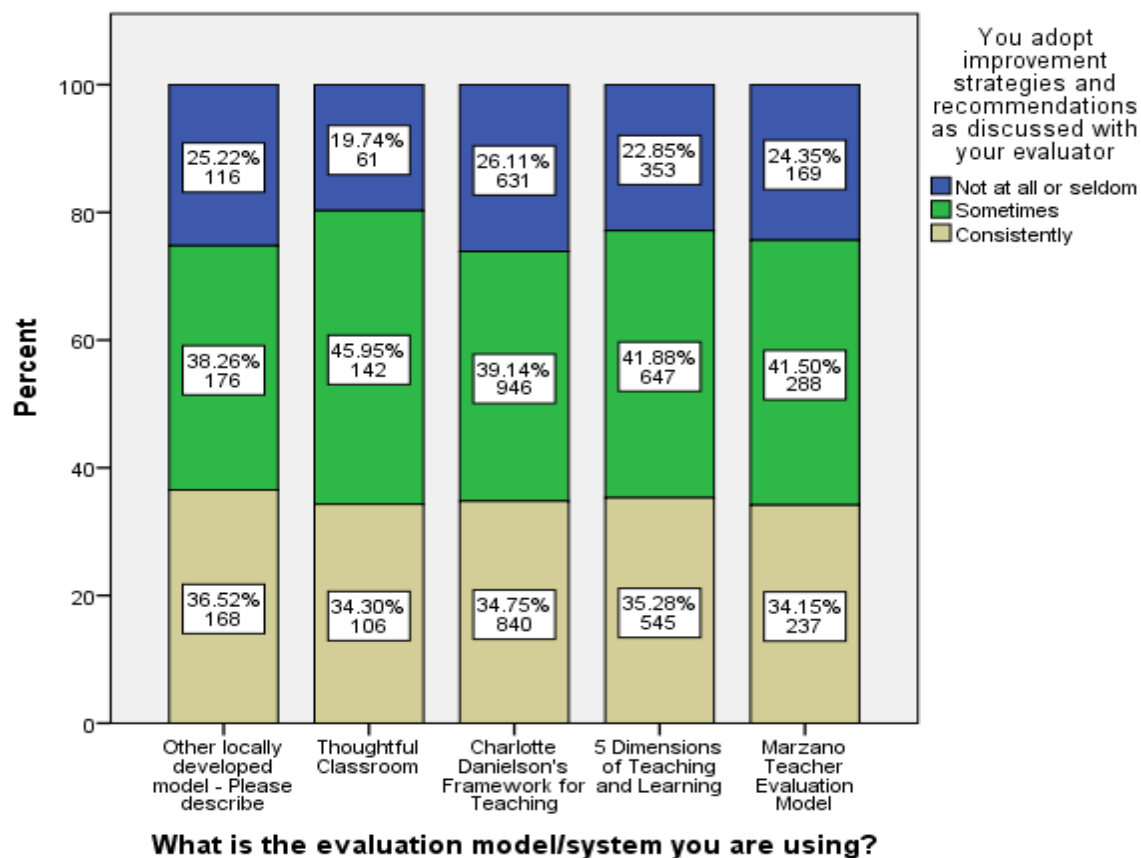


Figure 47: Q 8 x Q19 – Teachers: You adopt improvement strategies and recommendations as discussed with your evaluator sorted by model used (n = 5430)

September 30, 2017

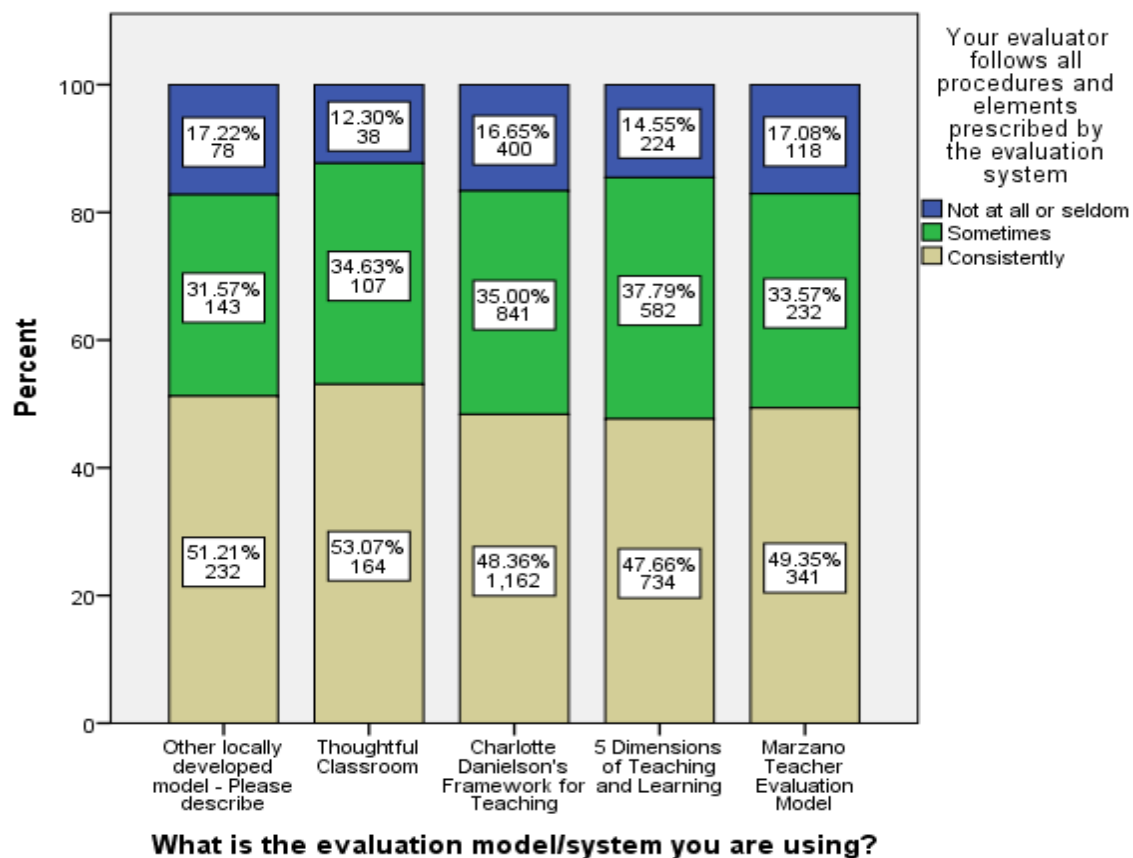
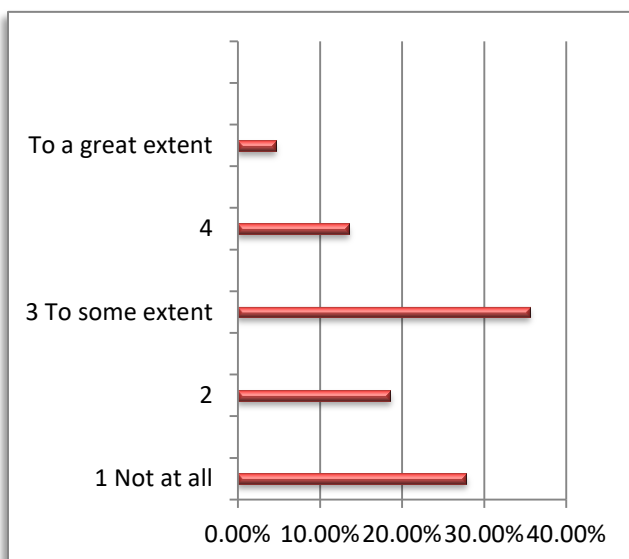


Figure 48: Q 8 x Q19 – Teachers: Your evaluator follows all procedures and elements prescribed by the evaluation system sorted by model used (n = 5430)

September 30, 2017

Figure 49: Q 20 – Teachers: To what extent are cultural competency skills included in the evaluation model you are using?



### Type and frequency of classroom observations visits

Best practice literature describes the importance of classroom observation visits. In some cases the primacy of scheduled long visits is described, preceded by a meeting between the teacher and evaluator to review the planned lesson and instructional strategies. Others cite the value of frequent short unscheduled visits. The Michigan legislation requires a minimum of two observations per year, with at least one of the two unscheduled<sup>20</sup>. We asked survey respondents, *How frequently were you observed in your evaluation this year?*

For each type of observation method – unscheduled short (less than 20 minutes), longer unscheduled, or scheduled (longer than 20 minutes) teachers most

### Best Practices in Action

#### Dearborn Public Schools

The district uses an open committee structure to engage in district-wide conversations about evaluation. The model allows for increased teacher commitment to the rubric and understanding its connection to classroom teaching. To improve inter-rater reliability, administrators are required to complete a challenging vendor-provided assessment. Ongoing calibration exercises are embedded in district administrator professional development throughout the school year.

Contact Person: Maysam Alie-Bazzi,  
Executive Director of Staff & Student  
Services

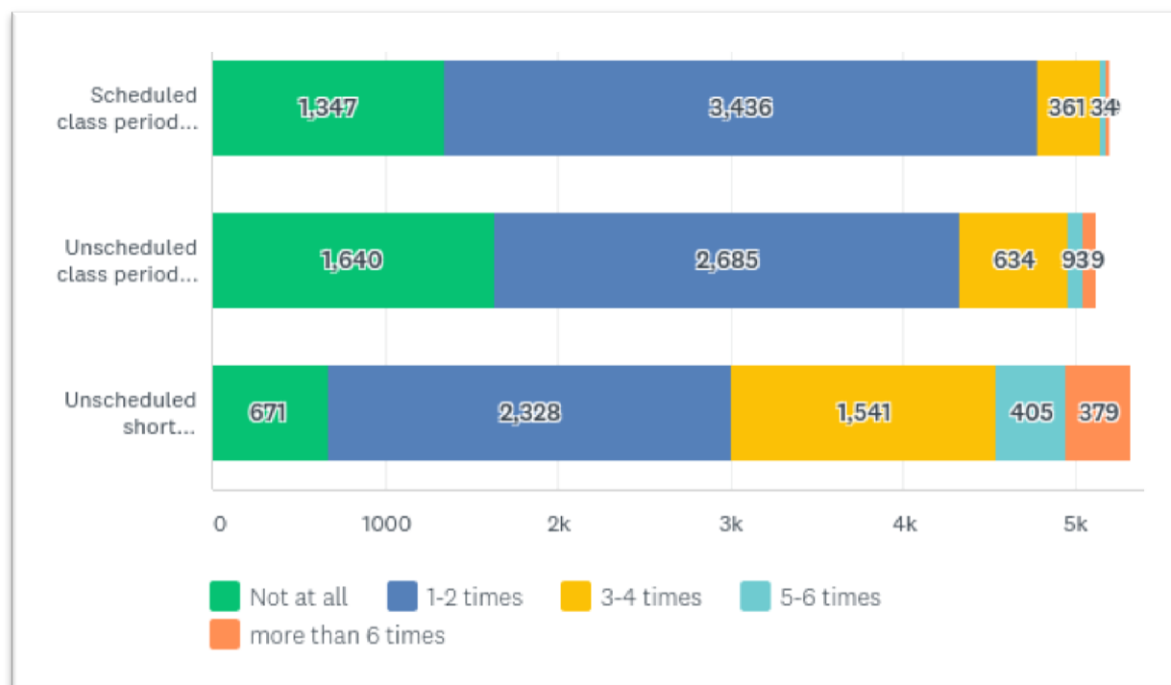
<sup>20</sup> Michigan Educator Evaluations Frequently Asked Questions(FAQs), page 8

September 30, 2017

frequently responded that they were observed 1-2 times.

Figure 50: Teachers: Frequency of teacher observations

**Q14: How frequently are you observed for your evaluation in a school year? Please respond to each type of observation. (5,537)**



	NOT AT ALL	1-2 TIMES	3-4 TIMES	5-6 TIMES	MORE THAN 6 TIMES	TOTAL	WEIGHTED AVERAGE
Scheduled class period or lesson (more than 20 minutes)	25.87% 1,347	65.99% 3,436	6.93% 361	0.65% 34	0.56% 29	5,207	1.84
Unscheduled class period or lesson (more than 20 minutes)	32.02% 1,640	52.43% 2,685	12.38% 634	1.82% 93	1.35% 69	5,121	1.88
Unscheduled short walk-through (up to 20 minutes)	12.60% 671	43.73% 2,328	28.94% 1,541	7.61% 405	7.12% 379	5,324	2.53

September 30, 2017

We looked at answers to this question cross-tabulated by the evaluation model respondents reported using. Those reporting using *Thoughtful Classroom* and *Marzano* did not report receiving more than 6 unscheduled short walk-through observations in high numbers.

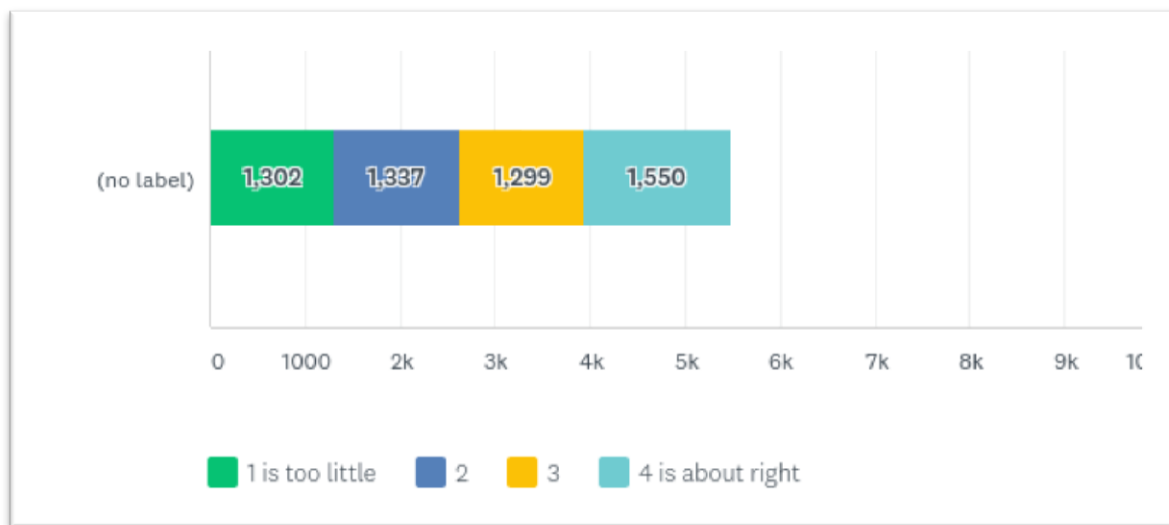
Teachers observed with *5 Dimensions* reported more than 6 unscheduled longer (more than 20 minutes) observations more frequently than those observed by other models. And teachers observed with *other* model reported scheduled observations of more than 6 times.

### *Quality of feedback – timely, accurate, useable, and fair*

Both survey responses and focus group discussion pointed to factors that impact the value of the evaluation and quality of feedback. Timeliness of feedback was important. Focus group participants pointed to the value of immediate feedback, i.e., within a couple of days rather than weeks, and valued detailed face-to-face discussions over scripted or prepackaged text posts delivered by email, text message or memo. Our literature review shows the importance of quality feedback. The value of immediate and targeted feedback was also described in practitioner and researcher whitepapers. We asked survey respondents about the amount of feedback that they received to promote their professional growth (Q15). In a four point forced choice question more than half responded in the highest two categories with 28.24 percent saying *is about right*.

Figure 51: Teachers: Teacher feedback to promote professional growth

**Q15: Evaluation processes allow for various forms of feedback. In your experience with evaluations, the amount of feedback you receive to promote your professional growth and development (n = 5488):**





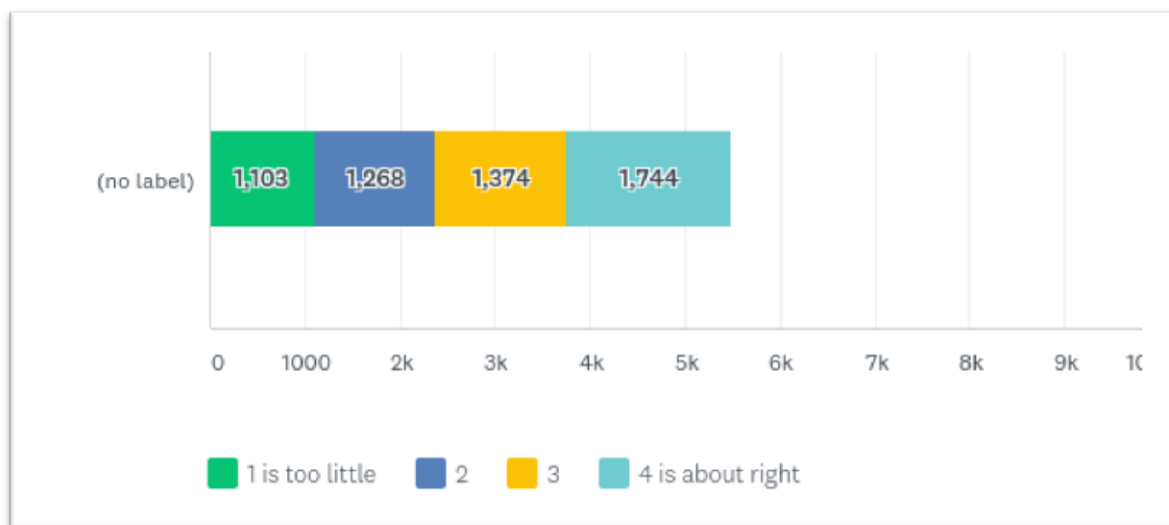
September 30, 2017

	NOT AT ALL	1-2 TIMES	3-4 TIMES	5-6 TIMES	MORE THAN 6 TIMES	TOTAL	WEIGHTED AVERAGE
Scheduled class period or lesson (more than 20 minutes)	25.87% 1,347	65.99% 3,436	6.93% 361	0.65% 34	0.56% 29	5,207	1.84
Unscheduled class period or lesson (more than 20 minutes)	32.02% 1,640	52.43% 2,685	12.38% 634	1.82% 93	1.35% 69	5,121	1.88
Unscheduled short walk-through (up to 20 minutes)	12.60% 671	43.73% 2,328	28.94% 1,541	7.61% 405	7.12% 379	5,324	2.53

And more than half gave a similar response regarding feedback received in their final evaluation; 31.77 percent said *is about right*. Teachers evaluated with *Danielson* and *5 Dimensions* responded *about right* more frequently.

Figure 52: Teachers - Feedback received

**Q16: In your experience with evaluations, the amount of feedback you receive via the final evaluation document, including comments and narrative from your evaluator (n =5,489):**



	1 IS TOO LITTLE	2	3	4 IS ABOUT RIGHT	TOTAL	WEIGHTED AVERAGE
(no label)	20.09% 1,103	23.10% 1,268	25.03% 1,374	31.77% 1,744	5,489	2.68

September 30, 2017

An assumption of the evaluation process is that it impacts educators' practice. We explored that assumption in four ways. We asked respondents about the extent to which their evaluation process impacts their practice in *a) providing innovation, personalized instruction and learning; b) fulfilling their role as a teacher; c) maintaining focus on teaching and learning, and d) addressing needs of historically underserved student populations in their classes*. Each of the four types of impact received their highest rating of three in a five point scale with weighted averages ranging from 2.68 to 2.89.

Figure 53: Teachers - Impact on teachers' practice

**Q17: Please indicate the extent to which your evaluation process impacts your practice in (n=5,501):**



September 30, 2017

	NOT AT ALL OR SELDOM	SOMEWHAT	CONSISTENTLY	TOTAL	WEIGHTED AVERAGE
Provide training in the protocols	51.45% 2,818	40.15% 2,199	8.40% 460	5,477	1.57
Reflect clear standards and expectations	18.18% 997	52.20% 2,863	29.63% 1,625	5,485	2.11
Focus on the most important areas of your professional practice	30.38% 1,666	48.50% 2,660	21.12% 1,158	5,484	1.91
Reflect learning outcomes of your students	24.08% 1,319	51.03% 2,795	24.89% 1,363	5,477	2.01
Provide you with information you need to increase your students' learning	46.83% 2,566	40.30% 2,208	12.87% 705	5,479	1.66
Provide you with information you need for your professional growth	39.31% 2,156	44.35% 2,432	16.34% 896	5,484	1.77
Incorporate the use of multiple data sources	33.89% 1,850	45.47% 2,482	20.64% 1,127	5,459	1.87
Include constructive feedback from your evaluator	23.88% 1,306	44.25% 2,420	31.87% 1,743	5,469	2.08
Allow adequate time for your evaluation process to be implemented as prescribed	33.65% 1,841	42.88% 2,346	23.47% 1,284	5,471	1.90
Provide accurate ratings of your practice and performance	29.90% 1,635	47.54% 2,600	22.56% 1,234	5,469	1.93

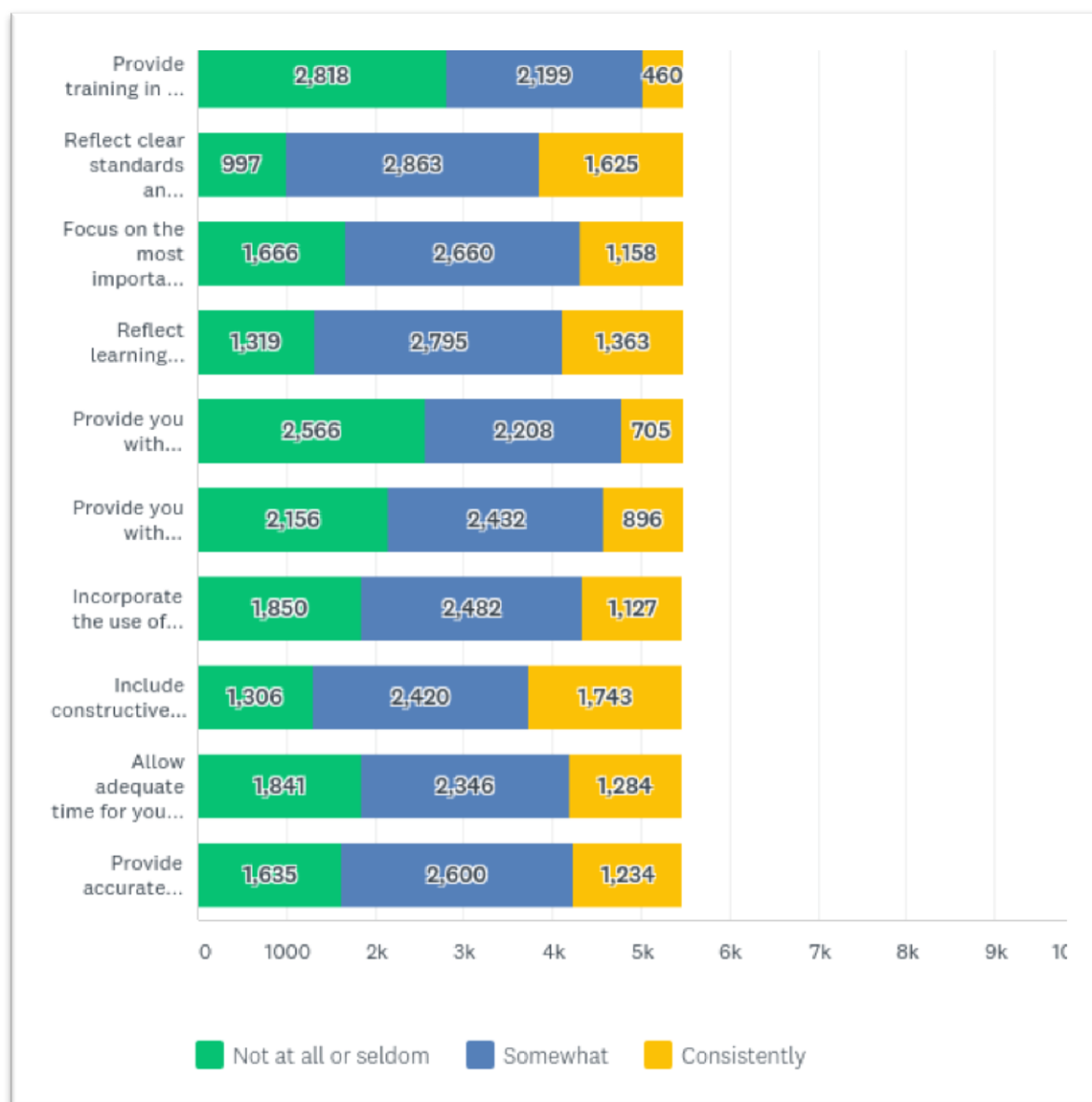
September 30, 2017

When disaggregated by model *other* showed more responses of positive impact in all four of the impact areas and *5 Dimensions* in three of the four impact areas. *Danielson* had comparatively higher responses of negative impact in all four areas and *Thoughtful Classroom* in three of four.

This is worth note because no matter the model there seems to be room for improvement in the essential area of evaluation impact on educators' practice.

Figure 54: Teachers - description of evaluation process

**Q18: To what extent does your evaluation process (5,505):**



September 30, 2017

	NOT AT ALL OR SELDOM	SOMEWHAT	CONSISTENTLY	TOTAL	WEIGHTED AVERAGE
Provide training in the protocols	51.45% 2,818	40.15% 2,199	8.40% 460	5,477	1.57
Reflect clear standards and expectations	18.18% 997	52.20% 2,863	29.63% 1,625	5,485	2.11
Focus on the most important areas of your professional practice	30.38% 1,666	48.50% 2,660	21.12% 1,158	5,484	1.91
Reflect learning outcomes of your students	24.08% 1,319	51.03% 2,795	24.89% 1,363	5,477	2.01
Provide you with information you need to increase your students' learning	46.83% 2,566	40.30% 2,208	12.87% 705	5,479	1.66
Provide you with information you need for your professional growth	39.31% 2,156	44.35% 2,432	16.34% 896	5,484	1.77
Incorporate the use of multiple data sources	33.89% 1,850	45.47% 2,482	20.64% 1,127	5,459	1.87
Include constructive feedback from your evaluator	23.88% 1,306	44.25% 2,420	31.87% 1,743	5,469	2.08
Allow adequate time for your evaluation process to be implemented as prescribed	33.65% 1,841	42.88% 2,346	23.47% 1,284	5,471	1.90
Provide accurate ratings of your practice and performance	29.90% 1,635	47.54% 2,600	22.56% 1,234	5,469	1.93

In our survey we asked both the teachers and administrators several questions about the process for evaluation. For teachers the response was not all together positive (Q18). When asked the extent to which the evaluation process provides training in the protocols, more than half of the teacher respondents replied *not at all or seldom* (51.45%). This finding is consistent with what is heard from

September 30, 2017

focus group participants. In contrast 59.36 percent of administrators who responded said *somewhat* and only 21.91 percent of administrators reported *not at all or seldom*. We saw earlier in this report that administrators report receiving more training than teachers, and their training is more likely to be provided directly by a vendor. As seen earlier, training was consistently cited as a barrier to implementation.

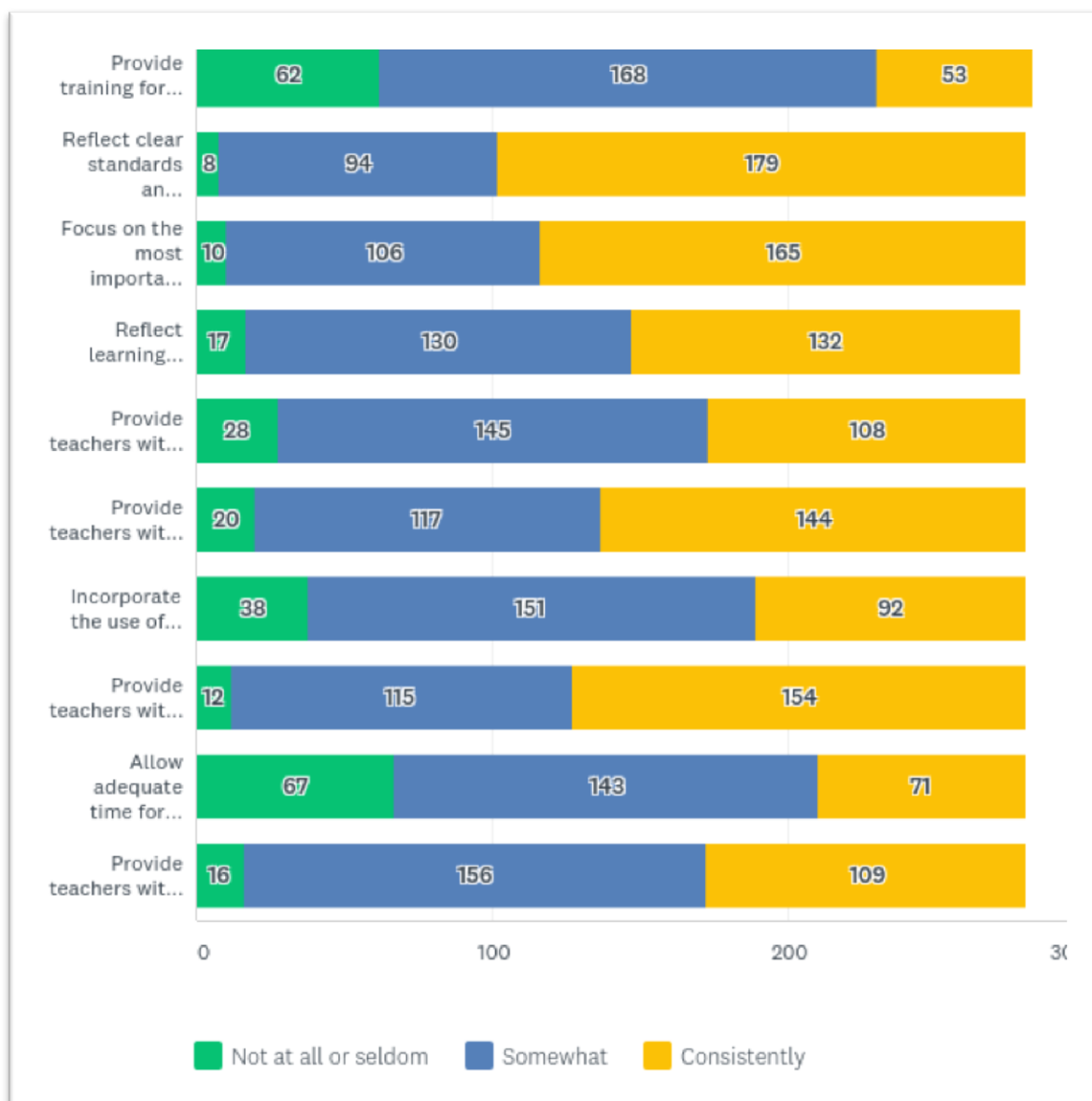
The survey asked the extent the evaluation process reflects clear standards and expectations. Here more than half (52.20 percent) of the teachers responding said *somewhat* and another 29.63 percent say *consistently*. However, 63.70 percent of administrators responded *consistently*. This is important because clear standards are universally cited among the best practices for evaluation.

Similar to asking whether the process represents clear standards, we asked teachers whether the process focused on the most important areas of their professional practice. Nearly half of teachers responded *somewhat* (48.50 percent) and another 30.38 percent said *not at all or seldom*. Only 21.12 percent responded *consistently*. When administrators were asked if the model used focuses on the most important areas of teachers' professional practice 58.78 percent responded *consistently* and only 3.56 percent responded *not at all or seldom*.

September 30, 2017

Figure 55 : Administrators - description of evaluation process

**Q 50: To what extent does the model you use for teacher evaluations (n=283):**



September 30, 2017

	NOT AT ALL OR SELDOM	SOMEWHAT	CONSISTENTLY	TOTAL	WEIGHTED AVERAGE
Provide training for teachers in the use of the protocols	21.91% 62	59.36% 168	18.73% 53	283	1.97
Reflect clear standards and expectations	2.85% 8	33.45% 94	63.70% 179	281	2.61
Focus on the most important areas of teachers' professional practice	3.56% 10	37.72% 106	58.72% 165	281	2.55
Reflect learning outcomes of teachers' students	6.09% 17	46.59% 130	47.31% 132	279	2.41
Provide teachers with information they need to increase their students' learning	9.96% 28	51.60% 145	38.43% 108	281	2.28
Provide teachers with information they need for their professional growth	7.12% 20	41.64% 117	51.25% 144	281	2.44
Incorporate the use of multiple data sources	13.52% 38	53.74% 151	32.74% 92	281	2.19
Provide teachers with constructive feedback and support from you as their evaluator	4.27% 12	40.93% 115	54.80% 154	281	2.51
Allow adequate time for evaluations to be conducted as prescribed	23.84% 67	50.89% 143	25.27% 71	281	2.01
Provide teachers with accurate ratings of their practice	5.69% 16	55.52% 156	38.79% 109	281	2.33

Figures 56 – 65 show administrators responses to survey question 50 regarding their evaluation process sorted by the model used.



September 30, 2017

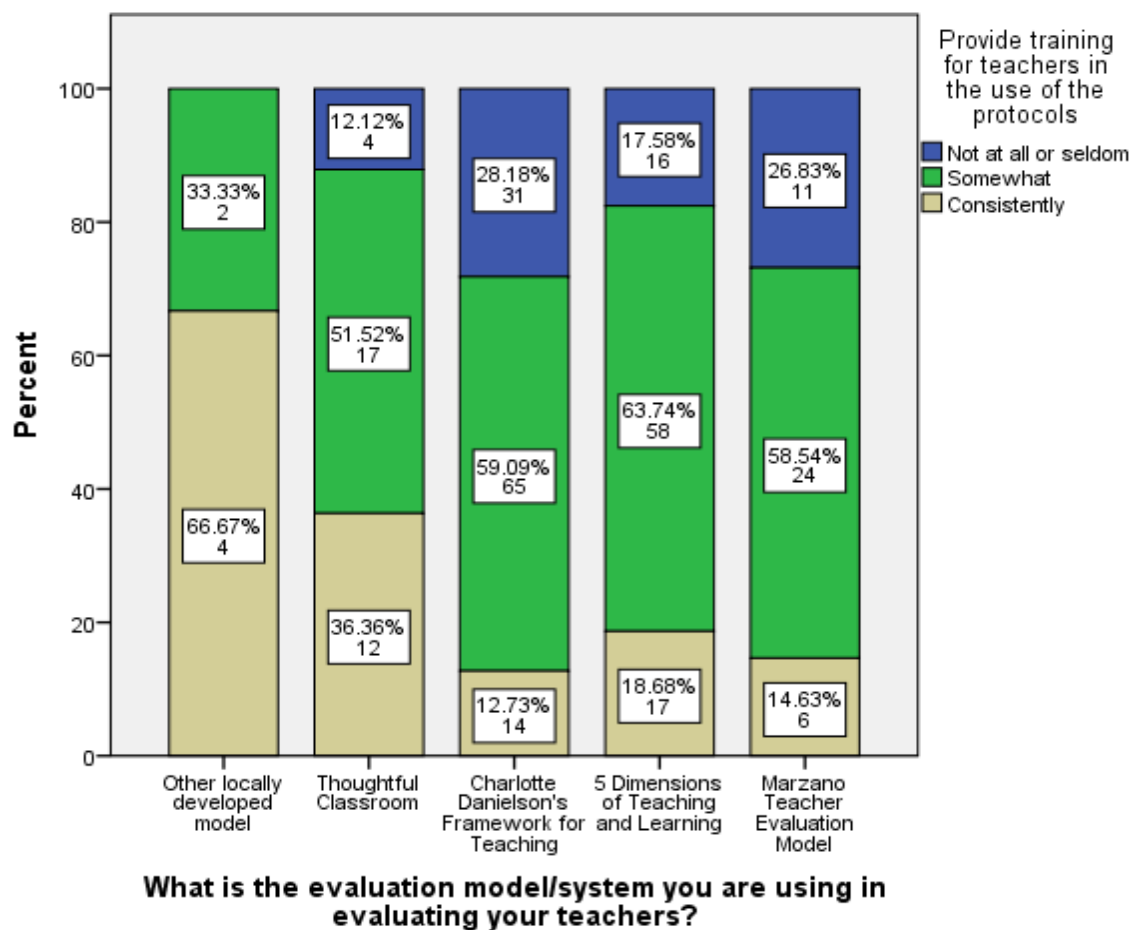


Figure 56: Q 42 x Q 50: Administrators: Provision of training for teachers in use of protocol sorted by model used

September 30, 2017

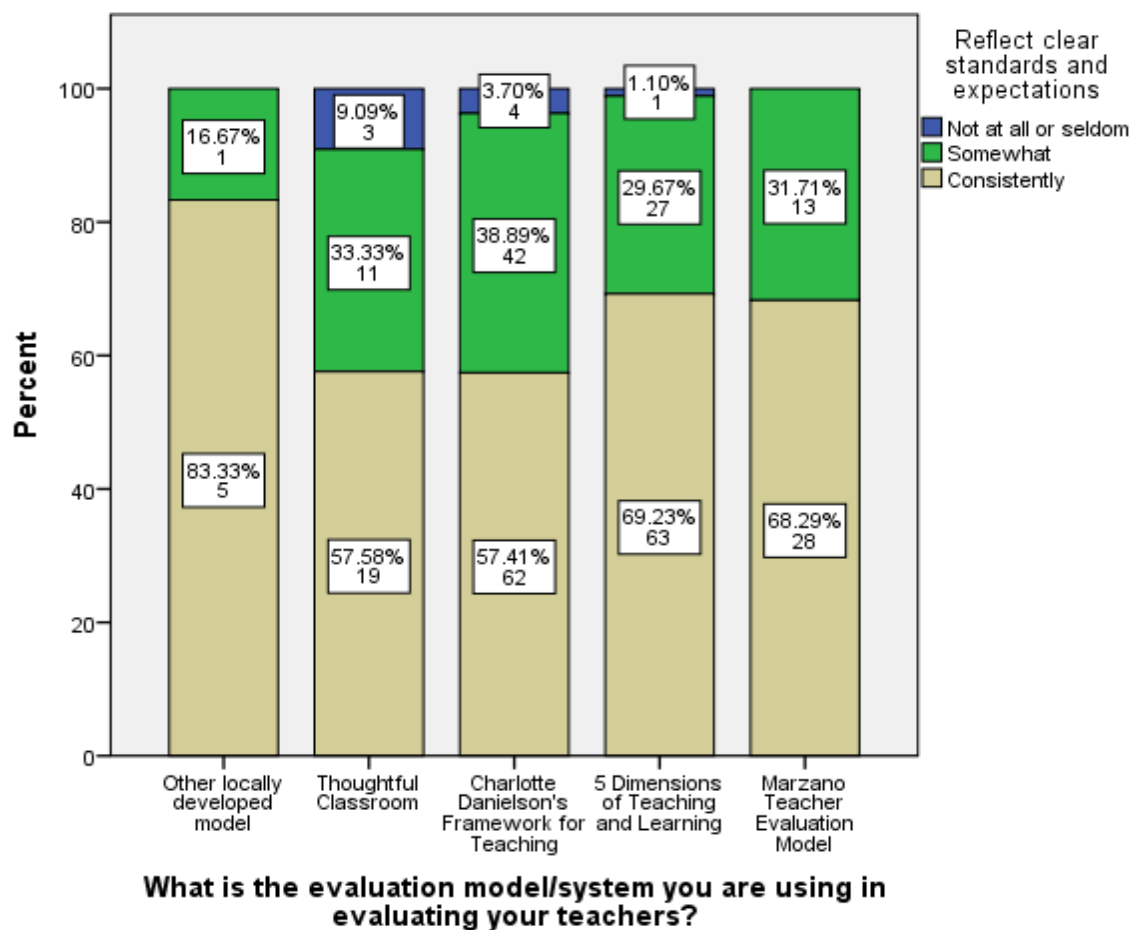


Figure 57: Q 42 x Q 50: Administrators: Reflects clear standards and expectations sorted by model used

September 30, 2017

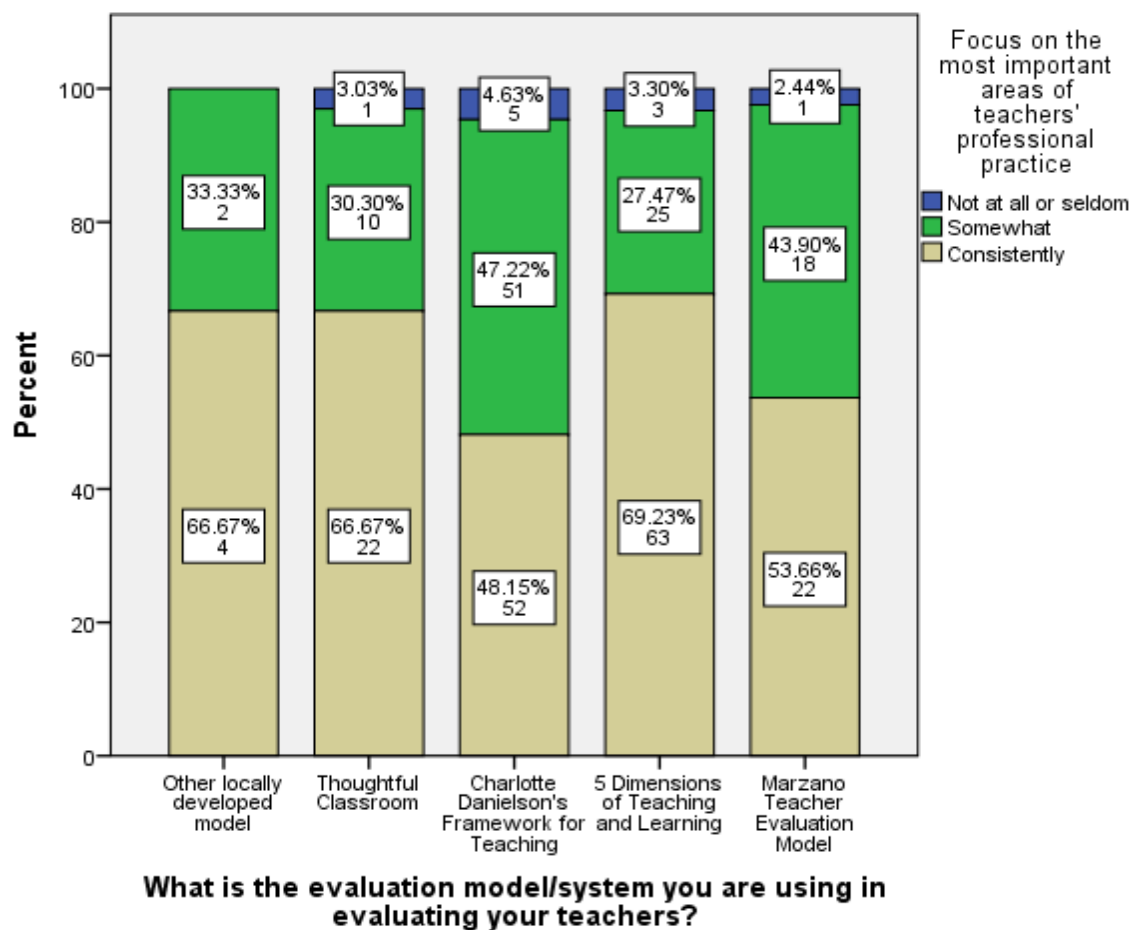


Figure 58: Q 42 x Q 50: Administrators: Focus on the most important areas for teachers' professional practice sorted by model used

September 30, 2017

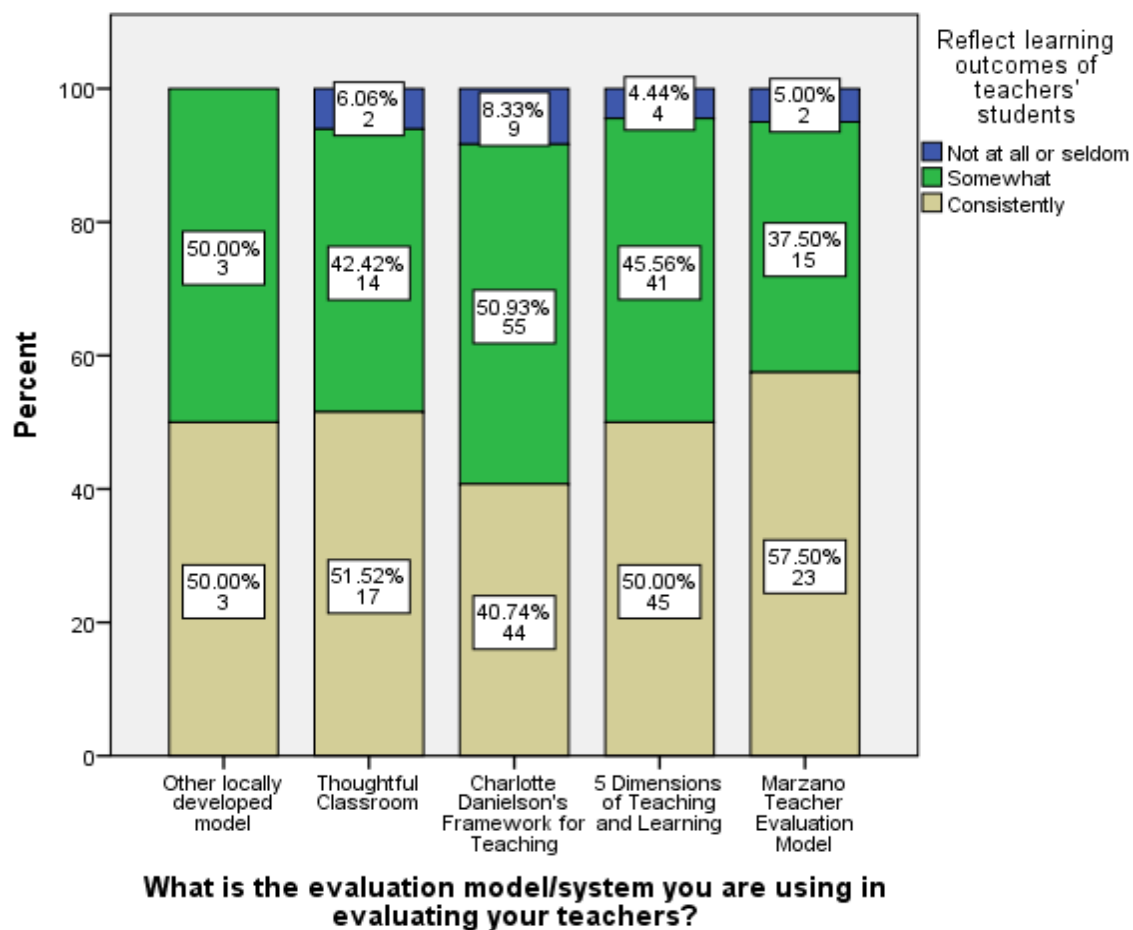


Figure 59: Q 42 x Q 50: Administrators: Reflect learning outcomes of teachers' students sorted by model used

September 30, 2017

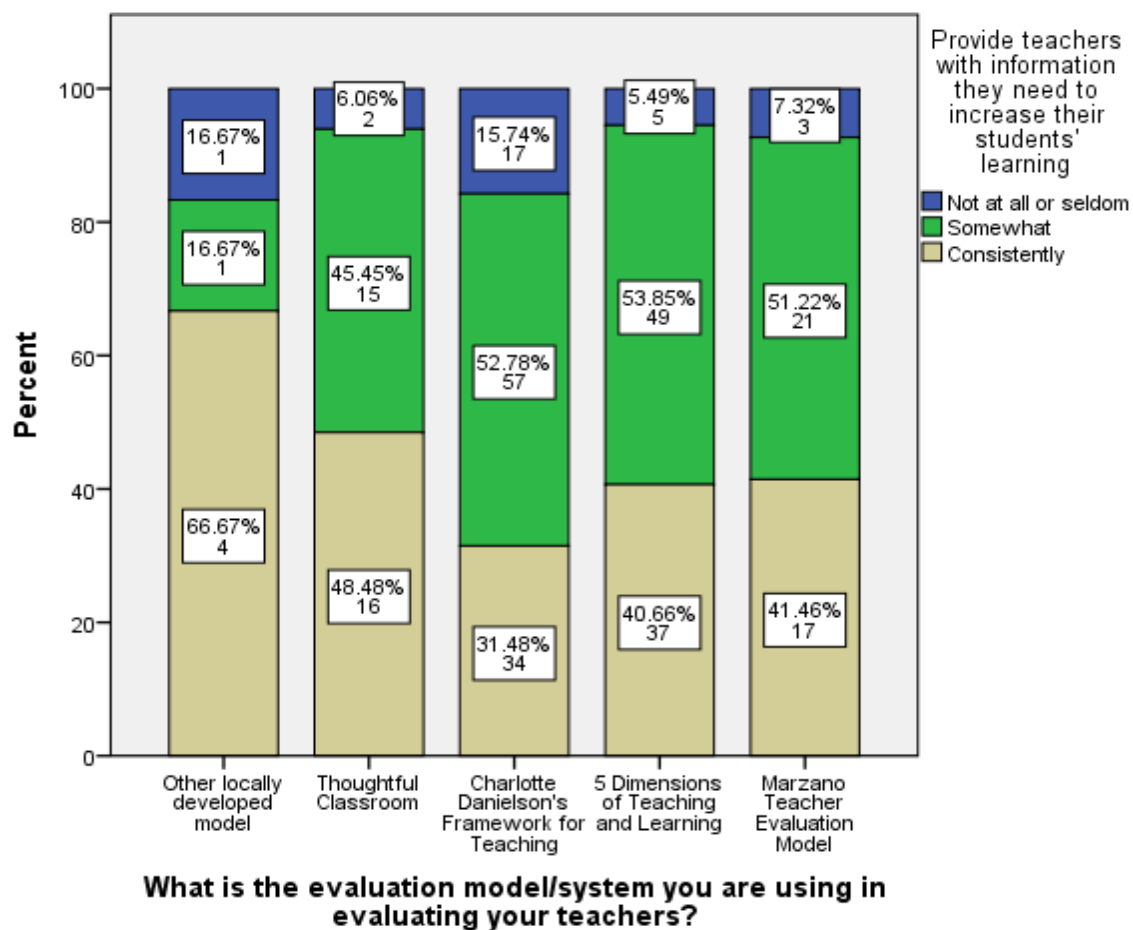


Figure 60: Q 42 x Q 50: Administrators: Provide teachers with information they need to increase their student learning sorted by model used

September 30, 2017

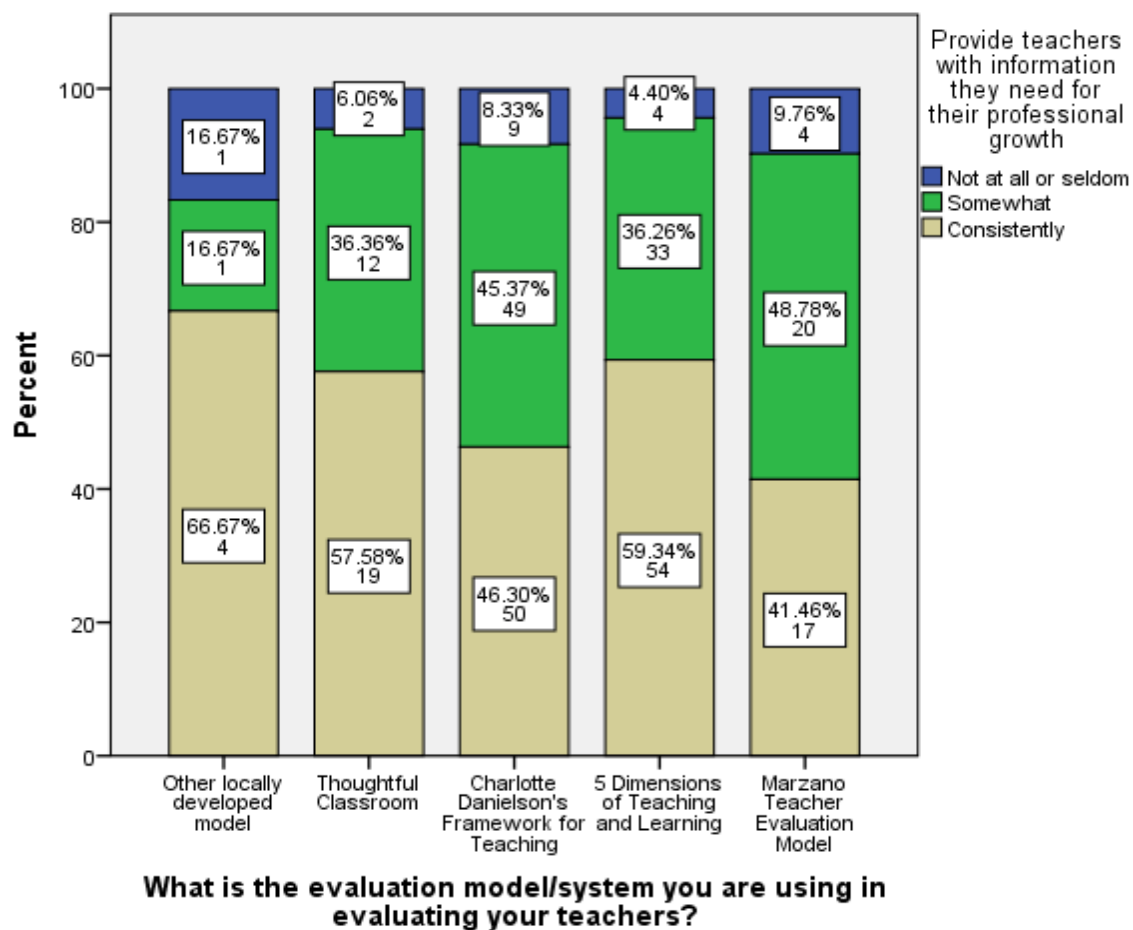


Figure 61: Q 42 x Q 50: Administrators: Provide teachers with information they need for their professional growth sorted by model used

September 30, 2017

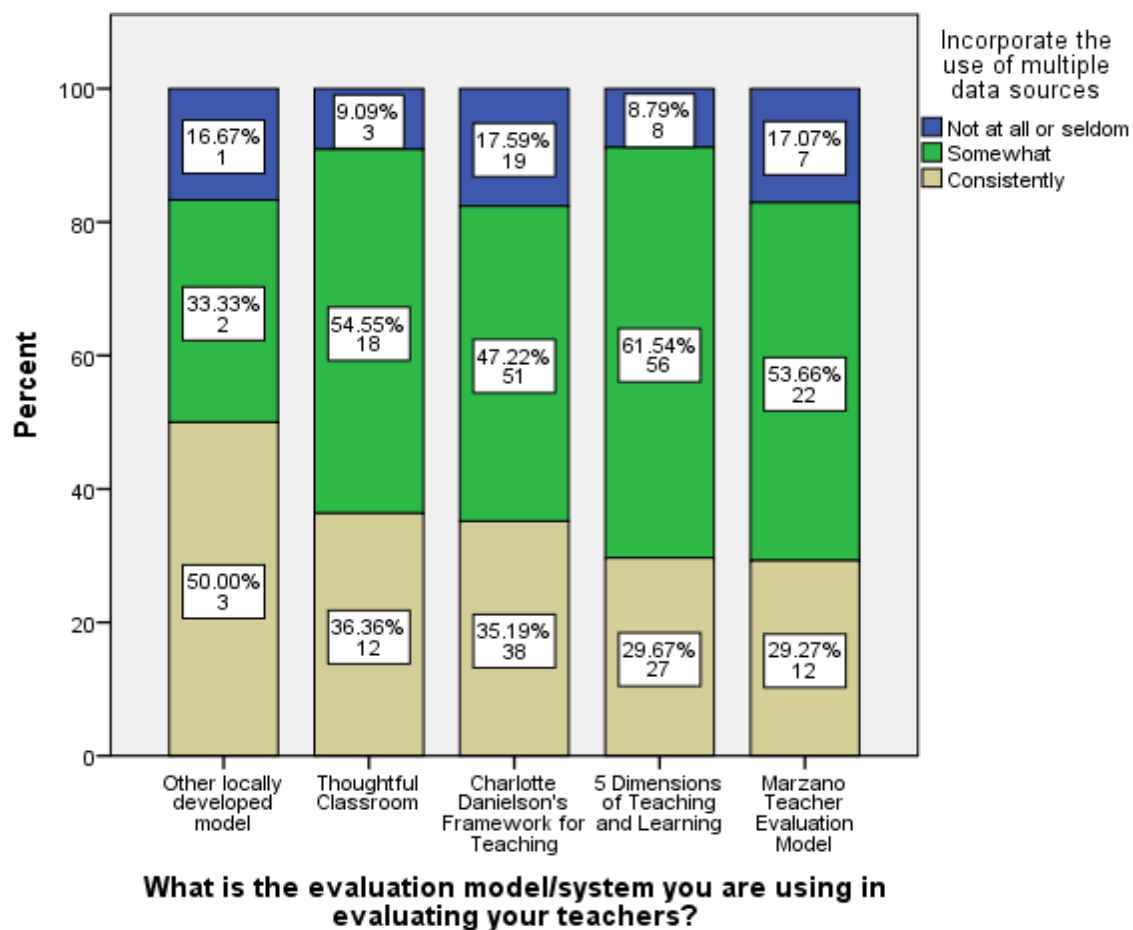


Figure 62: Q 42 x Q 50: Administrators: Incorporate the use of multiple data sources sorted by model used

September 30, 2017

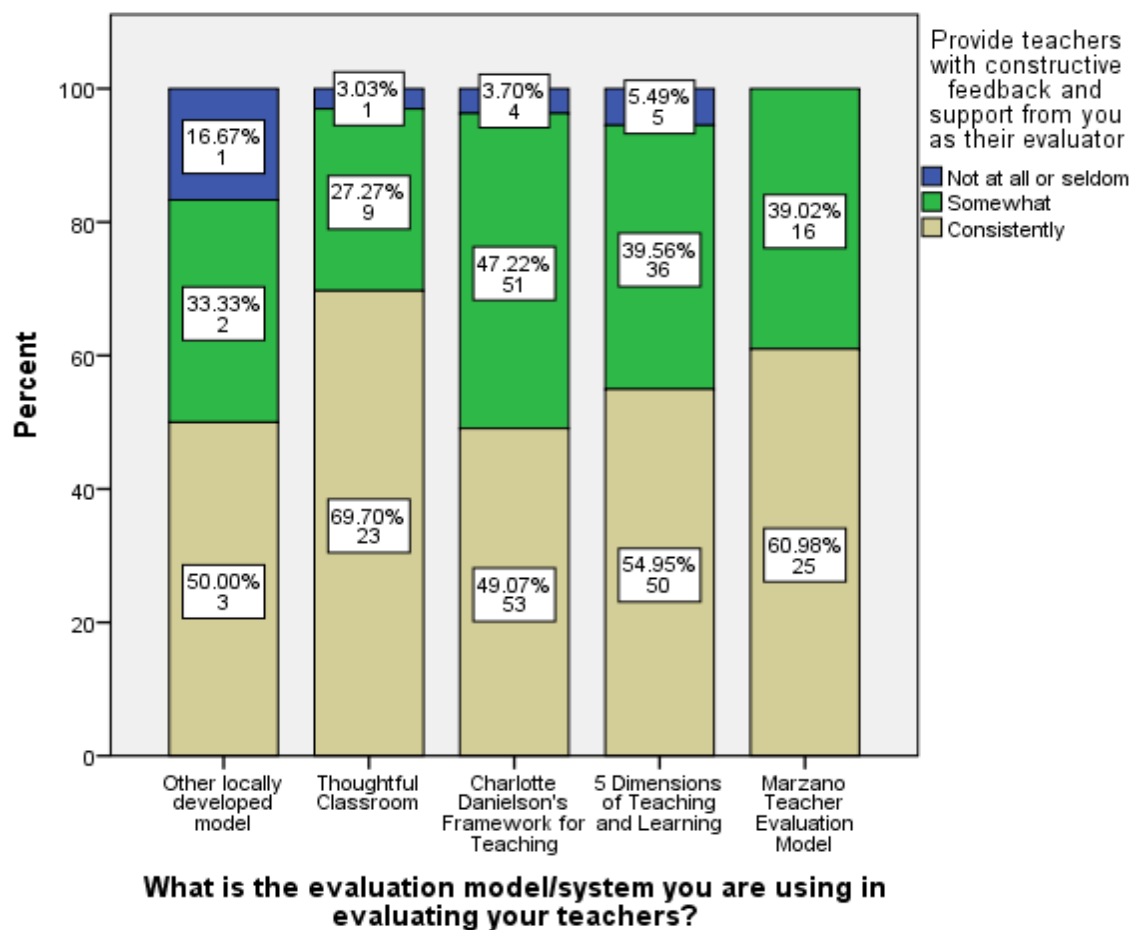


Figure 63: Q 42 x Q 50: Administrators: Provide teachers with constructive feedback and support from you as their evaluator sorted by model used



September 30, 2017

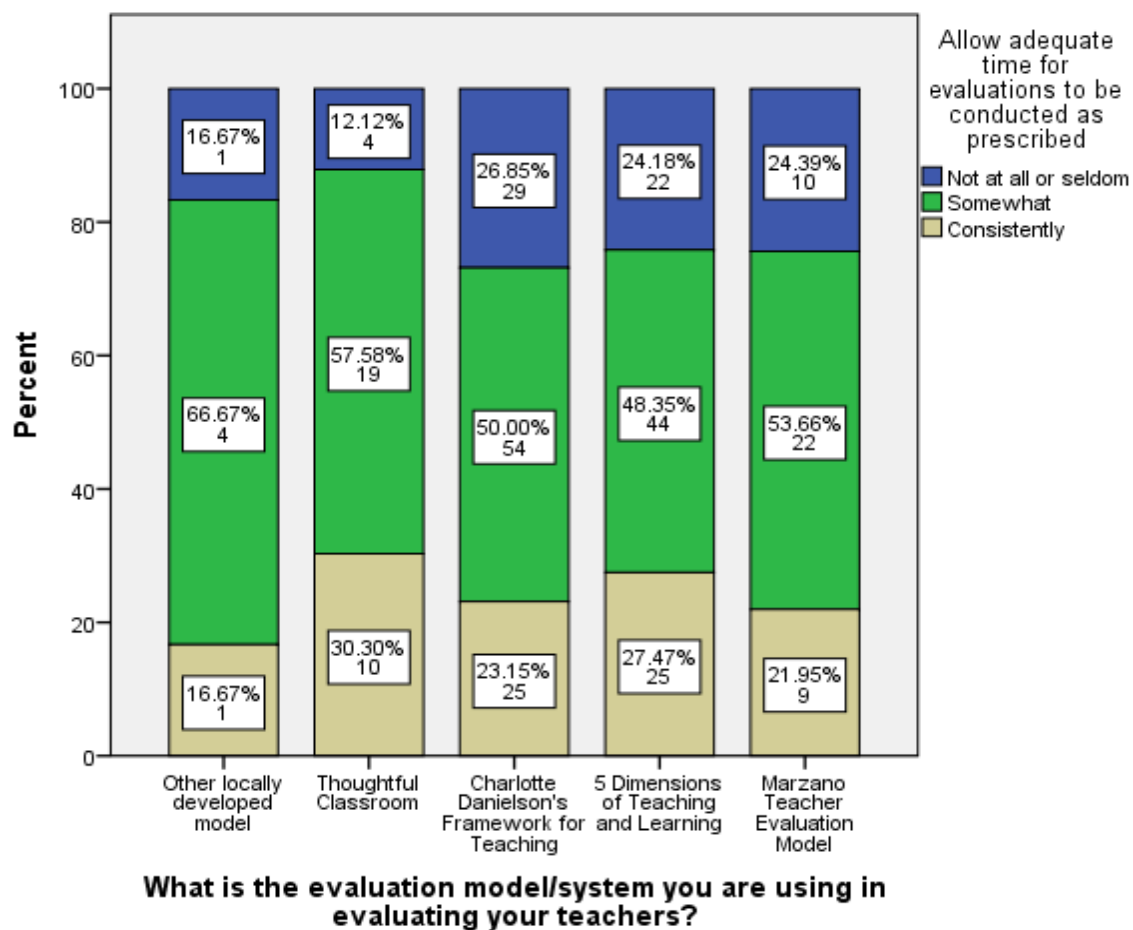


Figure 64: Q 42 x Q 50: Administrators: Allow adequate time for evaluations to be conducted as prescribed sorted by model used

September 30, 2017

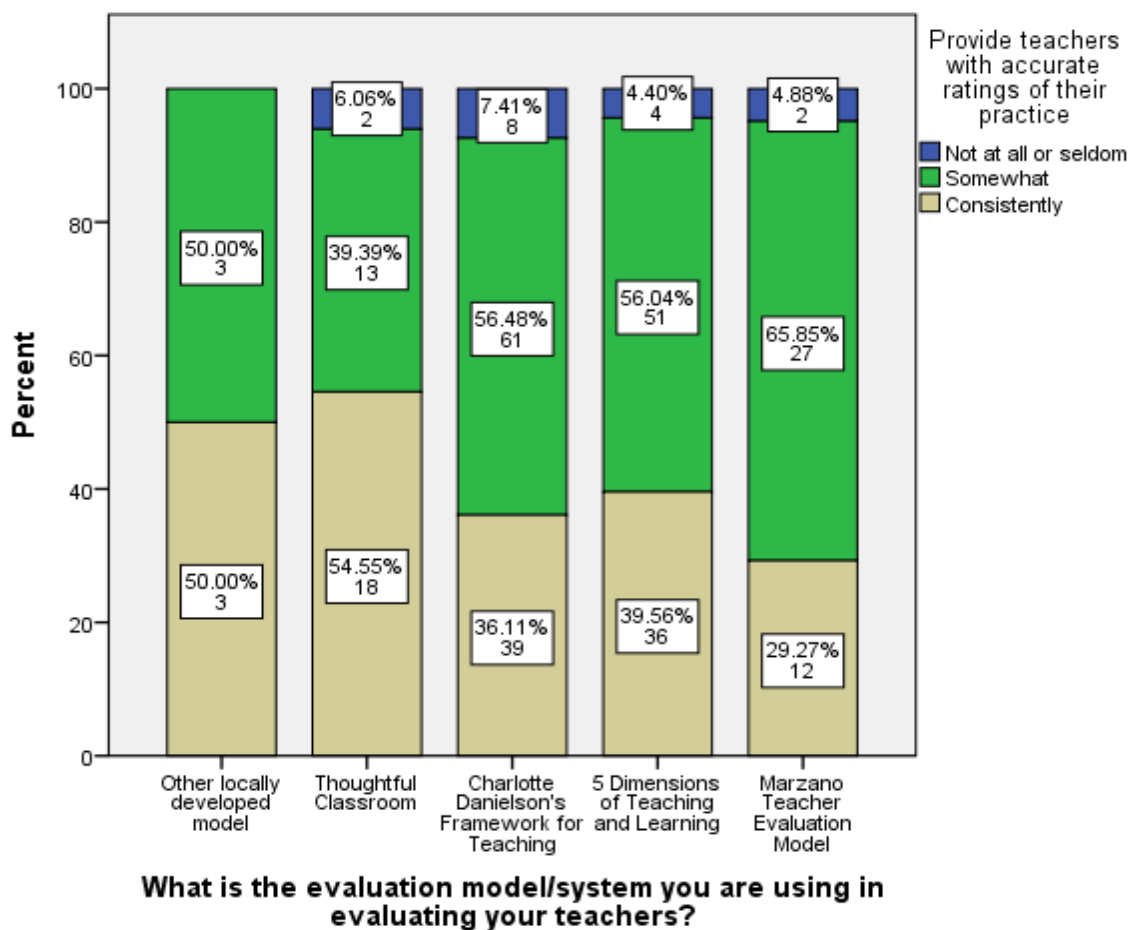


Figure 65: Q 42 x Q 50: Administrators: Provide teachers with accurate ratings of their practice sorted by model used

Usefulness of the evaluation process was cited by teachers and administrators alike as a barrier. This is borne out in the rating that teachers give the process for providing information needed to increase student learning; 46.83 percent said *not at all or seldom*, with another 40.3 percent saying *somewhat*, leaving a mere 12.87 percent responding that the process *consistently* provides the respondent with information needed to increase their students' learning. In contrast 38.43 percent of the administrators responding to our survey said that the model they use *consistently* provides teachers with information they need to increase their students' learning, and another 51.60% responded *somewhat*. Similarly, when teachers were asked if the process provides information needed for their professional growth 39.31 percent responded *not at all or seldom*, and 44.35 percent responded *somewhat*. Administrators were asked the extent that the model provides teachers information for their professional growth and 51.25 percent responded *consistently*. We also asked teachers the extent to which they perceived the process to include constructive feedback from their evaluator, 31.87 percent said *consistently* and 44.25 percent said *somewhat*, with 23.88% saying *not at all or seldom*. These are markedly different responses

September 30, 2017

when compared with administrators, who found that teachers are provided constructive feedback and support from [administrators] as evaluator *consistently* (54.80 percent) and *somewhat* (40.93 percent). Only 4.27 percent of administrators responded *not at all or seldom* compared with 23.88 percent of teachers. Perhaps not surprisingly when asked the extent their evaluation process provides accurate ratings of their practice and performance 47.54 percent of teachers said *somewhat*, 22.56 percent said *consistently*, and 29.90 percent said *not at all or seldom*. 38.79 percent of administrators found that the process provided accurate ratings of teacher practice *consistently*, 55.52 percent responded *somewhat* and only 5.69 percent of administrators responded *not at all or seldom*. This may compare with the rating the respondent expects and may also be reflective of the high rates of Highly Effective ratings found statewide.

One focus group principal described the time that he dedicates evaluating his total staff thusly:

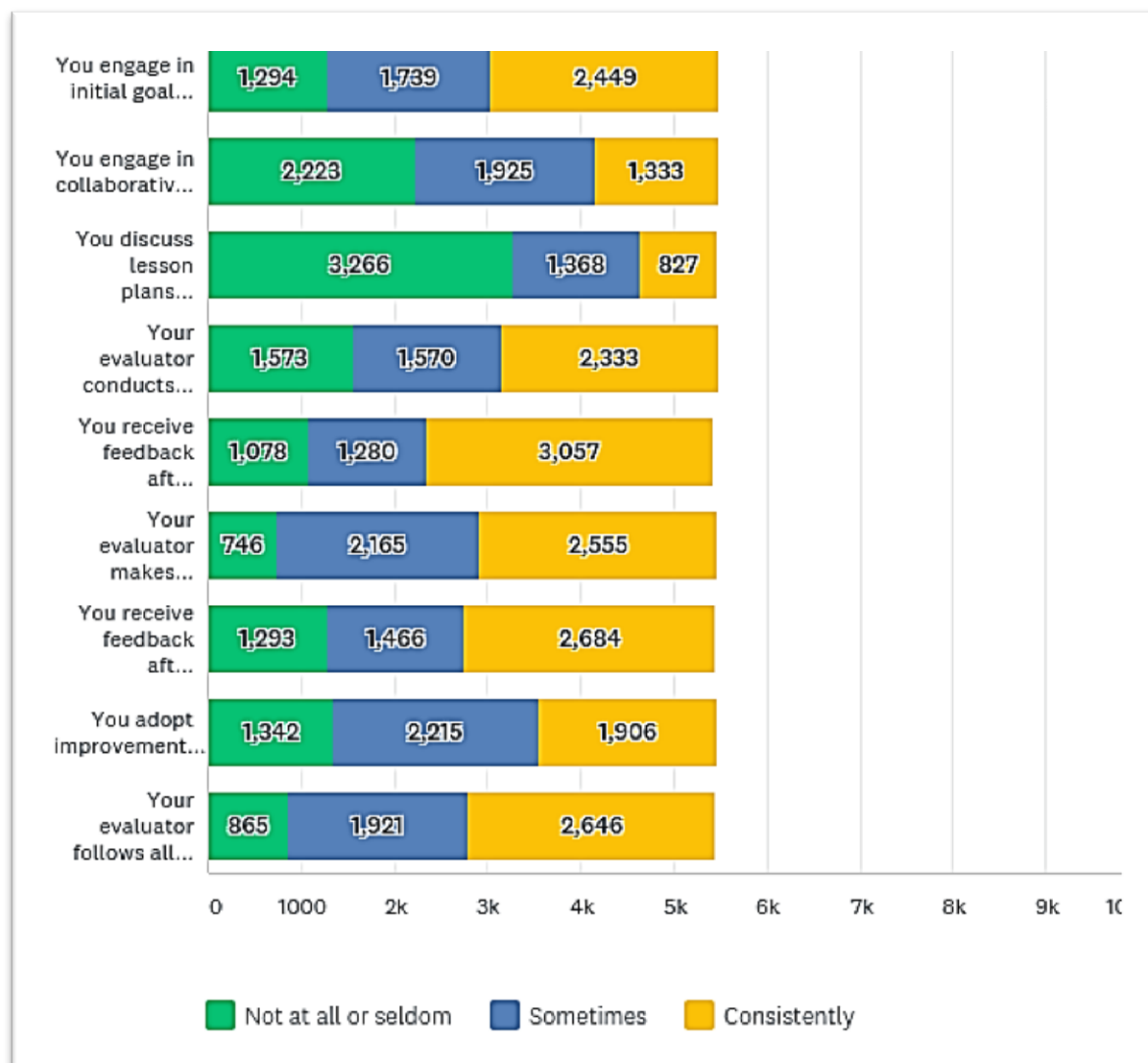
*“ [I] don’t have time to devote to teachers in need and coaching. [We are] jumping through hoops for high performing teachers.”*

To meet the stated purpose of evaluation to improve professional practice these numbers need to improve.

The disparity between low levels of usefulness and the extensive time commitment required can lead to a compliance mindset – just fill out the check-sheet and turn it in. Some focus group participants worried that with the demands of submission of multiple end of year reports evaluation became more a matter of compliance – just one more thing to check off the to-do list – and less about educator growth and improvement.

September 30, 2017

Figure 66: Teachers - description of evaluation

**Q19: To what extent do each of the following occur in your evaluation process? (n =5,488)**

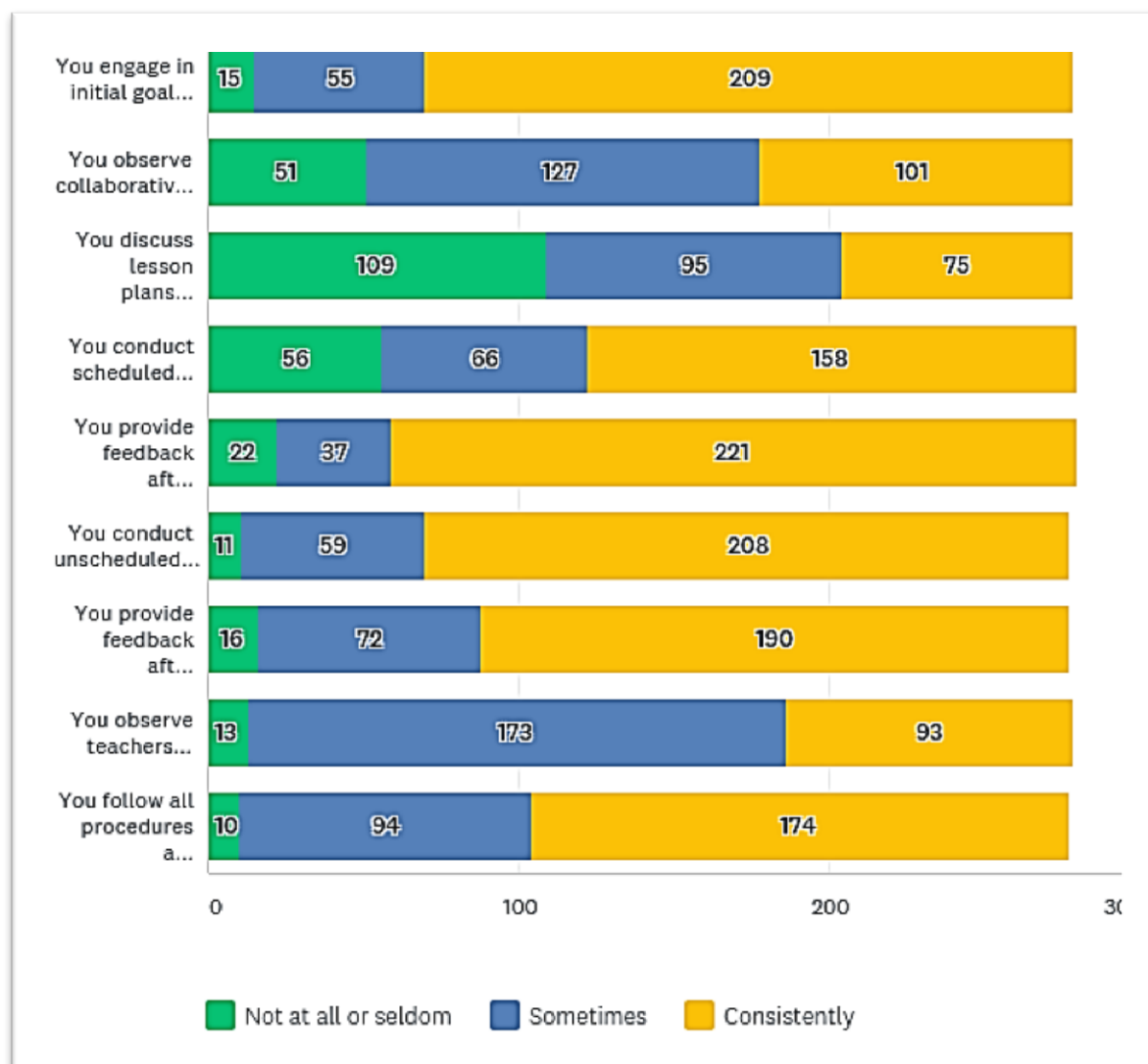
September 30, 2017

	NOT AT ALL OR SELDOM	SOMETIMES	CONSISTENTLY	TOTAL	WEIGHTED AVERAGE
You engage in initial goal setting with your evaluator	23.60% 1,294	31.72% 1,739	44.67% 2,449	5,482	2.21
You engage in collaborative goal setting with your colleagues	40.56% 2,223	35.12% 1,925	24.32% 1,333	5,481	1.84
You discuss lesson plans with your evaluator before each scheduled observation	59.81% 3,266	25.05% 1,368	15.14% 827	5,461	1.55
Your evaluator conducts scheduled observations	28.73% 1,573	28.67% 1,570	42.60% 2,333	5,476	2.14
You receive feedback after each scheduled observation	19.91% 1,078	23.64% 1,280	56.45% 3,057	5,415	2.37
Your evaluator makes unscheduled observations	13.65% 746	39.61% 2,165	46.74% 2,555	5,466	2.33
You receive feedback after each unscheduled observation	23.76% 1,293	26.93% 1,466	49.31% 2,684	5,443	2.26
You adopt improvement strategies and recommendations as discussed with your evaluator	24.57% 1,342	40.55% 2,215	34.89% 1,906	5,463	2.10
Your evaluator follows all procedures and elements prescribed by the evaluation system	15.92% 865	35.36% 1,921	48.71% 2,646	5,432	2.33

Teachers' perception of evaluation process sorted by the model used appears beginning on page or this report. The next few charts show the administrator perceptions of evaluation processes sorted by the model used.

September 30, 2017

Figure 67: Q 51 – Administrators - description of evaluation

**Q51: To what extent do each of the following occur in the evaluations you conduct? (n = 280)**

September 30, 2017

	NOT AT ALL OR SELDOM	SOMETIMES	CONSISTENTLY	TOTAL	WEIGHTED AVERAGE
You engage in initial goal setting with your teachers	5.38% 15	19.71% 55	74.91% 209	279	2.70
You observe collaborative goal setting among teachers and their colleagues	18.28% 51	45.52% 127	36.20% 101	279	2.18
You discuss lesson plans with teachers before each scheduled observation	39.07% 109	34.05% 95	26.88% 75	279	1.88
You conduct scheduled observations	20.00% 56	23.57% 66	56.43% 158	280	2.36
You provide feedback after each scheduled observation	7.86% 22	13.21% 37	78.93% 221	280	2.71
You conduct unscheduled observations	3.96% 11	21.22% 59	74.82% 208	278	2.71
You provide feedback after each unscheduled observation	5.76% 16	25.90% 72	68.35% 190	278	2.63
You observe teachers adopting improvement strategies and recommendations as discussed with you	4.66% 13	62.01% 173	33.33% 93	279	2.29
You follow all procedures and elements prescribed by the evaluation system	3.60% 10	33.81% 94	62.59% 174	278	2.59

The following charts, figures 68-76, show administrators answers to survey question 51 about features of the evaluation process sorted by the model used.

September 30, 2017

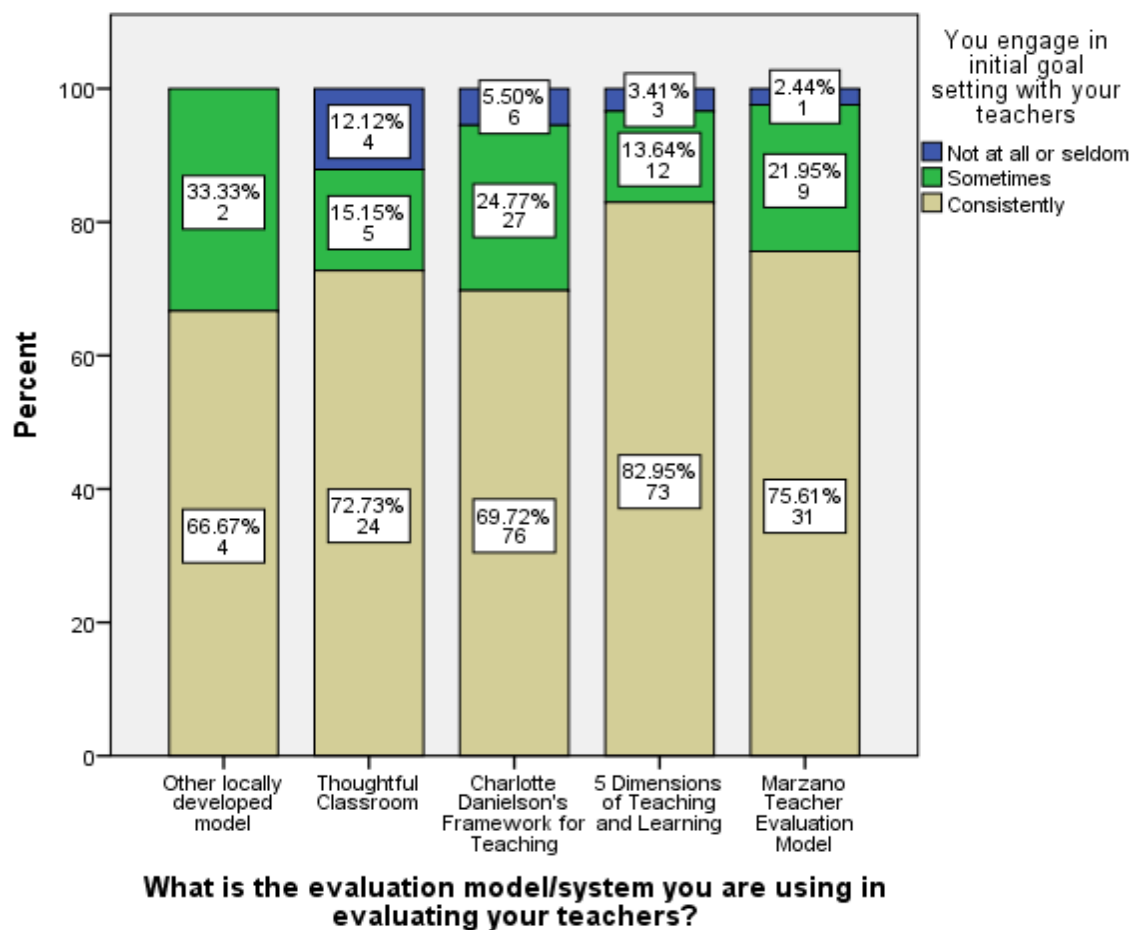


Figure 68: Q 42 x Q 51: Administrators: Engage in initial goal setting with your teachers sorted by model used. (n =280)



September 30, 2017

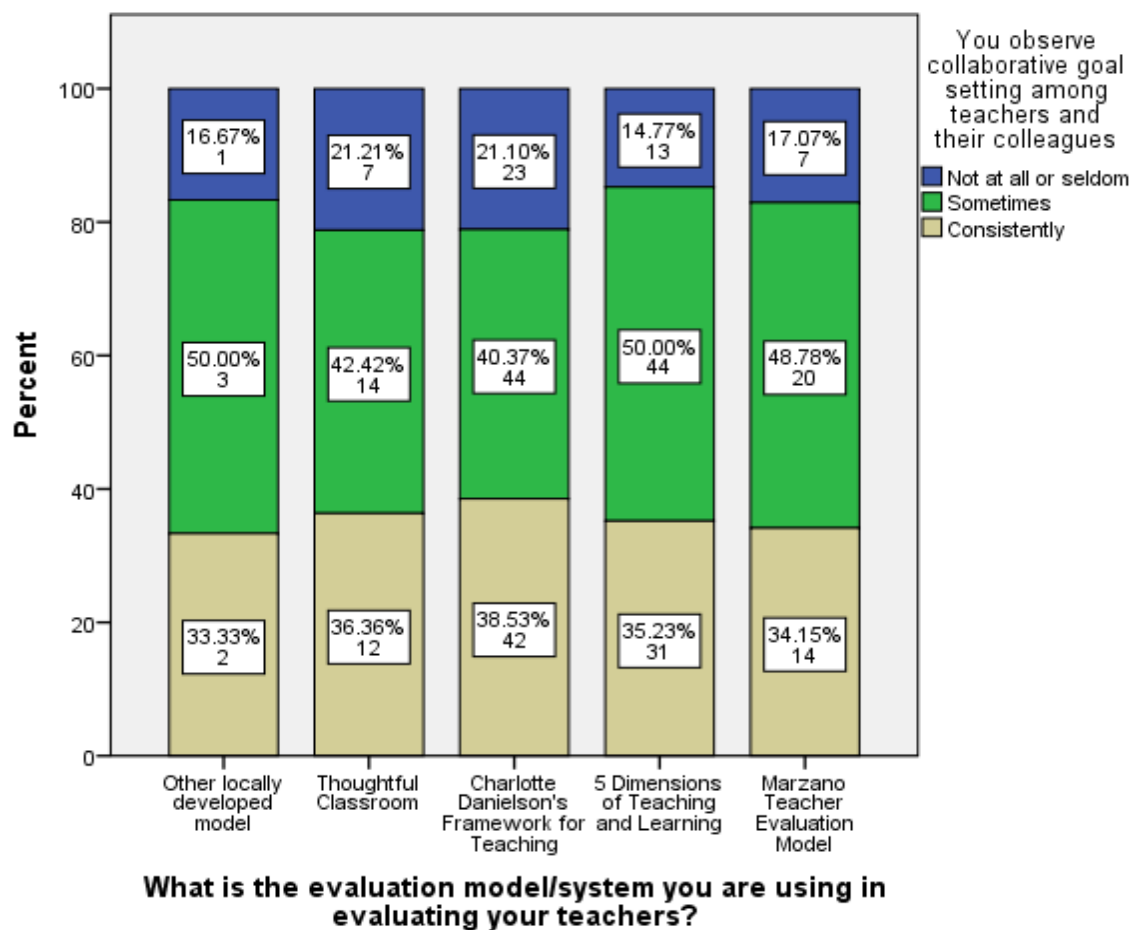


Figure 69: Q 42 x Q 51: Administrators: Observe collective goal setting among teachers and their colleagues sorted by model used (n =280)

September 30, 2017

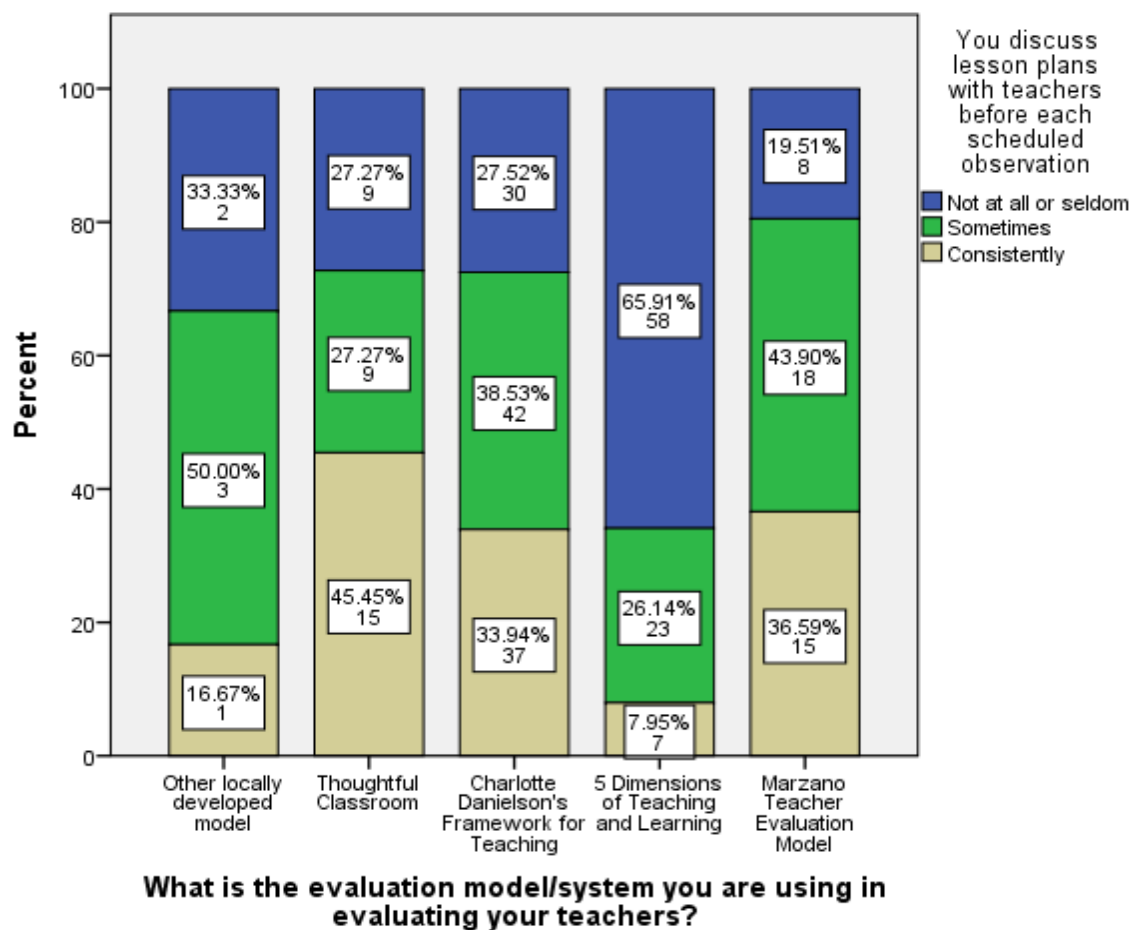


Figure 70: Q 42 x Q 51: Administrators: You discuss lesson plans with teachers before each scheduled observation sorted by model used ( n=280)

September 30, 2017

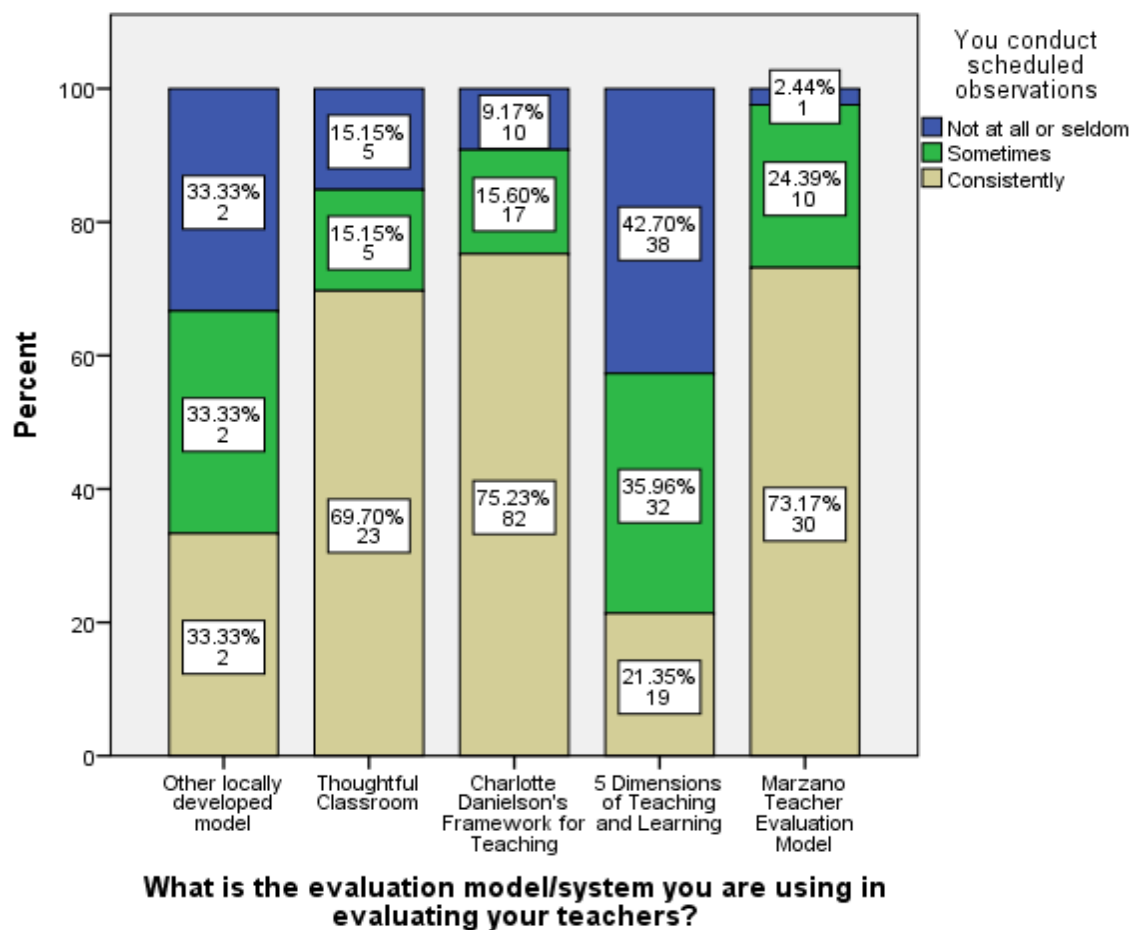


Figure 71: Q 42 x Q 51: Administrators: Conduct scheduled observations sorted by model used (n =280)

September 30, 2017

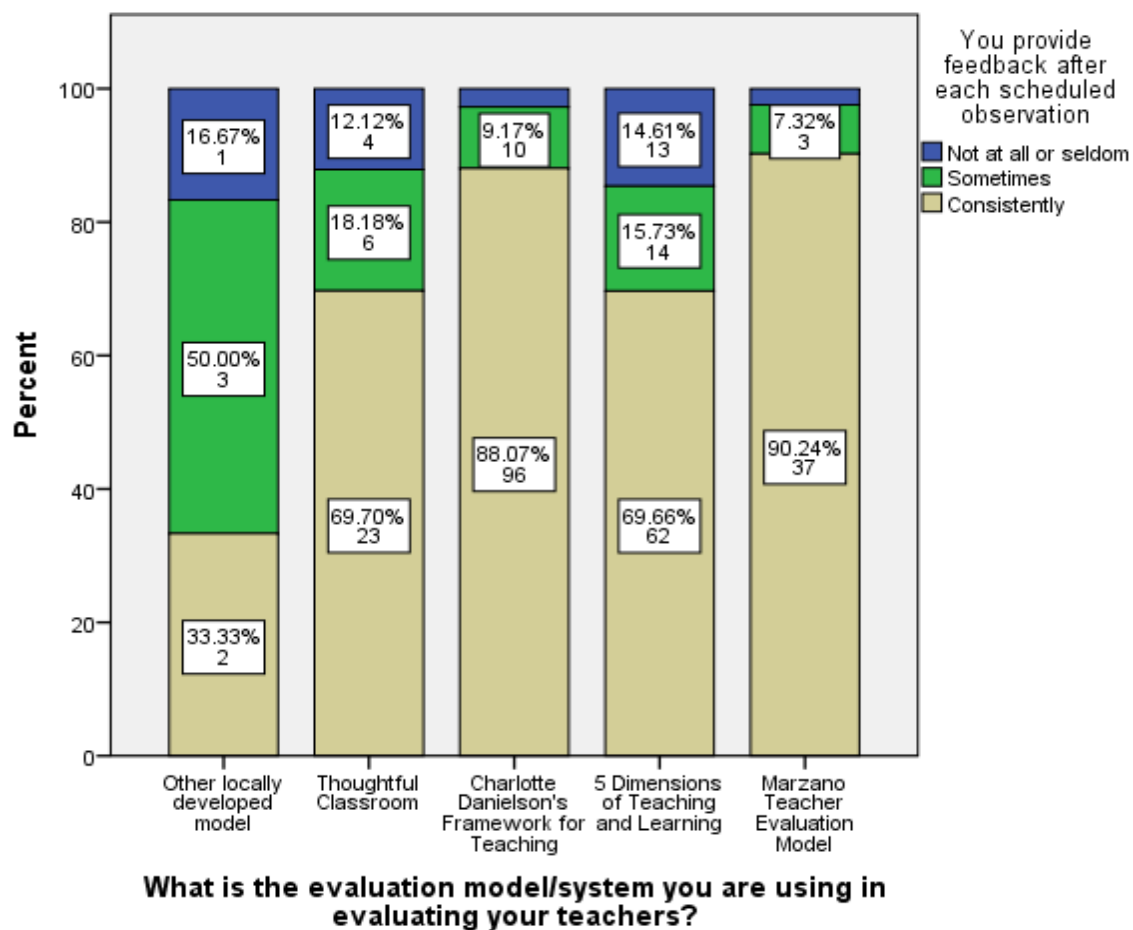


Figure 72: Q 42 x Q 51: Administrators: Provide feedback after each scheduled observation sorted by model used (n =280)

September 30, 2017

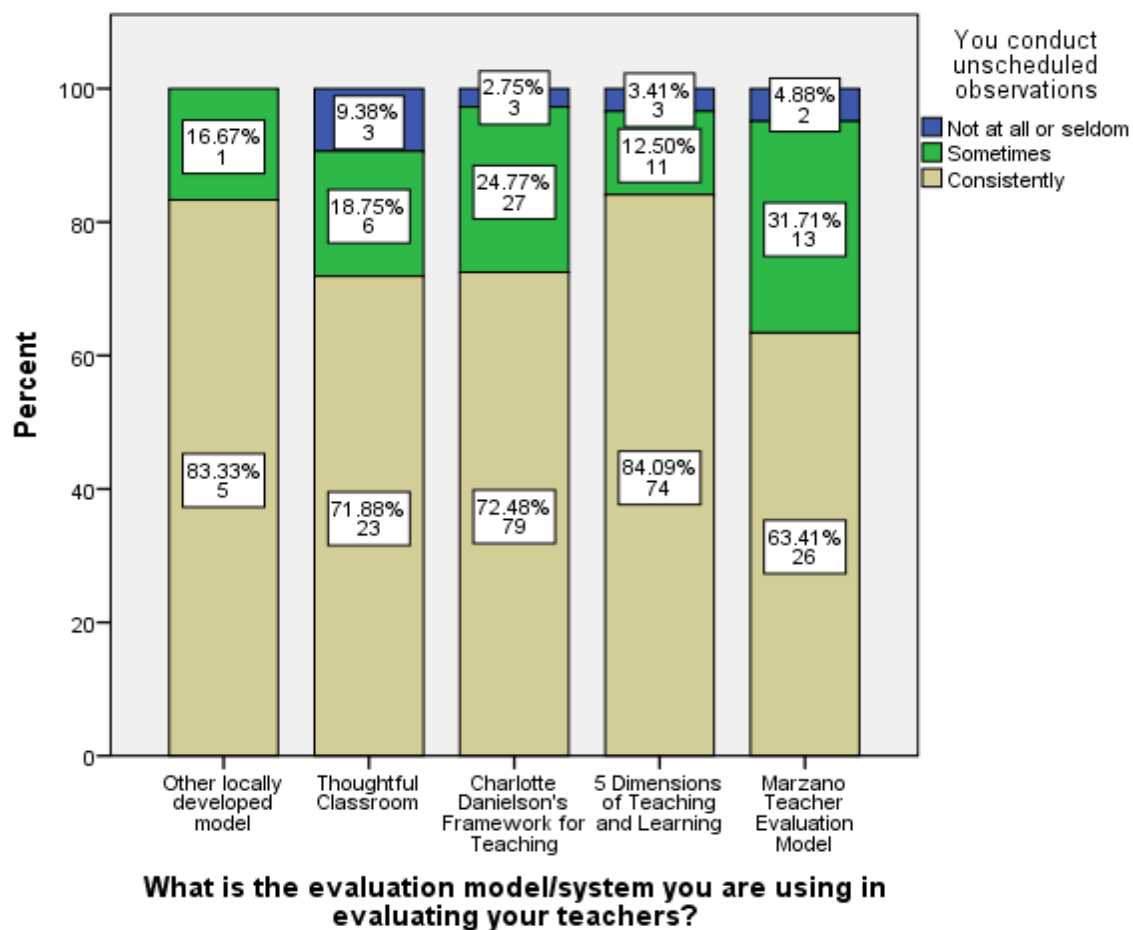


Figure 73: Q 42 x Q 51: Administrators: Conduct unscheduled observations sorted by model used (n =280)

September 30, 2017

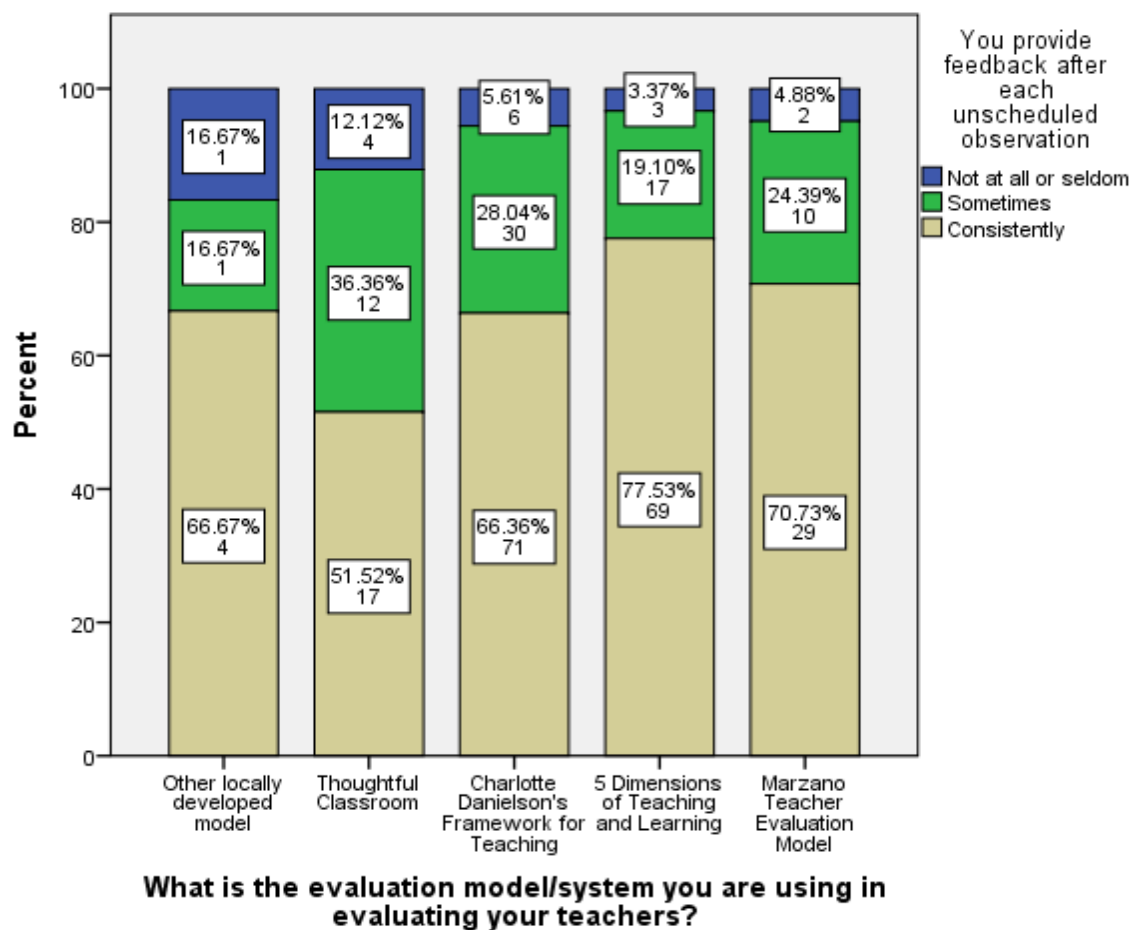


Figure 74: Q 42 x Q 51: Administrators: Provide feedback after each unscheduled observation sorted by model used (n =280)

September 30, 2017

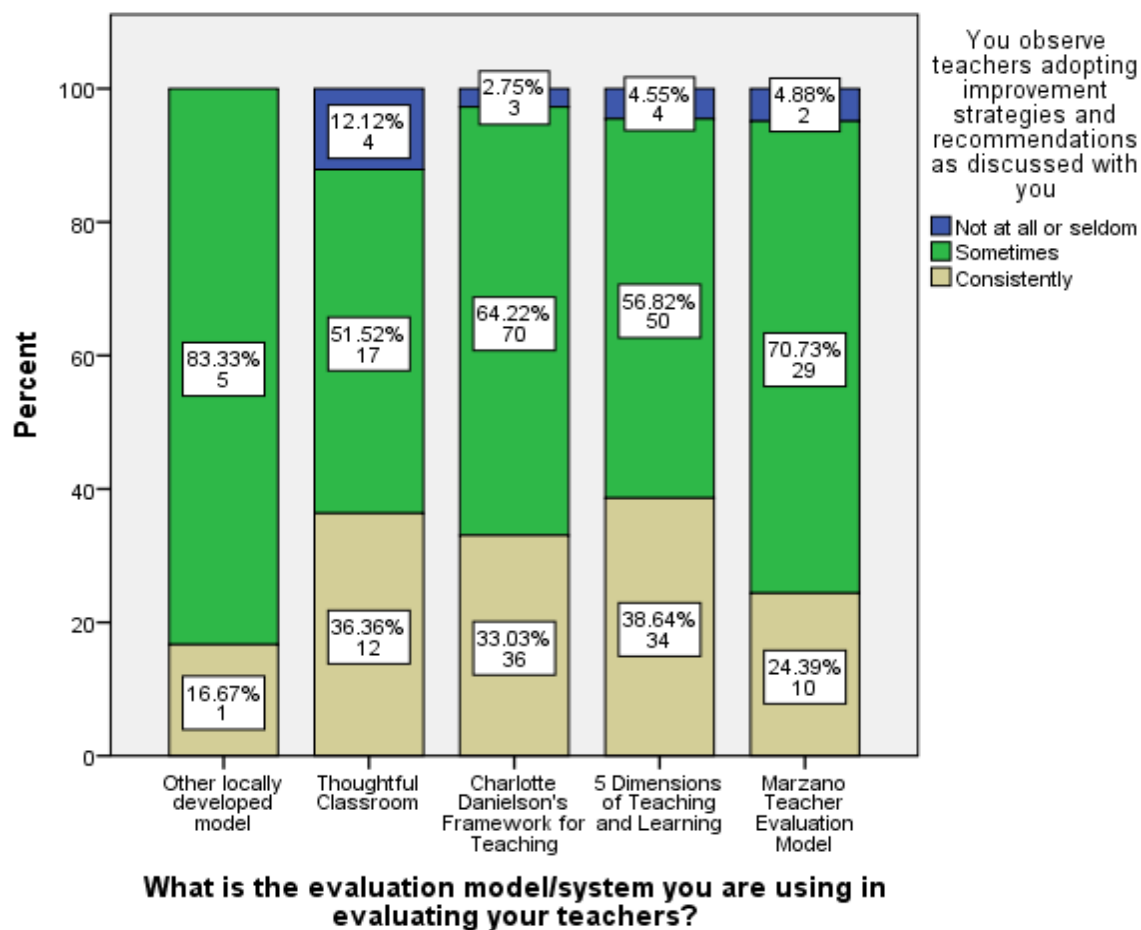


Figure 75: Q 42 x Q 51: Administrators: Observe teachers adopting improvement strategies and recommendations as discussed with you sorted by model used (n =280)

September 30, 2017

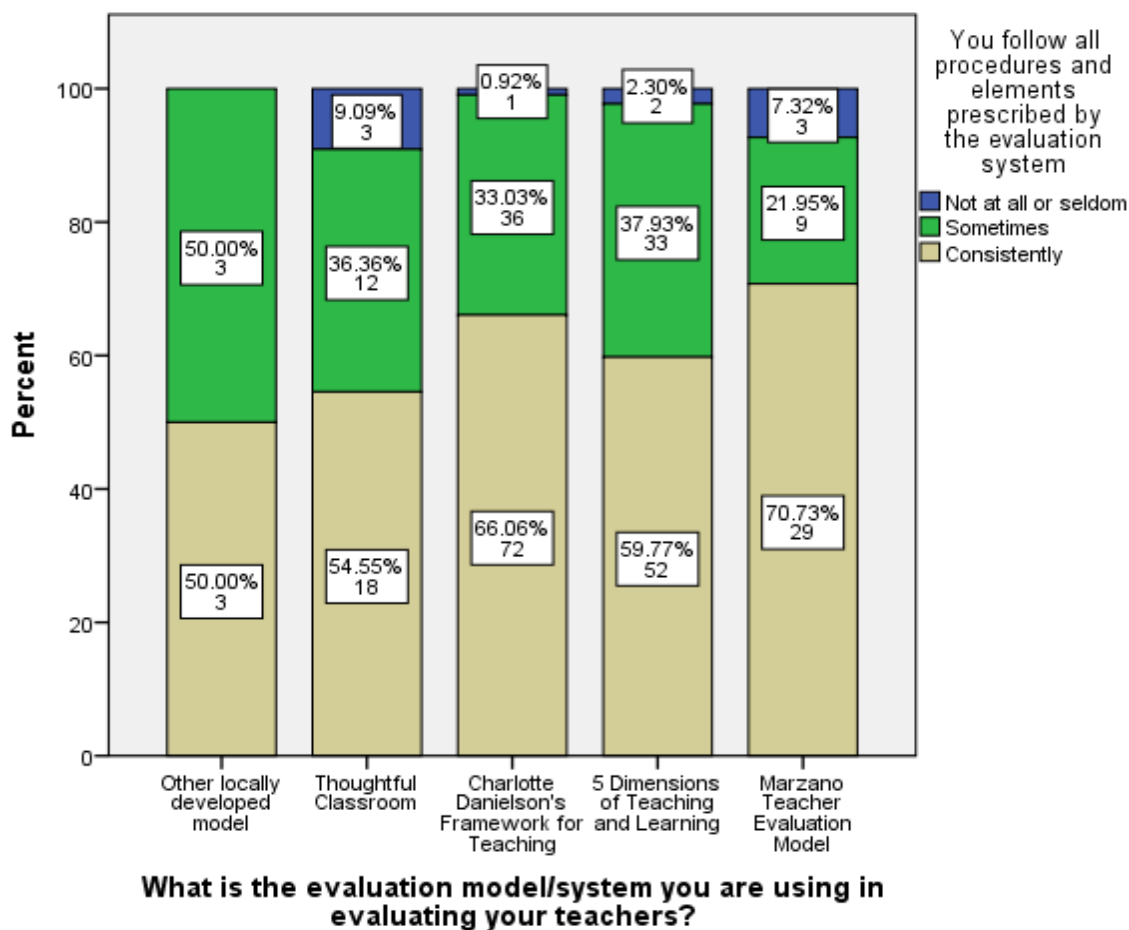


Figure 76: Q 42 x Q 51: Administrators: Follow all procedures and elements prescribed by the evaluation system sorted by model used (n =280)

### *The evaluation process (Q19 and Q51)*

Our literature review describes a series of practices found to be evidence based best practices. Survey question 19 asked teachers the extent to which these practices are found in their evaluation process. One process cited in the literature is an inclusive / collaborative process that involves teachers in goal setting along with their evaluator. 44.67 percent of the teachers responding to the survey report engaging along with their evaluator in initial goal setting *consistently*. However, 40.56 percent responded *not at all or seldom* did they collaborate with colleagues in goal setting. In contrast, 74.91% administrators responded *consistently* engaging in initial goal setting with their teachers; only 5.38% of administrators responded *not at all or seldom*.

During the focus groups we heard conflicting views about the value of planned scripted observations that some labeled *dog and pony shows*. Others, however, described the value of sharing the intent of



September 30, 2017

the lesson and instructional methods used with the evaluator before the observation. In the survey we asked the teachers extent they discuss lesson plans with their evaluator prior to each scheduled observation. Well over half of teachers responding, 59.81 percent, report *seldom or not at all*. We asked administrators the extent they discuss lesson plans with teachers prior to scheduled observations and 39.07 percent of administrators replied *not at all or seldom*. 42.60 percent of the teacher respondents said their evaluator *consistently* conducts scheduled observations – 56.43 percent of administrators said that they conduct scheduled evaluations *consistently*, and 46.74 percent of teachers say their evaluator *consistently* conducts unscheduled observations, 74.82 percent of administrators say they consistently conduct unscheduled evaluations. During focus groups participants spoke of the value of short frequent unscheduled observations to the evaluation process.

Whether scheduled or unscheduled, teachers valued swift and detailed feedback. 56.45 percent of the teacher respondents to our survey say they *consistently* receive feedback after each scheduled observation, 49.31 percent after each unscheduled observation. 78.82 percent of administrators tell us that they *consistently* provide feedback after each scheduled observation and 68.35% after each unscheduled.

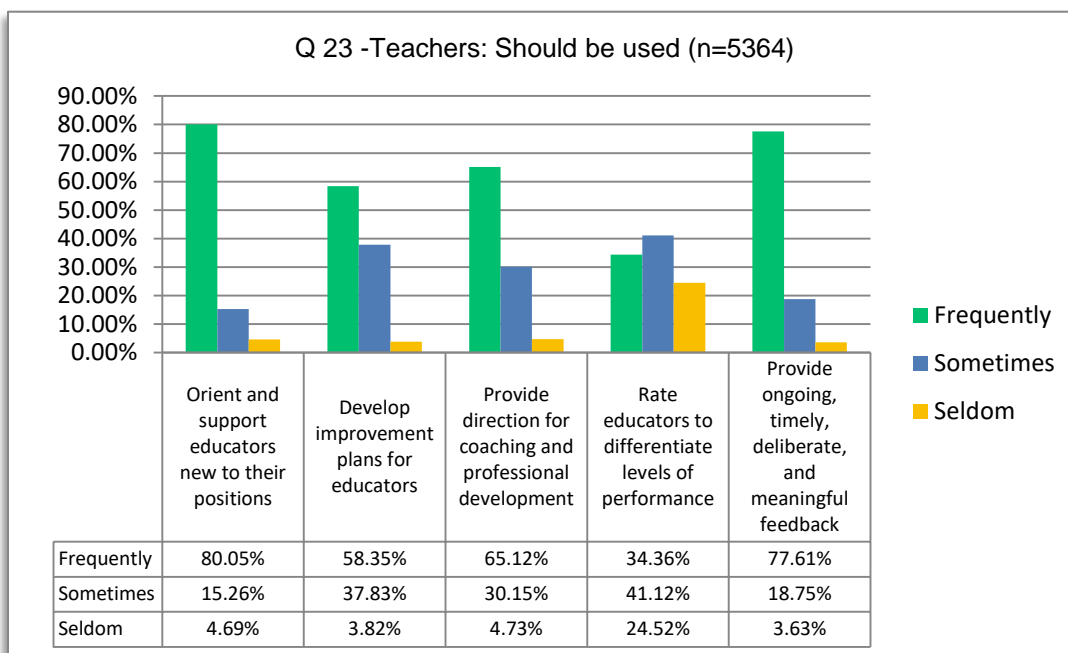
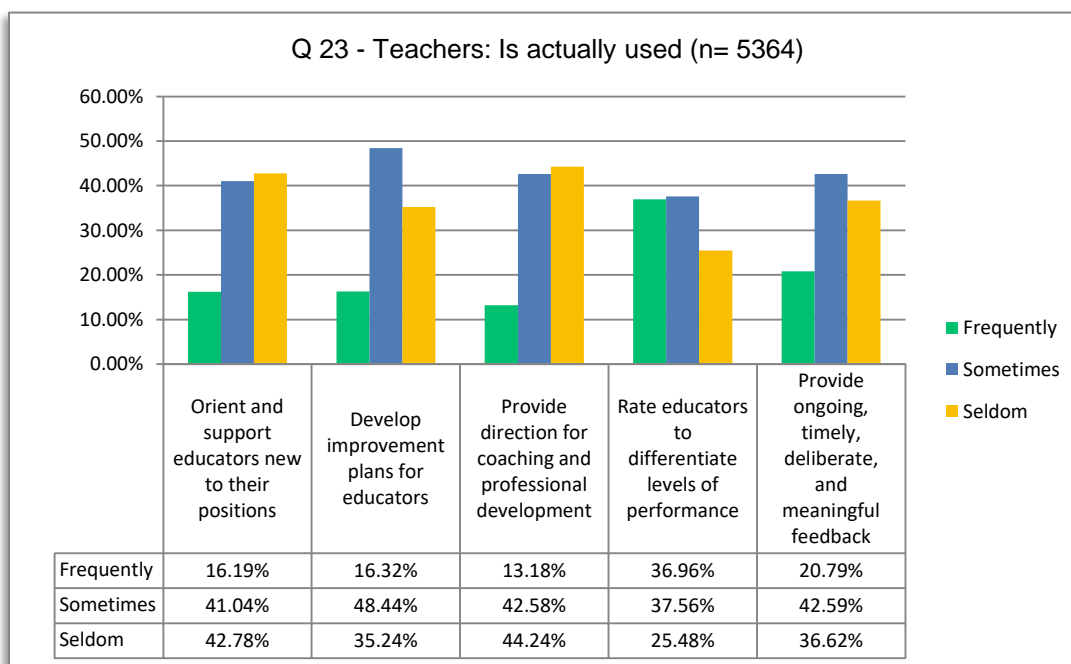
Notably, nearly 20 percent of teachers report *not at all or seldom* receiving feedback no matter if the observation is scheduled (19.97 percent) or unscheduled (23.76 percent). Administrators report *not at all or seldom* providing feedback to observations at far lower rates –for scheduled observations 7.86 percent and for unscheduled observations 5.76 percent.

A central component to educator evaluation usefulness as a tool for professional growth is measured by its capacity to lead to adoption of improvement strategies, professional development and changes in practices. We asked teachers the extent to which they adopt improvement strategies and recommendations as discussed with their evaluator. 40.56 percent of the responding teachers report *sometimes*, and 34.89 percent say *consistently*. A full 24.57 percent say *seldom or not at all*. This finding is reinforced when we asked respondents the frequency educator evaluations should be used for specific purposes in the evaluation process and how frequently it is actually used. Respondents said that evaluations should be most frequently used to *orient and support educators new to their positions* (80.05%) and *provide ongoing, timely, deliberate, and meaningful feedback* (77.61%). In contrast when asked the extent these two purposes are actually used, *support for new teachers* is actually used 16.19% and *ongoing and meaningful feedback* 20.79%.

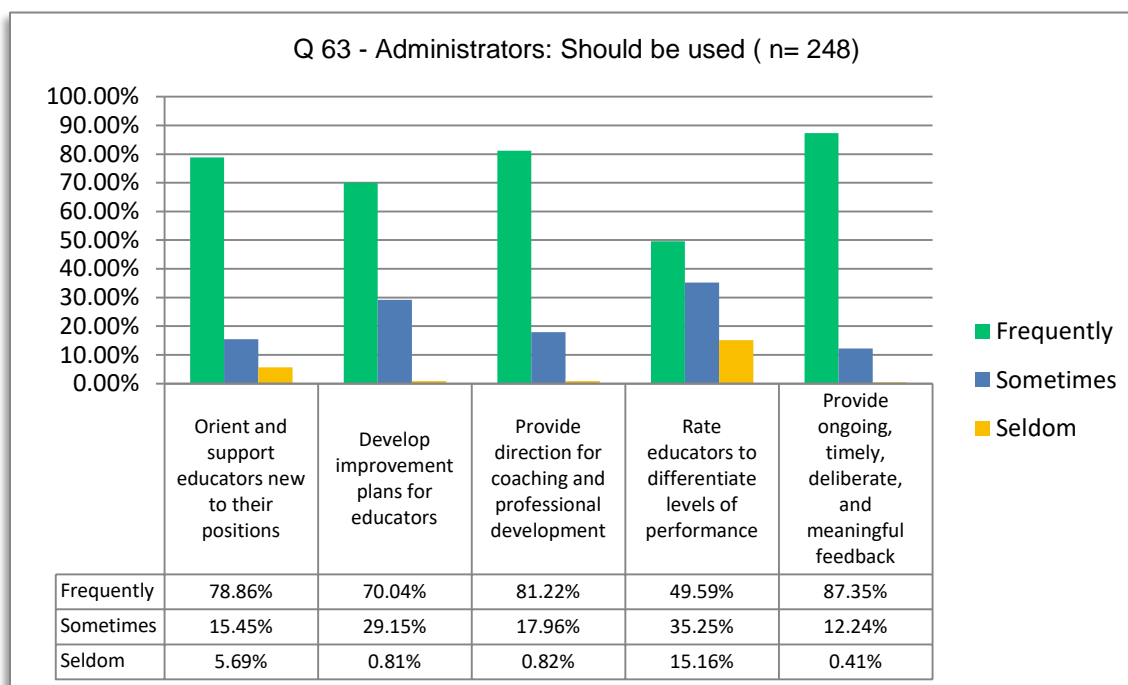
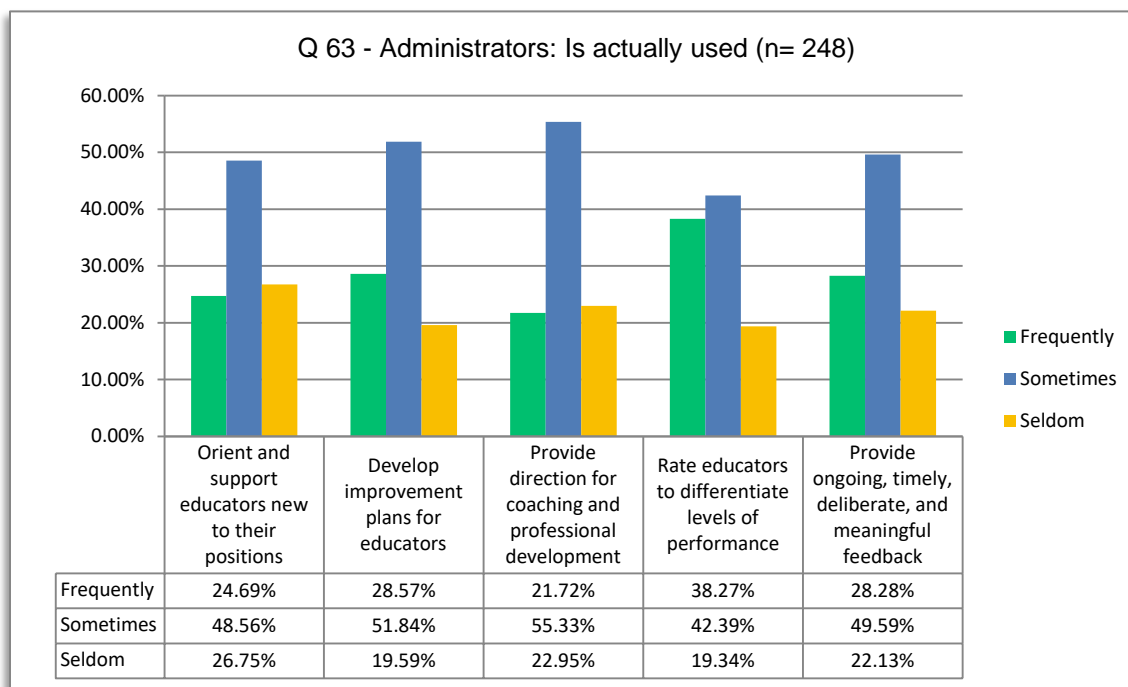
Notably, regarding use of evaluation for rating of educators to differentiate levels of performance, there was little difference in respondents' observation of the extent it should be used and their judgement of its actual use. Specifically, 36.98 percent responded that educator evaluation should be used to differentiate performance frequently, and 24.52 percent said it should be seldom used for this purpose. In respondents' judgement, actual use was frequently 36.96 and seldom 25.48 percent. For teachers, desired and actual use was much the same. When administrators were asked the same question we find that the difference between desired and actual use in each area is more noticeable.

September 30, 2017

In the following four charts for survey questions 23 and 63 both teachers and administrators are asked the frequency each of the following **should be used**; followed by frequency of **actual use**.

Figure 77: Q 23 - Teachers: **Should be used**Figure 78: Q 23 – Teachers: **Actually used**

September 30, 2017

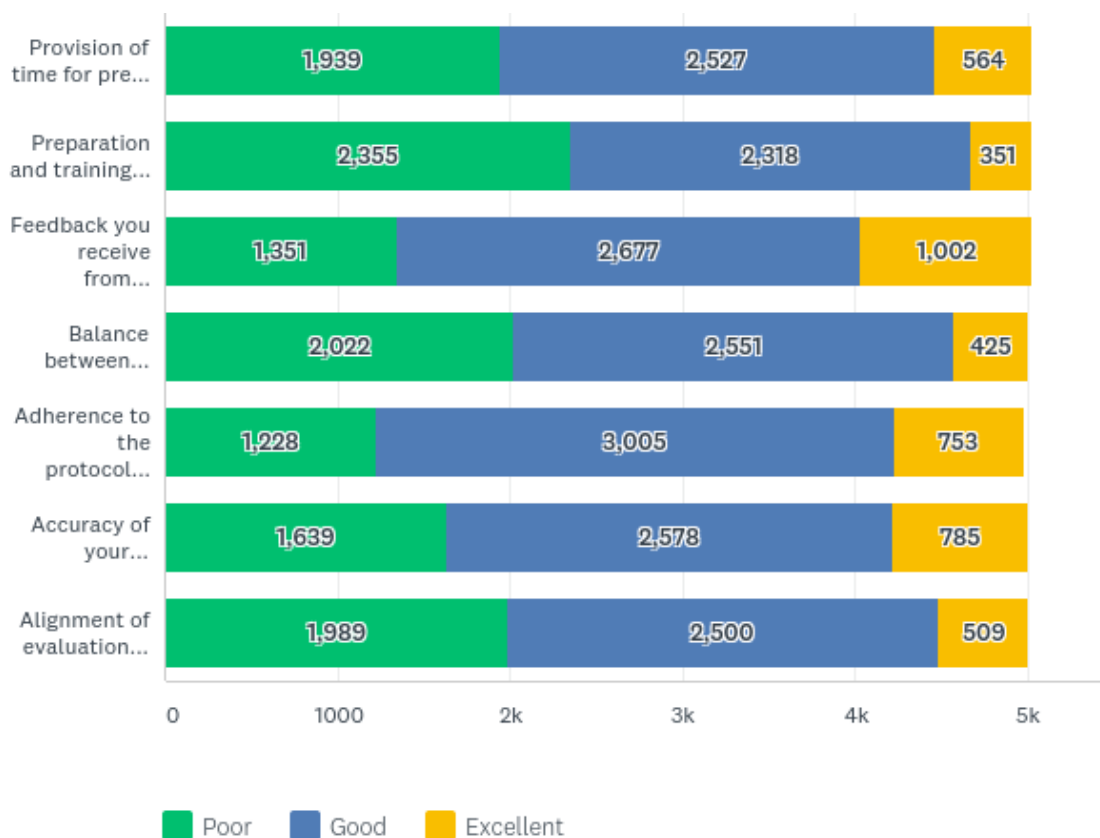
Figure 79: Q 63 – Administrators: **Should be used**Figure 80: Q 63 – Administrators: **Actually used**

September 30, 2017

As we consider the quality of the evaluation process we find that adherence to the protocols of the evaluation system the respondent is using and the quality of feedback received from their evaluator were rated as *good* or *excellent* by the vast majority of respondents. For teachers the response is 75.37% and 73.14% respectively. We find even higher ratings from administrators with 91.45% finding adherence to the protocols of the evaluation system the respondent is using *to be good* or *excellent* and 90.17% rating the quality of feedback *to be good* or *excellent*. However, we observe that 40% or more of teacher respondents found the alignment of evaluation outcomes to the teacher's subject areas (39.80%), the balance between formative and summative feedback in the evaluation process (40.46%), and preparation and training for the evaluation process (46.88%) to be *poor*. More than 60% of administrators found these same three features to be *good*. Specifically, 69.10% of the administrators rated alignment of outcomes as *good*; 70.09% of the administrators rated the balance between formative and summative assessment as *good*; and 64.96% of the administrators rated preparation and training as *good*. Once again we see a marked difference between teacher and administrator perceptions.

Figure 81: Q 25 – Teachers: Quality of process

**Q25: From your perspective, please indicate the quality of the following elements of your evaluation process. (n = 5,041)**



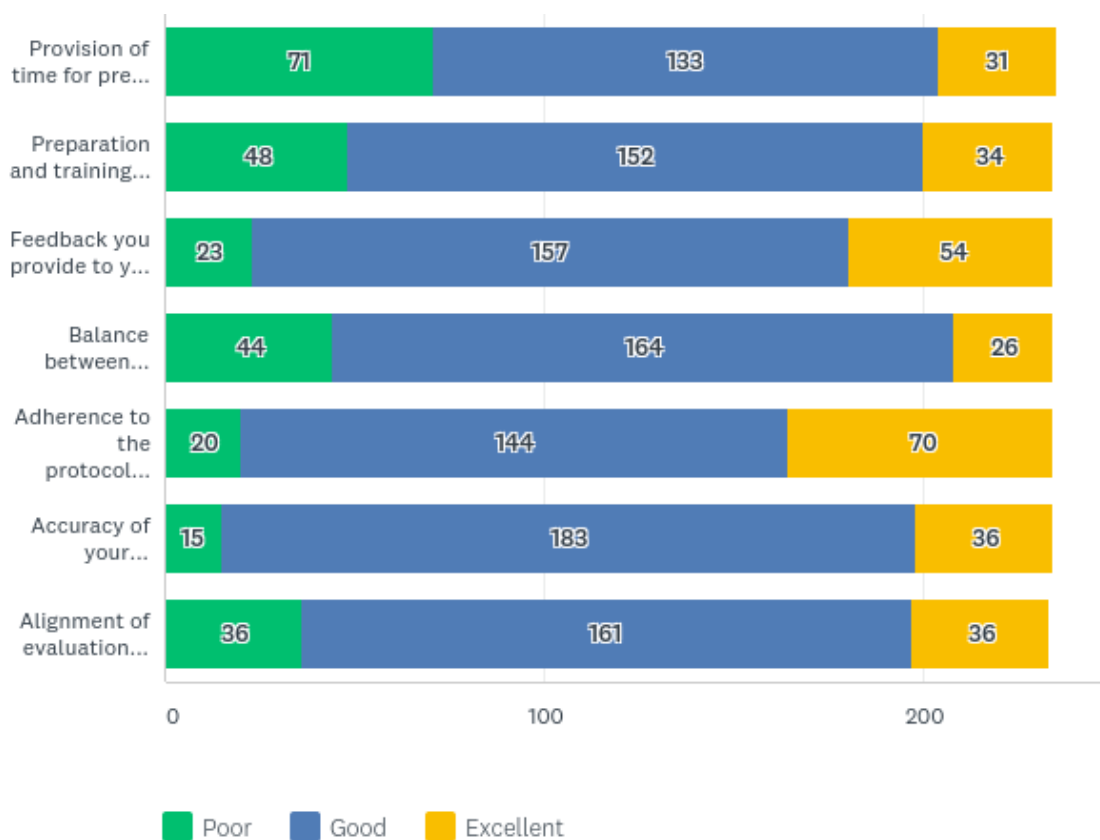
September 30, 2017

	POOR	GOOD	EXCELLENT	TOTAL	WEIGHTED AVERAGE
Provision of time for pre and post conferences	38.55% 1,939	50.24% 2,527	11.21% 564	5,030	1.73
Preparation and training of you and your evaluator for the evaluation process	46.88% 2,355	46.14% 2,318	6.99% 351	5,024	1.60
Feedback you receive from your evaluator	26.86% 1,351	53.22% 2,677	19.92% 1,002	5,030	1.93
Balance between formative and summative feedback in the evaluation process	40.46% 2,022	51.04% 2,551	8.50% 425	4,998	1.68
Adherence to the protocols of the evaluation system you are using	24.63% 1,228	60.27% 3,005	15.10% 753	4,986	1.90
Accuracy of your performance reflected in the evaluation outcome	32.77% 1,639	51.54% 2,578	15.69% 785	5,002	1.83
Alignment of evaluation outcomes to your subject area or instructional setting	39.80% 1,989	50.02% 2,500	10.18% 509	4,998	1.70

September 30, 2017

Figure 82: Q 65 - Administrators: Quality of process

***Q65: From your perspective, please indicate the quality of the following elements of the process you use in conducting teacher evaluations. (n =235)***



September 30, 2017

	POOR	GOOD	EXCELLENT	TOTAL	WEIGHTED AVERAGE
Provision of time for pre and post conferences for evaluatees	30.21% 71	56.60% 133	13.19% 31	235	1.83
Preparation and training of you and your evaluatees for the evaluation process	20.51% 48	64.96% 152	14.53% 34	234	1.94
Feedback you provide to your evaluatees	9.83% 23	67.09% 157	23.08% 54	234	2.13
Balance between formative and summative feedback in the evaluation process	18.80% 44	70.09% 164	11.11% 26	234	1.92
Adherence to the protocols of the evaluation system you are using	8.55% 20	61.54% 144	29.91% 70	234	2.21
Accuracy of your evaluatees' performance reflected in the evaluation outcome	6.41% 15	78.21% 183	15.38% 36	234	2.09
Alignment of evaluation outcomes to your evaluatees' subject area or instructional setting	15.45% 36	69.10% 161	15.45% 36	233	2.00

### *Beyond Requirements and Compliance*

Setting aside for the moment whether individual schools and districts have met legal requirements and compliance guidelines, one must ask if the policy intent has been or is being met by the system as designed and implemented to date. Is the system an impactful strategy for assuring quality education and accountability? Is the system worth the cost in time, resources and impact on organizational culture? Does it compete with other worthy goals? - The answer from focus groups and survey, white papers and the literature seems mixed. Many valued the feedback that well implemented observations and feedback the required models can produce, but many also questioned the time spent in paperwork, the lack of training, the imprecision of the value added measures and the drawing away from quality feedback and mentoring of teachers and their career development. The Brookings Report<sup>21</sup> makes a harsh conclusion...

*“... the system is spending \$1.4 billion a year to observe teachers [in the US]. This is spending a lot of money to find that nearly all teachers are effective and to generate teacher feedback that does not improve student learning.”*

<sup>21</sup> Brookings Report, Teacher Observations have been a waste of Time and Money, Mark Dynarski, December 8., 2016

September 30, 2017

***Beyond the classroom***

The major portion of this evaluation addresses questions raised by MDE and questions about the implementation and impact of the educator evaluation system as designed in accordance with PA 173. A more overarching line of inquiry can be raised regarding the systemic impact of the design and implementation of the evaluation system. Here questions about the theory of change (TOC) and alignment of the evaluation system in the context of other major contextual elements of education in Michigan can be explored. Anthony Bryc, and Pennie Foster-Fishman, et al<sup>22</sup>, and others raise the questions of systemic design and effect on effective education reform and initiatives. This view allows one to explore questions of system capacity, root causes, and other systemic issues that may act as barriers to impact. One can effectively argue that for educators' evaluation to have positive impact on teaching and learning system change must be embedded into the implementation of the intervention / model.

**Activity Question 2 – Barriers**

Question 2 posed by MDE asks for the barriers that may impact implementation of the educator evaluation system in Michigan. Barriers to implementation of the educator evaluation system and compliance with PA 173 requirements were on the minds of the educators participating in focus groups and the practitioners who wrote the white papers for this evaluation. Barriers were also revealed in our survey of Michigan educators.

The survey asked this question:

*“Each of the following may be barriers to effective implementation of the evaluation process. Based on your personal experience with the evaluation process, mark each with a value 1-5, with 1 being least significant and 5 being very significant as a barrier to an effective evaluation process.” (Q 24 and Q 64)*

**From Brookings article by Mark Dynarski, December 8, 2016<sup>1</sup>**

***“Teacher observation scores and student test scores show little correlation...evidence about teacher knowledge and practice is weakly and inconsistently related to student achievement. Observations are fundamentally about teacher practice. The finding is saying observations and test scores are measuring different things.”***

Top rated barriers identified in the survey by teacher respondents were **evaluation skills** of those conducting evaluation (3.92 weighted average), requirement for **student growth indicators** (3.86 weighted average), **time** (3.83 weighted average), the **usefulness of evaluation outcomes for teaching**

<sup>22</sup> Accelerating How E We Learn to Improve, Anthony Bryc, Educational Researcher Vol. 44. No. 9, pp 467-477, 2015 and the ABLe Change Framework: A Conceptual and Methodological Tool for Promoting Systems Change, by Pennie G. Foster-Fishman and Erin R. Watson, AM Journal of Community Psychology, July 2011



September 30, 2017

**and learning** (3.5 weighted average), and the **degree of implementation of evaluation model** (3.45 weighted average).

Top rated barriers to conducting evaluation identified by administrators were **time** (4.54 weighted average), requirement for **student growth indicators** (3.55 weighted average), requirement for all **educators to be evaluated every year** (3.46 weighted average), the **number of observations expected** to be conducted by your evaluator for your evaluation (3.23 weighted average), and the **rating system** (3.22 weighted average).

Responding to a survey question about barriers to their own evaluation administrators rated most highly **time** (3.79 weighted average), **training** (3.15 weighted average), degree of **implementation of the evaluation model** (2.88 weighted average), **usefulness of evaluation outcomes for teaching and learning** (2.83 weighted average), and requirement at all **educators be evaluated each year** (2.83 weighted average).

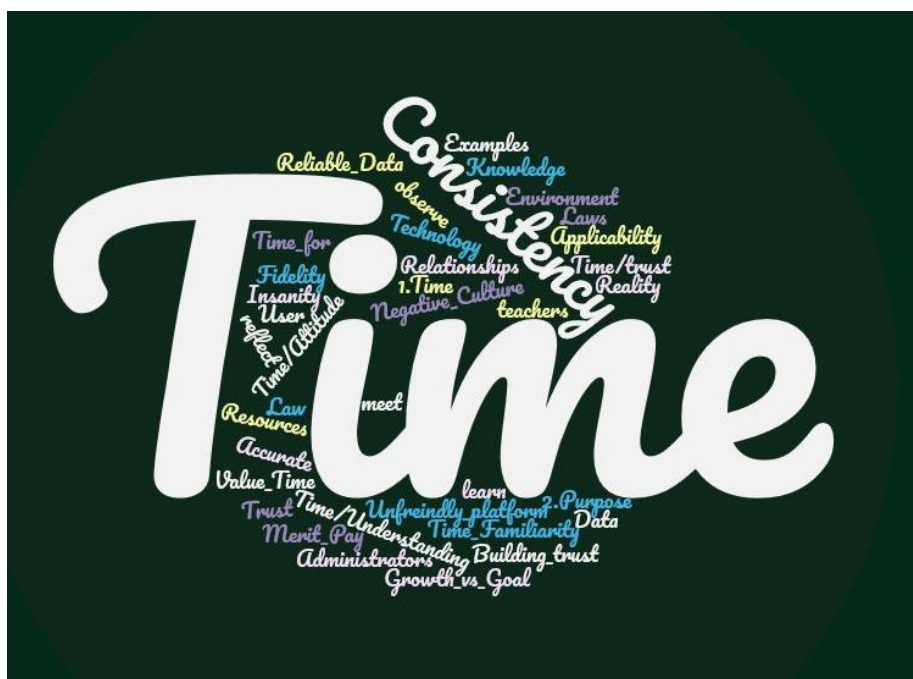
The chart below shows the top rated barriers as answered by teachers when considering their evaluation (Q 24 - columns one and two). The question was also asked of administrators regarding the evaluations that they conduct (Q64 - columns three and four), and regarding the evaluations that are conducted of them as evaluatee (Q69 - columns five and six). Note that time was rated among the top rated barriers by teachers and administrators in their role as evaluator and evaluatee.

*Figure 83: Highest rated Barriers on 1 to 5 scale (Survey questions 24, 64 and 69)*

Teachers(Q 24)	Weighted average	Administrators as evaluators (Q 64)	Weighted average	Administrators as evaluatees (Q 69)	Weighted average
Difference in <b>evaluation skills</b> among evaluators	3.92	<b>Time</b> to fulfill all requirements	4.54	<b>Time</b> to fulfill all requirements	3.79
Requirement for <b>student growth measures</b> to be incorporated into evaluations	3.86	Requirement for <b>student growth measures</b> to be incorporated into evaluations	3.55	<b>Training</b> on the evaluation model	3.15
<b>Time</b> to fulfill all requirements	3.83	Requirement that <b>all educators be evaluated</b> each year	3.46	Degree of <b>implementation</b> of evaluation model	2.88
<b>Usefulness</b> of evaluation outcomes for teaching and learning	3.55	<b>Number of observations</b> expected to be conducted by your evaluator for your evaluation	3.23	<b>Usefulness</b> of evaluation outcomes for teaching and learning	2.83
Degree of <b>implementation</b> of evaluation model	3.45	The <b>rating system</b> used in Michigan (Highly Effective, Effective, Minimally Effective, Ineffective)	3.22	Requirement that <b>all educators be evaluated</b> each year	2.83

September 30, 2017

Overall the barriers ranked highest are consistent with the barriers described during focus groups. When we asked focus group participants for one word to describe barriers to implementation they told us this.



One focus group participant described barriers succinctly:

*“Time and resources. A valid, reliable evaluation takes hours of observation and relationship building. It is often difficult for single administrators to work the process well.”*

It is worth highlighting that when considering the evaluation outcomes that both teachers and administrators receive, they describe the *usefulness of the evaluation outcomes for teaching and learning* as a high ranking perceived barrier. Similarly, focus group participants questioned if the feedback that they received was adequate.

*“The best [professional development] experience that I ever had in 20 years is ‘learning lab’ teachers from my same subject area. Both new and senior teachers met to talk about how they do what they do.” – Focus group participant*

As described earlier few survey respondents find the results of the evaluation to be impactful to their professional growth and teaching practice. These findings challenge the intent of evaluation as a vehicle for improvement of teaching and learning.

*“Evaluators need more time and practice with evaluation tools and processes to gain traction toward being competent in the evaluation process. Evaluatees need more feedback from formative processes to develop teaching and learning efficiency”. – Focus group participant feedback survey*

September 30, 2017

### *Issues related to Cultural Competence, Diversity and under-served or under-resourced schools and students*

The Request for Proposal from MDE specifically asked about issues related to cultural competence, diversity and under-served or under-resourced students and schools. In Michigan this issue can be viewed from myriad perspectives. The population of students and teachers in Michigan's schools experience diversity in race, ethnicity, urban versus rural, income disparities, languages spoken, immigration and newcomer status, religious affiliation and more. We asked about the issue during our interviews with researchers and practitioners. For some the educator evaluation system as implemented in Michigan, as in other states, has the potential for differential or disparate impact. Disparate impact may be experienced in the systemic differences in schools and districts that result in differences in teacher turnover rates, student mobility rates, English Learner populations, special education identification and services, mainstreaming and accommodation impact, high absenteeism, poverty rates, hiring practices and more. Some argue that teaching and learning are more challenging in these environments and therefore conditions need to be considered when measuring performance and growth. Others question if differences in standards, expectations, and measures of performance success are themselves likely to negatively impact and suppress student outcomes. In other words: Do student growth indicators adequately correct for these differences to produce a more valid measure of teacher impact on learning /student growth? Should they?

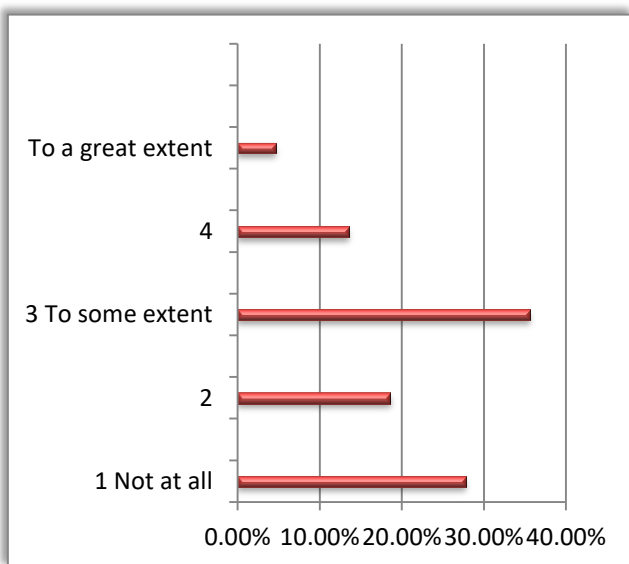


Figure 84: Q20 – Teachers: To what extent are cultural competency skills included in the evaluation model you are using?

*“Systems of formal evaluation could much more effectively foreground culturally relevant pedagogies by incorporating the practices and mindsets suggested into the evaluation rubrics, directly.” – from Ann Blais whitepaper*

Focus group participants described the differences that the population of students that they serve present, in the main citing high-poverty rural populations that have fewer in-school and out-of-school resources. The survey asked *“the extent to which your evaluation process impacts your practice in*

September 30, 2017

*addressing needs of historically under-served student populations in your classes”*. (Q 17) Responding to a five-point forced choice scale with one being *negatively impacts* and five *positively impacts*, 39.28 percent selected three for a weighted average of 2.68.

Figure 85: Impact on teachers’ practice

**Q17: Please indicate the extent to which your evaluation process impacts your practice in:**

(n = 5,501)



September 30, 2017

	1 NEGATIVELY IMPACTS	2	3	4	5 POSITIVELY IMPACTS	TOTAL	WEIGHTED AVERAGE
Providing innovative, personalized instruction and learning	14.45% 794	18.62% 1,023	40.72% 2,237	16.20% 890	9.99% 549	5,493	2.89
Fulfilling your role as a teacher	17.76% 975	18.78% 1,031	36.14% 1,984	16.90% 928	10.42% 572	5,490	2.83
Maintaining your focus on teaching and learning	18.68% 1,025	17.79% 976	33.32% 1,828	18.36% 1,007	11.85% 650	5,486	2.87
Addressing needs of historically under-served student populations in your classes	18.36% 1,003	22.60% 1,235	39.28% 2,146	12.02% 657	7.74% 423	5,464	2.68

Focus group participants also spoke of diversity in the context of serving students who receive special education services; particularly students in self-contained or center-based programs. Focus group participants also described the measurement challenges faced with English learner students and highly mobile students new to the school.

Questions related to diversity highlight the need for a systems view when considering the role of educator evaluation. The process cannot / should not be divorced from the overall needs and intent of the educational system. So as diversity represents the characteristics and needs of a system the response and measures of both individual and systemic performance must include impact related to diversity. This systemic focus should be seen in not only the evaluation measures and feedback, but also in the professional development, coaching, materials, instructional and school organizational practices – that is, all aspects of the teaching and learning environment of schools. We have not seen sufficient evidence that this is yet the case.

### Activity Question 3 – Measures of Student Growth using alternate methods/ tools and processes

Perhaps the most controversial feature of educator evaluation nationwide is inclusion of student growth indicators in the evaluation process. Discussion nationally was not only focused on whether student growth measures should be used but also on how they should be used, how they should be weighted,

September 30, 2017

which measures should be used, and how to apply statistical models to improve the reliability and validity of the measures.

Michigan Public Act 173 requires that student growth measures be a part of educator evaluations. The legislation provides for local options in selection of the measures used and how they are identified. However the legislation does require:

*“The legislation requires that evaluations be conducted annually, and that they incorporate student growth as a significant component, beginning at 25% in the 2015-2016 school year and growing to 40% in 2018-2019. Teachers and administrators with three consecutive highly effective ratings may receive biennial reviews in place of annual reviews.”<sup>23</sup>*

To understand the variety of student growth measures used in Michigan the survey asked teachers and administrators the student growth measure used for teacher evaluations.

Survey Question 29 (teachers): *What student growth measurement tools do you use in your evaluation? Mark all that apply.*

Survey Question 74 (administrators): *What student growth measurement tools do you use in the evaluations you conduct? Mark all that apply.*

Figure 86: Student Growth Measures Used

	Reported by Teachers (Q 29 n = 4797)	Reported by Administrators (Q 74 n = 205)
<b>M-Step / MI-Access</b>	32.29 %	36.59 %
<b>Student Learning Objectives (SLOs)</b>	25.39 %	24.88 %
<b>Teacher-made assessment (pre/post- tests, etc)</b>	61.48 %	61.95 %
<b>Alternative assessment purchased (NWEA, AIMSweb, ACT, etc</b>	50.91 %	65.37 %
<b>Locally developed assessment (e.g., by district, ISD/RESA)</b>	19.01 %	29.27 %
<b>Externally developed common assessment item bank (Fluence, INSPECT, etc)</b>	4.52 %	7.80 %
<b>Student growth percentile</b>	25.35 %	27.80 %
<b>Do not know /Unsure</b>	4.17 %	0.98 %

It is important to remember that only a portion of Michigan’s teachers are covered by M-Step – that is those teaching in tested subject matter and in grade levels tested. This necessitates alternate measures

<sup>23</sup> Michigan Educator Evaluations At-a-Glance. Michigan Department of Education, [www.michigan.gov/mde](http://www.michigan.gov/mde)

September 30, 2017

for a large swath of the educators in Michigan. In these cases we see educators turning to teacher-made tests, commercial assessment products, and student learning objectives (SLOs).

*“In two recent evaluations of SLO implementation, teachers reported that the SLO process provided them with the opportunity for data use, that the SLO process was an empowering aspect of their evaluation, and that they engaged in their evaluation more actively after SLO implementation.”<sup>24</sup>*

Focus group participants questioned the quality of teacher-made-tests, and expressed considerable concern for fairness related to use of M-Step as a measure of student growth. Some raised fairness issues related to those for whom M-Step data was available versus grade levels and subject matter for which there is no state data. This concern was raised in the practitioner white papers as well. Some educators described using M-Step scores as part of all classrooms’ outcome measures, even if not part of the testing cycle, as a way of distributing responsibility for student learning and fairness. Notably some focus group participants valued use of pre-post test data and teacher generated data. Others saw value in assessments developed district- and ISD-wide. Some of the focus group participants worried that the system could be “gamed” by design of pre-post assessments to guarantee student learning gains and thus more favorable evaluation. One focus group participant shared an anecdote about a

teacher whose students performed remarkably well on the pre-test, she allegedly lamented to the students “How will I ever show growth on the post-test?”

### ***Best Practices in Action***

#### ***Kalamazoo RESA, WoodsEdge Learning Center***

Educators in Kalamazoo RESA’s center-based program addressed the challenge of adapting the classroom observation rubrics of 5 Dimensions for their classrooms of students with severe impairments. Through focus groups and meetings of professional staff, the team developed adaptations to the evaluation model that apply to classroom observations and administrative feedback in evaluations.

Contact Person: Tom Zahrt, Assistant Superintendent of HR & Communication; Aubree Spencer, Principal

During focus groups unique issues were described by vocational education administrators for whom standardized student measures were not available. All things considered the administrator valued the opportunity for professional development that design and application of student measures for the career technical education instructors provided, many of whom had not been trained as teachers.

A similar concern was expressed by an administrator asked to observe and evaluate a teacher in a specialized elective area. The administrator said that they did not know about instructional practices for teaching orchestra, he had to take the teachers’ word for it. In some of the descriptions of practices teachers and administrators described peer-to-peer teams that conducted classroom observation and

<sup>24</sup> Implementation of SLOs: Recommendations for Decision Makers, Michigan Department of Education Office of School Improvement, October 2106 -

September 30, 2017

feedback. Other focus group participants challenged this as an option due to the high stakes nature of evaluation. The high stakes evaluation bred competition rather than collaboration among colleagues. To the extent that this is the case it is an unfortunate unintended outcome of the educator evaluation system.

*“States use different measures and put them together in different ways. Thus, while most schools report that they have an educator evaluation systems in place, the reality of what those systems are like – and how they are experienced by teachers – varies wildly, from occasional “walk throughs” to multiple, highly structured classroom observations conducted with trained observations conducted with trained observers, from a focus struggling teachers to a focus on schoolwide continuous improvement.” – from Suzanne Wilson’s white paper*

During focus groups concern was raised about growth indicators for students receiving special education service. Participants said that although some used Mi-Access others looked to IEPs as the growth indicator. Discussion followed that use of IEP goals could result in depressing IEPs in pursuit of growth. When considering both observation methods and growth measures one focus group participant put it this way, *“As always with special education the standard is set and we adapt.”* Educators described the challenge of implementing the observation models in self-contained classrooms and center based programs of students with severe needs (e.g., autism, severe cognitive impairment, multiply impaired). They asked: what do each of the instructional strategies and domains described in the various evaluation models look like in practice in these classrooms?

Campbell’s Law – “The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.”

Student growth is a central feature of the Michigan model for educator evaluation and much on the minds of the educators of the state. In 2017-2018 25 percent of the annual year-end evaluation must be based on student growth and assessment data; and beginning in 2018-2019 40 percent of the annual year-end evaluation was to be based on student growth and assessment data<sup>25</sup>. That threshold has recently been postponed (confirm). Nevertheless, concerns remain. Survey respondents were asked:

*In Michigan legislation regarding educator evaluation has established the weight that student growth measures will carry. In your opinion, how much weight should student growth measures carry in your evaluation?*

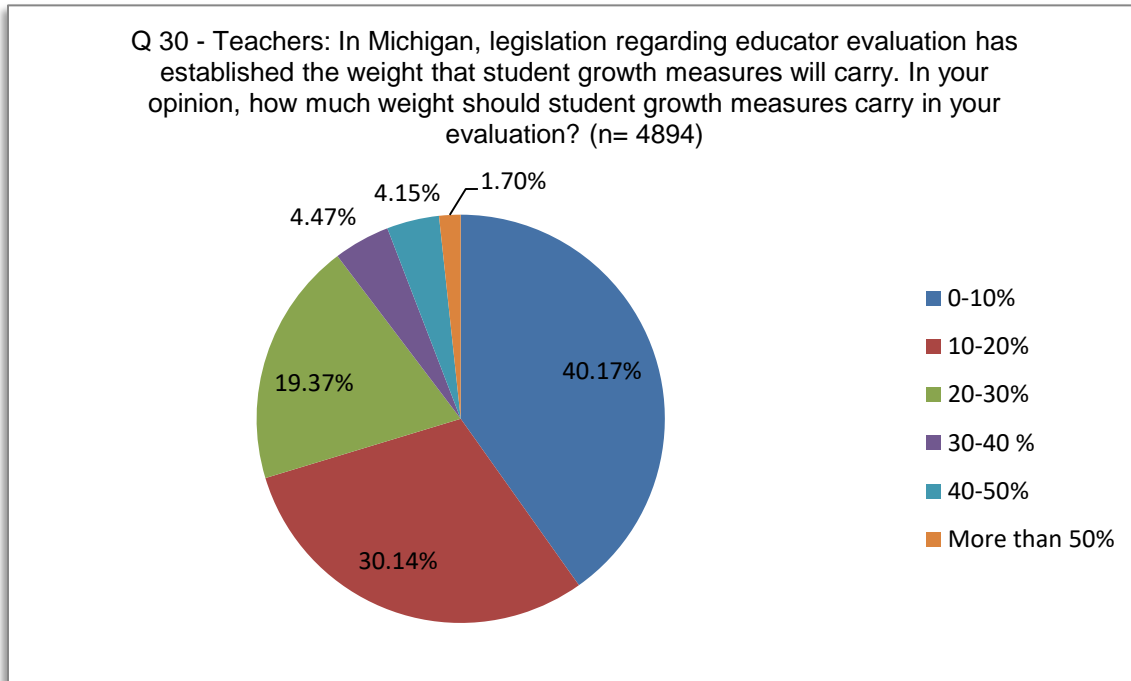
<sup>25</sup> Michigan Educator Evaluation at-a-Glance or PA 173 of 2015



September 30, 2017

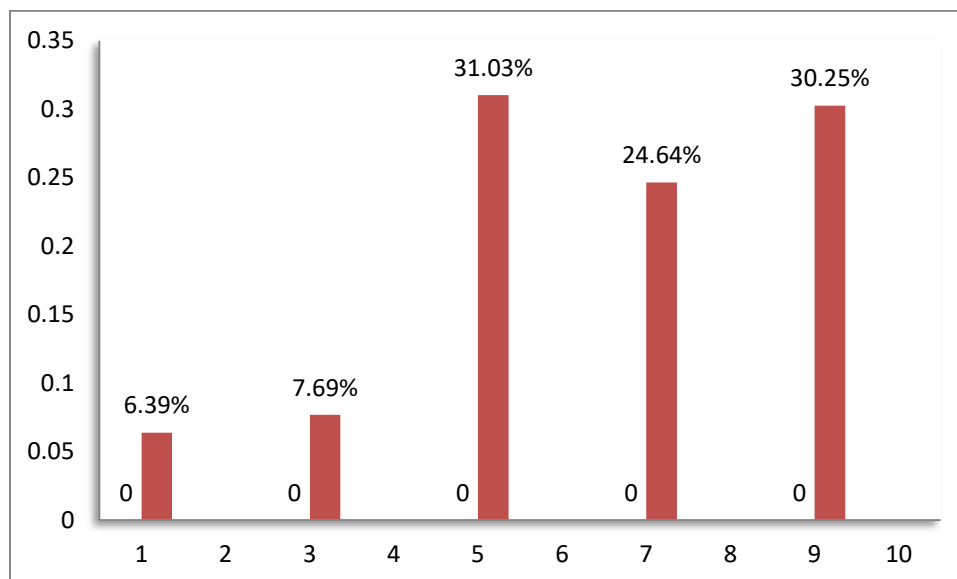
90 percent of teachers responding indicated weights of 30 percent or below.

Figure 87: Q 30 – Teachers: Weight student growth measures should carry



We later follow-up to ask, “To what extent do you know how student growth measures impact your end-of-year evaluation rating?” On a five-point scale 30 percent said they fully know.

Figure 88: Q 33 – Teachers: To what extent do you know how student growth measures impact your end-of-year evaluation rating? (n = 4853)



September 30, 2017

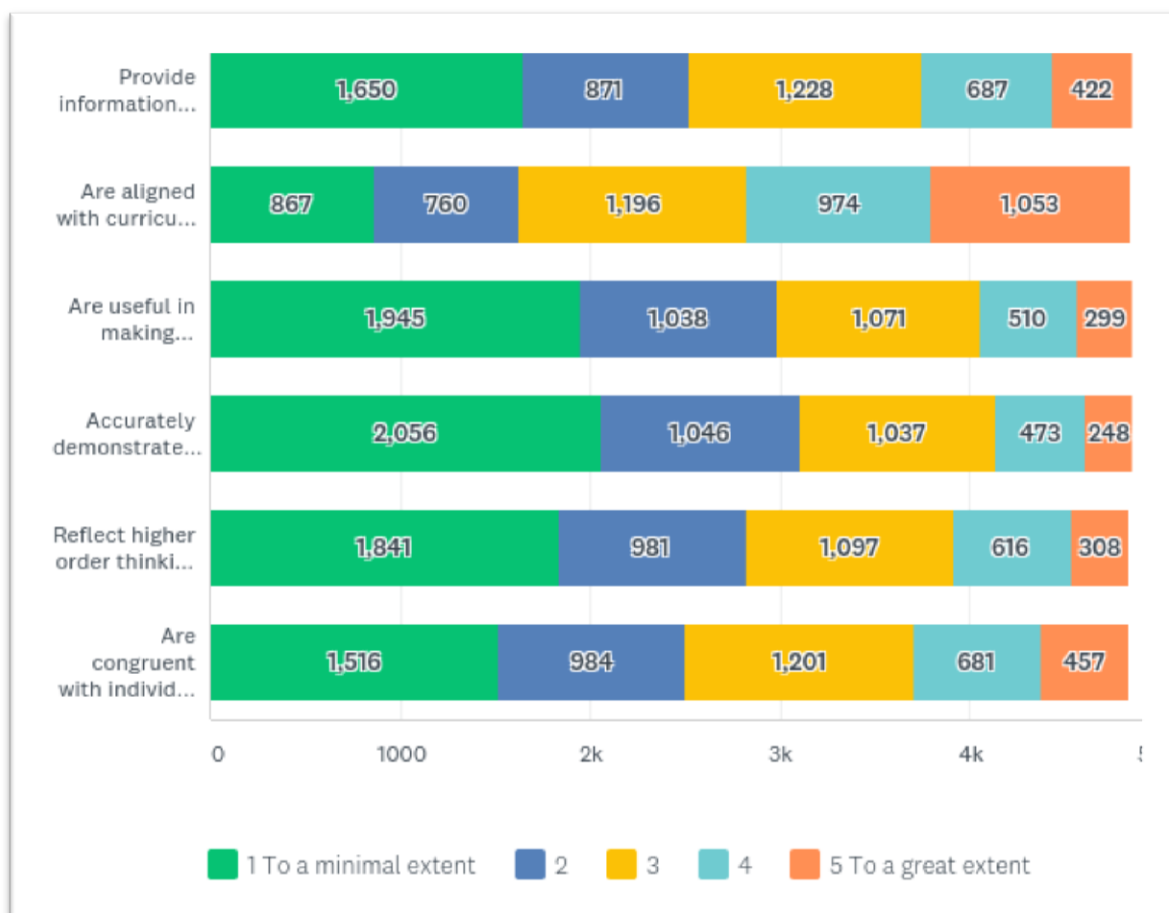
Our white papers, focus group participants and interviews raised questions about the level of understanding of student growth data and the ability of educators to effectively use the data for improvement of teaching and learning.

We wanted to learn about teachers' perceptions of growth measures in their evaluation process. More than a third of teachers responded that student growth measures provided information useful to teaching and learning. However, on a five-point scale 65 percent said student growth measures are aligned to curriculum expectations (rating 3 or above). More than 60% find student growth measures minimally useful in making judgments about their performance as an educator, and demonstrate their effectiveness in promoting student achievement (ratings of 1 and 2). Likewise, more than 50 percent of teachers rated *minimally* (ratings of one or two) the extent student growth measures reflect higher order thinking skills and problem solving skills of their students, and are congruent with individual student learning objectives.

Figure 89: Teachers: Assessment of Student growth measures

**Q31: Teachers - The student growth measures used in your evaluation process:**

(n= 4,872)



September 30, 2017

	1 TO A MINIMAL EXTENT	2	3	4	5 TO A GREAT EXTENT	TOTAL	WEIGHTED AVERAGE
Provide information useful to teaching and learning	33.96% 1,650	17.93% 871	25.28% 1,228	14.14% 687	8.69% 422	4,858	2.46
Are aligned with curriculum expectations	17.88% 867	15.67% 760	24.66% 1,196	20.08% 974	21.71% 1,053	4,850	3.12
Are useful in making judgments about your performance as an educator	40.00% 1,945	21.34% 1,038	22.02% 1,071	10.49% 510	6.15% 299	4,863	2.21
Accurately demonstrate your effectiveness in promoting student achievement	42.30% 2,056	21.52% 1,046	21.34% 1,037	9.73% 473	5.10% 248	4,860	2.14
Reflect higher order thinking skills and problem solving abilities of your students	38.01% 1,841	20.26% 981	22.65% 1,097	12.72% 616	6.36% 308	4,843	2.29
Are congruent with individual student learning objectives	31.33% 1,516	20.33% 984	24.82% 1,201	14.07% 681	9.44% 457	4,839	2.50

On a four-point scale 46 percent of teachers found that student growth measures aligned with school improvement goals *very often* or *almost always*. But more than half found that was *not usually* or *only some of the time* an accurate indication of their student growth over time, used as a basis for discussion with their evaluator, useful in planning instruction, appropriately weighted in their evaluation, or an accurate measure of their impact on student achievement.

September 30, 2017

Figure 90: Q32: Teachers - Please indicate the extent to which student growth measures used in your evaluation process are:

(n = 4,851)



September 30, 2017

	NOT USUALLY	SOME OF THE TIME	VERY OFTEN	ALMOST ALWAYS	TOTAL	WEIGHTED AVERAGE
an accurate indication of your students' growth over time	23.54% 1,140	49.74% 2,409	20.01% 969	6.71% 325	4,843	2.10
used as a basis for discussion with my evaluator	29.30% 1,416	34.75% 1,679	24.48% 1,183	11.47% 554	4,832	2.18
useful in planning for instruction	25.89% 1,250	37.61% 1,816	25.99% 1,255	10.52% 508	4,829	2.21
appropriately weighted in your evaluation	43.59% 2,097	36.25% 1,744	15.09% 726	5.07% 244	4,811	1.82
an accurate measure of your impact on student achievement	43.66% 2,111	38.10% 1,842	13.77% 666	4.47% 216	4,835	1.79
aligned with school improvement goals	18.21% 876	35.61% 1,713	31.60% 1,520	14.57% 701	4,810	2.43

*“For all measures of student growth, it is important to look at multiple years of data whenever possible. Measures of student growth are known to vary from year to year, even if the instruction is fairly stable. Whenever possible, the measures of student growth should be consistent of students at the beginning of the school year.” – from Robert Floden’s white paper<sup>26</sup>*

In Michigan recent changes in M-Step caused focus group participants to worry. They expressed fear over changing versions of the test and access to results in a manner that is useful and timely for impact on instruction. Focus group participants also worry that frequent changes in the state test negate multiple year trend data needed for decision making.

*“... [T]o the extent that the evaluation is used for educator improvement, emphasis should be on the observation system, rather than on student test data. Although Michigan legislation mandates heavy use of student test data for making personnel decisions about educators whose evaluation are extremely low, for most educators, the student test data, with their statistical adjustments, are difficult to understand and give only crude indicators of particular areas of strength and weakness. Hence they lack transparency, and probably lack rigor. Even for educators low on the scale, principals tend to be hesitant to put too much weight on student*

<sup>26</sup> For full text of white paper see appendix to this report.

September 30, 2017

*learning data, at least until they have taken steps to get teachers to improve, focusing on areas of weakness reveled through observations.” - from Robert Floden’s white paper*

Several of the articles reviewed in our literature review raised questions about implementation of VAM and other measures of student growth. Concerns are raised on both technical /psychometric and practical grounds. Rowan and Raudenbush<sup>27</sup> talk about the impact of effect size in definitions of performance ratings in VAM. Differences in a very small number of students (one or two) can make the difference between a rating of met, nearly met, or did not meet (the ratings categories used in the Rowan example). This difference can also be seen in the SLO sample by MDE<sup>28</sup>. For a standard elementary classroom of 25 or so students, reduced by students not meeting enrollment cut off, one or two students can make a difference between not met, nearly met or met.

Figure 91: Hypothetical sample: Classroom of 25 with 22 meeting enrollment standard (i.e., year-long enrollment)

Percentage of Students meeting growth targets	Number of students this represents in our sample classroom ( <i>need to round to whole number</i> )	Descriptive rating	Numerical rating
90-100	19.8-22	Exceeded	4
80-89	17.6-19.8	Met	3
70-79	15.4-17.6	Nearly met	2
Below 70	Fewer than 15.4	Not met	1

For example in this hypothetical classroom a difference of 2.2 students makes the difference between a ranking of *nearly met* and *exceeded*. Similarly the difference of 2.4 students makes the difference between *not met* and *met* rankings.

The practitioner and research communities alike ask, what are the implications of dramatic changes in rating based on a small number of students?

This difference is an example of what Rowan refers to as the fallacy of imprecision or *false precision*. Are the student outcomes data precise enough for high stakes decisions? Do student growth percentiles improve this problem of small numbers making great impact on cut scores?

*“...[W]hile there is evidence of predictive validity in the performance measurement system reformers want to use in current teacher evaluations, the evidence reviewed... suggests that any particular indicator of teacher performance (whether it be derived from classroom observations*

<sup>27</sup> Teacher Evaluation in American Schools, Brian Rowan and Stephen Raudenbush, AERA-Handbook of Research on Teaching, chapter 19, January 2016

<sup>28</sup> The Implementation of Student Learning Objectives: Recommendations for Decision Makers; Michigan Department of Education and Great Lakes Comprehensive Center at American Institutes for Research, October 2016, p 15

September 30, 2017

*or value-added measures of teacher effects on student achievement) will contain substantial measurement error and distortion.”<sup>29</sup>*

To what extent do teachers and principals deeply understand the intricacies of the appropriate use of this data and its limitation? Rowan<sup>30</sup> and focus group participants suggest that the knowledge is not deep enough.

When asked how effective student growth data in educator evaluations is, one focus group participant concludes:

*[We are at the...] “very beginning of implementation, but I think it will get better.”*

## Guiding Principles for Educator Evaluation System

The review of the Michigan Educator Evaluation System is led by a review of guiding principles. We note the design principles described by the MCEE process. These principles help to define efficacy of the design and implementation of the system. Our conversations with educators during focus groups and interviews, as well as discussions in researcher and practitioner white papers remind us of some strongly held principles about teaching and learning. Not the least of which is the dual function of teaching as both an art and a science.

**Figure 92: MCEE principles and their status today**

Guiding Principles for Michigan’s Educator Evaluation System <sup>31</sup> (Summarized)	Observations of status – based on focus group, survey and interview results and document review
<p><b>To ensure a strong and improvement-focused system, the MCEE developed the following design principles for Michigan’s educator evaluations:</b></p> <ul style="list-style-type: none"> <li>• <b>The system should support clear and rigorous expectations for educational practice in teaching and administration.</b></li> </ul>	<p>Focus group participants describe little statewide congruence in standards for teaching and administration. The distribution of effectiveness ratings gives some information regarding the rigor of the system.</p> <p>Some focus group participants questioned whether evaluators could identify the best in teaching practices if seen – Focus group participants said that there was inadequate preparation for observing and providing feedback for teaching of highly specialized courses in which the evaluator may have little knowledge</p>
<ul style="list-style-type: none"> <li>• <b>The resources—money, time, attention—required to</b></li> </ul>	<p>The current federal budget proposal calls for elimination of</p>

<sup>29</sup> Teacher Evaluation in American Schools, Brian Rowan and Stephen Raudenbush, AERA-Handbook of Research on Teaching, chapter 19, January 2016

<sup>30</sup> Ibid.

<sup>31</sup> From Michigan Council for Educator Effectiveness: Final Recommendations, July 2013, p 5-6, See document for full statements of principles.

September 30, 2017

<b>implement the educator evaluation system should be carefully balanced.</b>	Title II funding to schools, a major source of professional development (PD) funds. Both survey respondents and focus group participants describe time as a major barrier to evaluation. Time impacts conducting the observations, coaching, providing quality feedback, and, experiencing professional development and training.
<b>• The system should be designed to support efforts toward improvement rather than compliance.</b>	Improvement of teaching practice rests on several factors; trust, quality and frequency of feedback, link between feedback and effective instructional practices, mentoring /coaching for teacher continuous professional development /development of practice
<b>• The system should develop local capacity to improve teaching and learning at all levels and in all LEAs.</b>	What is needed for this – consistency, inter-rater reliability, alignment of curriculum, aligned professional development system for improvement
<b>• The system for evaluating educators' performance should be based on multiple measures.</b>	Some focus group participants described the models as overwhelming in the number of components reviewed. Some questioned if there was genuine understanding of the assessment results used for measurement and their impact on teaching and learning.
<b>• The system should be designed to improve teaching and learning in the state, and structured to support ongoing educator learning and development.</b>	Opinions here vary with some describing the process as compliance driven with little information that improves teaching and learning / teaching practices. The link between evaluation outcomes and teacher training, professional development and coaching is not apparent.
<b>In addition, the MCEE agreed on five specific parameters necessary in order to ensure that the system contributes to improvement of instruction and student learning:</b>	
<b>1. A rigorous educator evaluation system...is dependent on clearly articulated student standards.</b>	The goals for student learning in Michigan continue to develop. Educators in the focus groups and white papers questioned whether there is adequate continuity to support effective evaluation
<b>2. Similarly, an educator evaluation system that can support educational improvement depends on the articulation of clear standards of professional practice.</b>	Focus group participants and white papers cite lack of standards of professional practice as barrier in Michigan
<b>3. To support improvement, educator evaluation must provide specific substantive feedback that is accompanied by targeted and specific professional learning opportunities</b>	Focus group discussion suggests that this principle has not yet been achieved. The participants questioned the quality of feedback. Some described feedback as more focused on compliance than useful professional conversations. Lack of timely feedback, and superficial feedback were cited as barriers to effectively using the system to improve teaching and learning. No evidence was seen of a link between professional development and evaluation results on individual, schools or district levels. More review is needed here.



September 30, 2017

<p><b>4. Individual educators' evaluations must be treated as confidential personnel information. The goal is improvement, not embarrassment.</b></p>	<p>There is little evidence that educator evaluations are used to guide grade level, school level, or district level professional development. Similarly there is little description of individual teacher centered PD that emerges from individual teacher evaluation results. Districts are required to notify parents of children who have been assigned to a classroom taught by teachers who has received a rating of ineffective (IE) more than two years in a row.</p>
<p><b>5. Educator evaluations must not be used to determine merit pay.</b></p>	<p>There is some evidence of merit pay systems for Michigan educators but we see no research or evaluation regarding their effectiveness for improvement of teaching and learning. [Lenawee ISD model]</p>

### Example of Best Practice-

#### Missouri's Essential Principles of Effective Evaluation Systems

**Principle 1:** Performance of educators is measured against research-based, proven expectations and performance targets consistent with the improvement of student achievement.

**Principle 2:** Multiple ratings are used to differentiate levels of educator performance.

**Principle 3:** A probationary period of adequate duration is provided to ensure sufficient induction and socialization through developmental support for new teachers and leaders.

**Principle 4:** Measures of growth in student learning across two points in time are included as a significant contributing factor in the evaluation of professional practice at all levels.

**Principle 5:** Ongoing, timely, deliberate, and meaningful feedback is provided on performance related to research-based targets.

**Principle 6:** Standardized, periodic training is provided for evaluators to ensure reliability and accuracy.

**Principle 7:** Evaluation results and data are used to inform decisions regarding personnel, employment determination, and human resource policies such as promotion, retention, dismissal, induction, tenure, and compensation.

### Why then How

To adequately address questions about the effectiveness of educator evaluations one must be reminded why the strategy was adopted in the first place. As noted earlier in this report educator evaluation is frequently cited as a central component to accountability structures for K-12 school systems and as a school improvement strategy. Our review of the literature, white papers and focus group results causes us to ask if these two goals may be at odds.

Words like *accountability* and *evaluation* have multiple definitions and can carry much weight in contemporary K-12 education circles. What is meant by accountability? For some accountability means to make judgment, rank, or assess. To be accountable therefore in this definition the K-12 system must screen, sort and select among educators. Others describe educator evaluation as a means of quality assurance. How does/to what extent does the Michigan educator evaluation system advance accountability? – Impede accountability? Are the models used for educator evaluation designed for or effective in supporting accountability?

September 30, 2017

What is the **theory of change** that underlies the educator evaluation system and strategies of the Michigan model for evaluation and legislation? Is this theory of change upheld by practice? For instance is the working purpose of evaluation in schools and in school systems to differentiate between effective and ineffective teachers or to support human resource decisions of reduction of workforce, termination, transfer and promotion? Why do school people conduct evaluations? What beliefs and assumptions do they bring to the process? Do these beliefs and assumptions make the evaluation system ineffective?

Schools nationwide and internationally (as seen in *Educational Researcher series*, May 2017<sup>32</sup>, and the NCEE report) exist in a context of calls for improvement and reform. These reform protocol, best practices, and accountability mandates need to be aligned and coherent for the education system to truly approach success. The question raised here and supported by interviews, focus groups, survey results and literature review; whether the reforms and education evaluation system in Michigan are adequately/sufficiently aligned and coherent. Hiebert and Stigler<sup>33</sup> conclude that U. S. system or policies focus on improving **teachers** rather than a systemic view or even focus on improvement of **teaching**. This focus leads to the defining component of accountability to be the ranking, screening, sorting and selection of teachers (i.e., human resource functions).

The overall model is built on assumptions about how teaching and learning, the system of schooling, and staffing operate and thus the impact of the evaluation process on that system. Are these assumptions valid? Are they universal? In what cases do they not apply?

To what extent does the theory of change that underlies the educator evaluation model reflect a systemic change model? Is systemic change included in the theory of change?

<sup>34</sup> When considering effective systemic change models successful implementation requires – readiness, capacity, diffusion, and sustainability. These features seem to be evident in varying degrees in the Michigan educator evaluation system.

### *General conclusions about expertise*

- **Elite performance is the result of at least a decade of maximal efforts to improve performance through an optimal distribution of deliberate practice**
- **What distinguishes experts from others is the commitment to deliberate practice**
- **Deliberate practice is**
  - ❖ **An effortful activity that can be sustained only for a limited time each day**
  - ❖ **Neither motivating nor enjoyable – it is instrumental in achieving further improvement in performance**

**From Dylan Wiliam, Institute of Education, London UK, 2010 <sup>1</sup>**

<sup>32</sup> Educational Researcher, p 154, May 2017

<sup>33</sup> Ibid. p 154,

<sup>34</sup> Foster-Fishman, op cit. p. 5

September 30, 2017

One can argue, as Hiebert and Stigler<sup>35</sup> do, that change is unlikely to occur through implementation of high-stakes evaluation systems that assume that teachers are neither working hard enough, or are not adequately motivated to teach effectively. They state: *“for the vast majority of teachers, improvement is a learning issue, not a motivation or effort issue”*<sup>36</sup>. They argue for assessment and evaluation models that feature information that teachers can use to improve. Meaningful monitoring and feedback with accompanying professional development and coaching are the most impactful to improvements in teaching and learning.<sup>37</sup> The frequency, timeliness and quality are critical to effectiveness of observations.

***Focus group and interview participants described a national landscape for education that leaves teachers feeling vulnerable and that their profession is under attack.***

Research based models produce strongest sustained improvement feature collaborative learning and feedback as provided by Communities of Practice, Professional Learning Communities, peer mentoring and collaborative teaching. We heard during focus groups that such close, collaborative and trusting relationships among colleagues are compromised when the school and district culture skews toward the competitive culture epitomized by rating systems. Models based on how adults learn and how organizations successfully implement change describe opportunities to collaborate and to try-out and receive feedback on new routines. To effectively learn and change, these models would argue, takes trust. Trust, according to focus group and survey respondents, is compromised by high stakes ratings systems. Such a high stakes system may have an unintended outcome of amplifying the vulnerability that some educators feel and impeding school improvement. For student achievement to be the primary focus, a mindset that values implementation and growth over ratings should guide local district and state policy making cultures.

*“Teachers need to be able to see what highly effective teaching looks like.” Focus group participant*

#### *The goal – support, not sanction*

When reviewing high-performing countries the National Center on Education and the Economy (NCEE) found that *“The goal of these teacher appraisal systems tends to be to provide information to help teachers improve their performance, rather than to identify and sanction low performers. ...These systems, by and large, do not expect to fire teachers later.”*<sup>38</sup>

<sup>35</sup> Hiebert and Stigler 2017 Educational Researcher ...

<sup>36</sup> Borko, 2004; Darling-Hammond & Sykes, 1999

<sup>37</sup> Hiebert and Stigler, op cit. p 173

<sup>38</sup> Policy Brief; Empowered Educators – How High-Performing Systems Shape Teaching Quality Around the World, National Center on Education and the Economy, 2016, p1

September 30, 2017

In her white paper Suzanne Wilson<sup>39</sup> tells us:

*“Even if the majority of Michigan’s teachers are effective, the preponderance of effective and highly effective scores meant that there was little specific and concrete guidance to help even the state’s best teachers continue to improve. When asked about 2011 reforms, one policymaker noted that the intent ‘was never about punishing teachers... [but] about improving performance.’ But improving performance means identifying growth areas, which in a high stakes policy environment can quickly be misinterpreted as ‘weakness’. Thus, here one sees the clash between a system intended to identify weak teachers versus designing a system to support good teachers as they continue to grow and learn.”*

This contrasts with the experience described by some of the focus group participants who reflected on their experiences and described need for the performance system to be supportive and not a **“got-ya”**.

A number of times during the focus groups and interviews we were reminded that Danielson and Marzano claim that their models were not intended for the high stakes professional evaluation and human resources decision making purposes for which they are now being used. In an April 18, 2016 commentary published in Education Week, Charlotte Danielson writes:

*“I’m deeply troubled by the transformation of teaching from a complex profession requiring nuanced judgment to the performance of certain behaviors that can be ticked off on a checklist. In fact, I (and many others in the academic and policy communities) believe it’s time for a major rethinking of how we structure teacher evaluation to ensure that teachers, as professionals, can benefit from numerous opportunities to refine their craft.”*

### *Standards of professional practice*

Standards in education are commonly talked about in two ways – curriculum standards and standards of instructional practice – in other words, content versus pedagogy. Both are central to educator evaluation systems, and both need to be clearly articulated for evaluation to be meaningful, fair and likely to support and improve teaching and learning. When considering curriculum standards specifically some in our focus groups and interviewed worried that curriculum standards may vary widely across school districts. Variance in curriculum standards across districts can impact the content validity of student growth measures. More telling is the weak link between SLO and other measures of student growth and M-Step progress. Weak alignment of the curriculum upon which assessments are based can result in scattered outcome results. At the root an effective educator evaluation system must effectively assess both what is taught and how it is taught.

---

<sup>39</sup> Full text of the white papers can be found in the appendix to this report.

September 30, 2017

Rowan, Wilson and others ask: where are the standards for good teaching? Common standards are essential to fair and effective evaluation, and focused professional growth. Commonly understood standards are a hallmark for effective and fair appraisal systems as evidenced in the principles from MCEE report and Missouri’s Essential Principles of Effective Evaluation Systems.

A critical question asked in focus groups, interviews and whitepapers is: do the evaluators recognize good teaching when they see it? The quality of feedback and validity of observation response is based on a belief that educators can identify best practice as they observe it, and similarly recognize the absence of best practices if they do not exist. The quality of the observation and feedback provided the educator evaluated are essential features cited in the literature, white papers and by focus group participants. Focus group participants who conduct evaluations described the challenges presented in cases of specialized classes where the evaluator may have little content knowledge or content specific pedagogical skill, for example orchestra. “We have to rely on the teacher’s word”. Administrators also describe the challenge of time that forces them to revert to checklist style feedback to *just get it done*.

One may ask if standards of effective practice are clearly stated, well known, consistent over time, and uniform across districts. Should they be? Going further we may also ask for the evidence that the standards or practice result uniformly in improved student performance on the measures used to assess teacher effectiveness. This question is posed in the literature in articles by Daniel Weisberg, et. al.<sup>40</sup>

What is not seen in this evaluation review is analysis of the actual content of the feedback given through the evaluation process. Further analysis should ask how specific the feedback given to educators is and how tied to student growth and best instructional practices, etc. Some of the models used by districts give protocol for providing feedback in the form of wonderings and “look-fors”. Others gravitate more closely to a checklist approach. Based on survey results educators say that the feedback they receive is not very useful to improving their teaching practice. Educators also report that evaluation results not well tied to professional development and growth. A close look at ongoing and end-of-year evaluation feedback and

### *Best Practices in Action*

#### *Oscoda Public Schools – Oscoda Area High School*

**The district builds trust through seeking input, providing immediate feedback after walk-throughs, and collaboration on the evaluation process. There is an Individual Development Plan for each teacher. The district is expanding use of SLOs and working to align multiple measures of student growth.**

**Contact Person: Scott Moore, Superintendent; Terry Allison, Principal**

<sup>40</sup> See “The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness”, 2009, The New Teacher Project, Daniel Weisberg, et al

September 30, 2017

discussions of evaluators and evaluatees may give insight into why there is such disconnect between the perceptions of usefulness of evaluation perceived by teachers and administrators.

### *Cycles of Professional Growth and Change*

To what extent is Michigan’s evaluation system (or the evaluation system of any state for that matter) designed for systemic and individual improvements over the professional life of a

#### *Best Practices in Action*

##### *Reading Community Schools, Reynolds Elementary School*

**The elementary school principal meets with each teacher to generate and reinforce positive guidelines and implementation of the evaluation model. The collaborative process creates a productive atmosphere and increases opportunities for quality of feedback given to teachers.**

**Contact Person: Chuck North,  
Superintendent; Dennis Ireland,  
Principal**

teacher, faculty, school and school system? As the system changes taking on new goals and mandates, new curriculum, outcomes, resource allocation, system and school leadership, governance, and public support and sentiment how does the evaluation system adapt and reflect these changes? Is the system as designed, or implemented nimble enough?

Despite popular criticism that schools are relics of the past and are unlikely to experience change, one hears descriptions from the focus groups, white papers and literature review of the relentless churn experienced in schools – changing administrators at the school and building levels, changes in board governance leadership, changes in classroom practices, changes in curriculum and assessment, changes in expectations and adoption of the NEXT NEW THING. The churn may be cited as reason for failure of the system to adequately focus or adapt to meaningful change, or may be cited as the root to the resistance heard in educators who complain of each innovation and mandate that *this too shall pass*.

This steady churn speaks to the need for improved inter-rater reliability among evaluators and across evaluation cycles. Training and calibration will need to be continuous given the rapid turnover in educators performing evaluations. The new administrator may bring new expectations, new models for what she looks for as best practices and differences in expectation of the normative rating of teachers. One focus group participant described it this way...

*“With our administrator change our former administrator was a **got-ya**... Now there is more of a focus on improvement – walkthroughs are more authentic”.—Focus group participant*

As the teaching workforce changes and becomes more or less senior, trained in specific methods, content, organizational values and culture, pedagogy and variation in pre-service skill development, one may ask how are these reflected in the evaluation design and methodology. Is the pre-service preparation of new, mid-career and senior teachers aligned with evaluation methodology and tools? Do

September 30, 2017

they reflect changing emphasis on methods like student directed learning, constructivism, project based learning, or other methodologies that may be seen in teacher instructional practice? How are these aligned and adaptive to variations, evolution and change?

As individual teachers progress from novice teacher to mid-career tenured teacher, and on to senior teacher, how do the methods and tools of evaluation adapt to meet current needs of individuals? How do they meet the needs of a faculty with individuals that fall along the full continuum of professional maturity? Concurrently, how do they adapt as the faculty evolves, and as the evaluator (principal) evolves? Should tools, methods and expectations for the novice evaluator be different than for the experienced? Do transforming / reforming schools require different evaluation methods than high performing?

*“In 21 years of teaching the evaluation is not what has made me a better teacher. It has been professional development. Not just one-day – that is what encourages me to be a better teacher.” – Focus group participant*

Is the evaluation system as designed and implemented sufficiently comprehensive to accommodate and inform the full cycle of career growth of educators from preservice, to novice, to emerging to experienced veteran? How do feedback and rating systems reflect this professional progression?

Research based standards caution that professional evaluation for human resource decision making purposes should not be intertwined with professional feedback for professional growth<sup>41</sup>, yet here we are with the same tools and protocol for human resource evaluation and professional growth and feedback. From a truly pragmatic and coherent academic environment perspective it is hard to see how these two functions cannot be mixed.

### ***The Perceived Cost of Action***

One cannot underestimate the impact of the fear of litigation experienced by school system leaders. Wrongful termination complaints and resulting legal fees and negative publicity can lead to extremely cautious decision making. There is also the real possibility that the educator found to be ineffective will indeed not be terminated in the end, thereby exposing the evaluator and causing great expense to the system. Some determine that it may be better for the organization and the educator to allow a struggling teacher to voluntarily leave the system or to counsel them out. These practices serve to depress the number of teachers rated as ineffective. However there are no known statistics to verify this popular belief. We do know that for whatever the reason 40 percent of teachers leave the profession in their first two years<sup>42</sup>. School systems have commonly followed a practice of progressive discipline whereby in the exception of an egregious act, the road to discipline is slow and contains several steps of documented feedback, direction and opportunity for corrective action. In some districts this progression may include voluntary or involuntary re-assignment with change in school, grade level or even subject

---

<sup>41</sup> Op cit MCEE report

<sup>42</sup> Beginning Teachers: Are They Still Leaving the Profession? Leslie Marlow, Duane Inman & Maria Betancourt-Smith  
The Clearing House: A Journal of Educational Strategies, Issues and Ideas Vol. 70 , Issue 4,1997



September 30, 2017

matter taught. These common practices along with litigation require extensive documentation and compliance with board of education, human resources and contract policies, processes, timelines and state laws. This reality may in some cases pose negative incentive to conferring a low rating upon an educator. Low ratings can open the door to challenges to a complex system of policies, rules, and timelines. The “safer” route may be to avoid the lowest rating, and to counsel the educator away from the school, the school system, or the profession as a whole. Similarly, some educators anticipating a low rating voluntarily seek employment at a school or system that they judge to be a better fit professionally. The result is high turnover in the first few years in the profession and low ineffective rates.

Perhaps in contrast to popular opinion few of the teachers or administrators in our survey rated contract provisions among the top of their list of barriers. Teachers’ ratings were 2.37 weighted average and administrators’ ratings were 2.46 weighted average on a five-point scale with a rating of one indicating of *not a significant barrier* to and five *a very significant barrier*. The rating represents the lowest choice made by teachers and the second lowest choice among administrators. See Figure 83 of this report for the chart showing the responses to questions 24 and 64.

### Rating Systems

In the focus groups and survey responses we see attention paid to the education evaluation model as a system of ratings. Much of the early media and research reports focused on the distribution of teachers judged to be in the highest category. We see this concern in fields other than education, but in education the discussion has led to state-wide adoption of the evaluation system. Two contrasting views are expressed; one the need for a method that reduces the arbitrary and subjective nature of evaluating educators. The second is distaste for emphasis on rating systems that judge teachers as highly effective, effective, minimally effective, and ineffective. A concern raised in performance assessments no matter the field is the proportion of those evaluated rated in the highest terms. In *The Widget Effect*, Daniel Weisberg et. al. describes this phenomenon on a national scale – state by state. In the 2017 *Revisiting the Widget Effect*<sup>43</sup> we find Michigan with 2.4% of its teachers rated in the lowest category (lower than the national average for that rating). New Mexico by contrast is an outlier with 28.7 rated in the lowest category. The majority of states find fewer than 4 percent below effective.

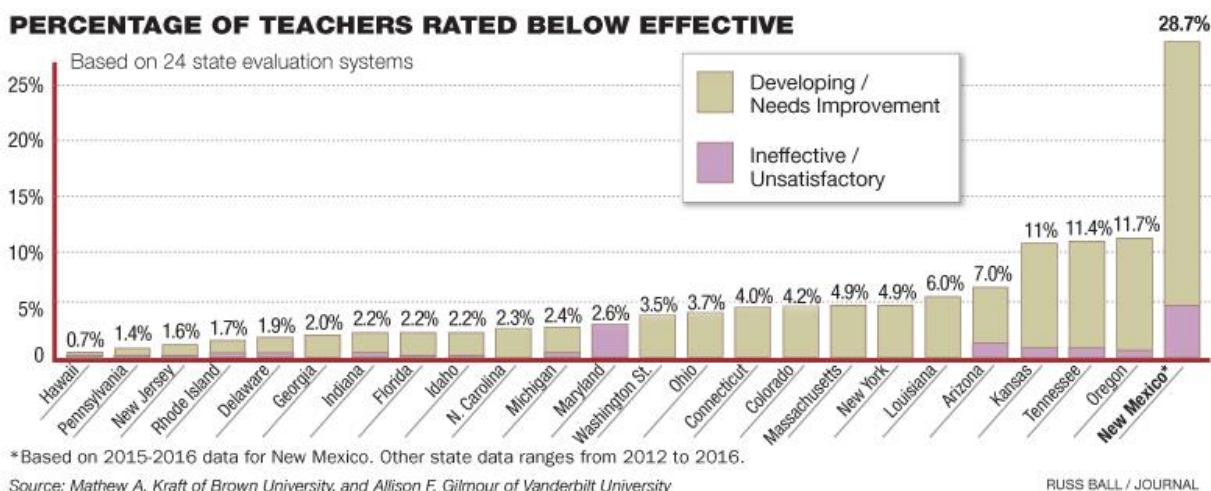
---

<sup>43</sup> *Revisiting the Widget Effect: Teacher Evaluation Reforms and the Distribution of Teacher Effectiveness*  
January, 2017



September 30, 2017

Figure 93: Multi-state review of teachers rated below effective



From Albuquerque Journal 2017

Matthew Kraft of Brown Univ. says *“The New Mexico system is very different than others” ...“Tough’ would be one way to describe it.”*<sup>44</sup>

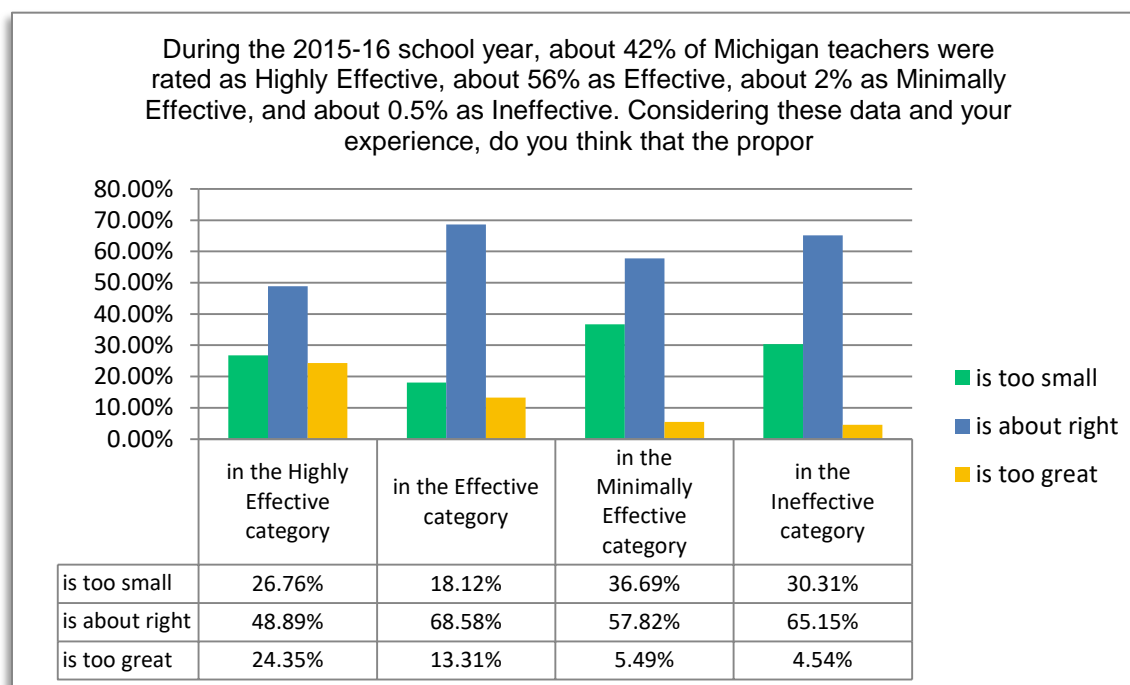
We asked educators if the distribution of ratings seen in Michigan in 2015-2016 was about right. Survey respondents said, in the main, that ratings are about right.

Prompted by similar data as seen by the *“Revisiting the Widgets Effect”* report Michigan survey respondents say that the number of teachers rated as minimally or ineffective was *too small* was between 30% and 37%; and a majority of respondents, 58% to 65%, found the percentage of teachers rated in the lowest two categories to be *about right*.

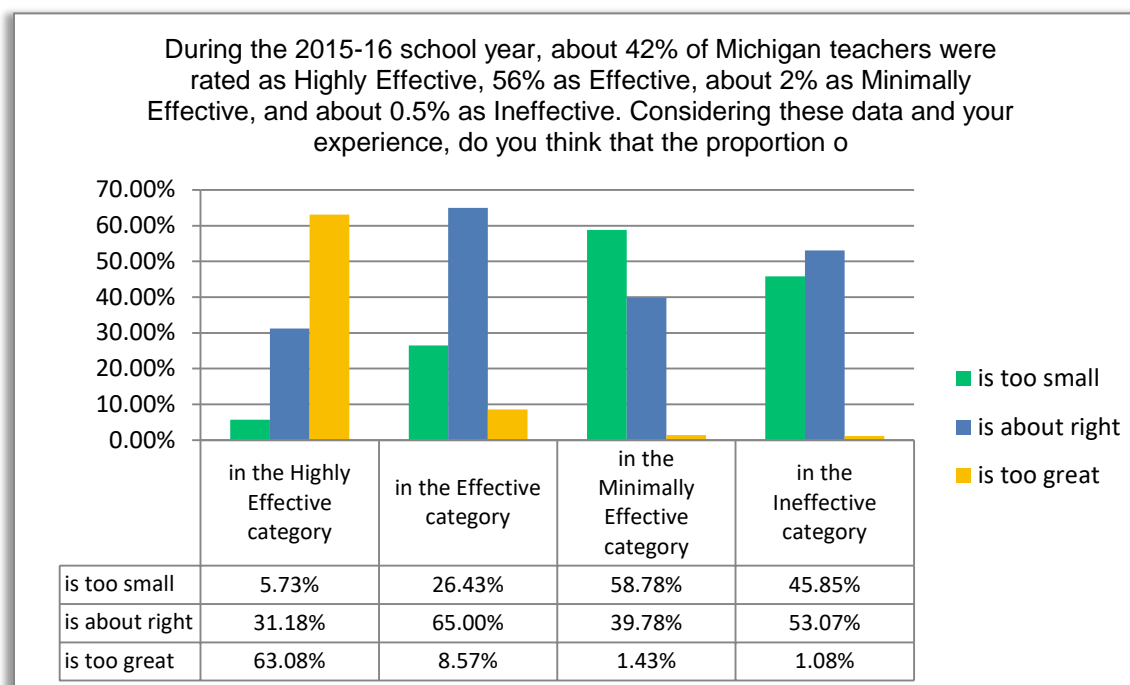
<sup>44</sup> Op cit *“Revisiting the Widget Effect: Teacher Evaluation Reforms and the Distribution of Teacher Effectiveness”* January, 2017

September 30, 2017

Figure 94: Q 22 Teachers – Perception about teacher ratings



When asked if their personal evaluation rating was right respondents said ...

Figure 95: Q 54 – Administrators: Perception about **teacher** ratings

September 30, 2017

Michigan's rating of teachers in the categories below effective has inched downward the three years that MDE has reported data.

Figure 96: Distribution of teacher effectiveness ratings, 2011-2012 vs. 2012-2013, vs. 2013-2014

Teachers	2011-2012 (N = 97,915)	2012-2013 (N = 98,616)	2013-2014 (N = 95,885)
<b>Highly effective</b>	22.6%	32.6%	37.9%
<b>Effective</b>	73.3%	64.5%	59.3%
<b>Minimally effective</b>	3.3%	2.4%	2.3%
<b>Ineffective</b>	0.8%	0.6%	0.5%
<b>Combined minimally effective and Ineffective</b>	4.1%	3.0%	2.8%

Similar patterns are seen for principal and assistant principal ratings.

Figure 97: Distribution of principals' and assistant principals' effectiveness ratings, 2011-2012 vs. 2012-2013, vs. 2013-2014

Principals and Assistant Principals	2011-2012 (N = 4,431)	2012-2013 (N = 4,524)	2013-2014 (N = 4,645)
<b>Highly effective</b>	20.9%	26.7%	29.6%
<b>Effective</b>	74.4%	70.1%	67.1%
<b>Minimally effective</b>	3.9%	2.7%	2.9%
<b>Ineffective</b>	0.8%	0.5%	0.4%
<b>Combined minimally effective and Ineffective</b>	4.7%	3.2%	3.3%

Focus group discussion shed more light on the conflicted views about ratings. *"I just wish that they would get rid of HE (highly effective – the highest rating possible). Just have two ratings. ..."* Another participant countered with *"Don't take away Highly Effective. That is what I aspire for."*

But others worried that if they accepted effective as the new normative rating they could not be fairly compared when competing for a job with someone from a district where Highly Effective was the norm.

September 30, 2017

Several factors serve to create forces to drive up ratings. By identifying educators as highly effective the district could evaluate highly effective educators less often, thus reducing the pressure to complete all evaluations each year. This could conceivably create an incentive to find more educators highly effective. Second, as districts are urged toward transparency high numbers of educators rated ineffective as compared with competing schools and districts may be seen as disadvantaging a district. Local district cultures also serve to entrench ratings of highly effective. If an educator and their colleagues have always in the past been highly effective they expect the rating to continue. Even worse, educators hold a common much stated belief that everyone knows who the poor performers in their school are. If they see these bottom tier educators rated effective they think... *I am better than that, I should not be grouped with this low performer so I deserve a higher rating than they get.* The result is to push ratings higher and higher.

Questions of fidelity of implementation reoccur throughout our review of literature, white papers, focus groups and interviews. Whether implementation is purely compliance, the result of fear of sanctions and “Got-Ya”, or to truly promote improvements in teaching practices and learning is

to be seen.

**The human face of policy – Street-level Bureaucrats - “policy implementation in the end comes down to the people (the street-level bureaucrats) who actually implement it”. Michael Lipsky**

Michael Lipsky says that –

policies are ultimately made by the street-level bureaucrats charged with implementing them for whom he coins the title “Street-level Bureaucrats” – Michael Lipsky<sup>45</sup>

### *Best Practices in Action*

#### *Grosse Pointe Public Schools*

##### *One district makes the HE rating a title to be earned*

The Grosse Pointe School district has established a unique approach to managing “Highly Effective” ratings, in which teachers apply for and earn the rating only through district-developed completion of personal growth and professional practice regimen similar to National Board certification. This defined process addresses the expectations for achieving the HE rating. It allows teachers to choose this pathway, with the principal’s agreement, at the beginning of a year if they want to go significantly above and beyond typical standards in their educator role. The Grosse Pointe schools use a merit pay system whereby all teachers rated effective or highly effective receive the same merit pay.

**Contact Person: Jon Dean, Deputy Superintendent**

*“Our findings reveal that the percentage of teachers rated as Unsatisfactory has not changed in the majority of states that have adopted new teacher evaluation systems. At the same time, we find considerable variation across states in the percentage of teachers rated in the category just below Proficient as well as those above. One primary hypothesis for these*

<sup>45</sup> M. Lipsky 2010, Street-level bureaucracy: Dilemmas of the individual in public service. 30<sup>th</sup> Anniversary Expanded Edition, Russel Sage Foundation

September 30, 2017

*findings emerges from Lipsky's work that "street-level" public-sector employees "cannot do the job according to ideal conceptions of the practice because of limitations of the work structure" <sup>46</sup>*

What have these street-level policy implementers told us about the results that we see? One concern that can be raised again is the need for an agreed upon standard of best instructional practices that is robust enough to reflect the myriad individual instructional needs and teaching conditions experienced in schools.

***"There is professional consensus that the number of teachers whose practice is below standard is very small, probably no more than 6 percent of the total, according to the Measures of Effective Teaching study and others. ...Given this landscape, it makes sense to design personnel policies for the vast majority of teachers who are not in need of remediation. And, given the complexity of teaching, a reasonable policy would be one that aims to strengthen these educators practice. Personnel policies for the teachers not practicing below standard – approximately 94 percent of them – would have, at their core, a focus on professional development, replacing the emphasis on rating with one on learning."***

Charlotte Danielson; *Charlotte Danielson on Rethinking Teacher Evaluation*, Education Week, Commentary, April 18, 2016

## Time

Over and over again educators spoke of the impact of time on the evaluation process. They spoke of the time it took to evaluate every teacher in their school, the time it took to provide meaningful feedback, the time it took to receive feedback -- preferably feedback in days not weeks. One option that is available but may not to be well known or understood is that teachers judged to be effective or highly effective during their two most recent evaluations have no minimum number of observations required<sup>47</sup>. This feature may result in a reduced number of evaluations to be conducted annually by the administrator. Arguably the option creates an incentive to rate educators effective and highly effective to reduce the pressure of large numbers of evaluations to conduct within a school year.

## Trust

Trust is an issue that occurs at several levels. Through focus groups and interviews educators raised the issue of trust as they describe barriers to effective evaluation – Trust between evaluators and those being evaluated; trust regarding how the evaluation would be used; trust regarding whether they would be fairly judged; trust in the system itself regarding whether it could be rigged by some to distort student gains. There are questions of trust regarding the intent of the evaluation process / system as designed and its results regarding whether effective teachers are accurately represented in the system with its current and planned weighting of student growth indicators. Another result of the

high stakes nature of the education evaluation system is the fear that was expressed that there would

<sup>46</sup> (2010, p. xvii). In *Revisiting the Widget Effect*; Matthew Kraft and Allison Gilmour, 2017, p 4 Quoting M. Lipsky

<sup>47</sup> Michigan Educator Evaluation Frequently Asked Questions, op cit

September 30, 2017

be consequences if respondents or their schools were found to be out of compliance with evaluation requirements.

*“Labels get in the way of trust.”* – focus group participant

Trust as an issue occurred in the conduct of the survey as some receiving the survey questioned how results would be used, by whom, and whether there could be negative consequences for responses from the state, district levels, or from individual administrators.

### Quality of Feedback

Focus group participants told us that much of the feedback that teachers receive and the observations that evaluators make is about classroom management. One can argue that without adequate classroom management instructional practices are doomed for failure. However, others may say that without appropriate instructional practices classroom management is undermined.

Both survey respondents and focus group participants describe concern for the skill of evaluators. The evaluation models ask for observation of best practices like student centered instruction and project based learning and promotion of higher order thinking skills. However, without deep level knowledge about how these practices are exhibited by teachers and students the evaluator may not be able to adequately recognize the skills and practices when seen. Further beliefs about the fit of instructional practices for different groups of students and classroom and school situations may serve to dilute implementation of practices.

### Links to professional development and professional growth

The literature of best practices, MCEE principles, and white papers emphasize the importance of connecting evaluation, feedback and professional development (professional growth). Purposeful, data driven and individualized systems are needed to guide educators to professional growth and improvement of teaching and learning.

#### *Looking at the wrong knowledge*

##### ❖ *The most powerful teacher knowledge is not explicit:*

- *That’s why telling teachers what to do doesn’t work.*
- *What we know is more than we can say.*
- *And that is why most professional development has been relatively ineffective.*

##### ❖ *Improving practice involves changing habits, not adding knowledge:*

- *That’s why it’s hard:*
  - *And the hardest bit is not getting new ideas into people’s heads.*
  - *It’s getting the old ones out.*
- *That’s why it takes time.*

##### ❖ *But it doesn’t happen naturally:*

- *If it did, the most experienced teachers would be the productive, and that’s not true (Hanushek & Rivkin, 2006).*

From: Dylan Wiliam, Institute of Education, London UK, 2010

September 30, 2017

In a Policy Brief by the National Center on Education and the Economy high performing schools from across the globe were reviewed. The policy brief concludes with 3 Lessons Learned.<sup>48</sup>

- 1) “Teacher appraisal is designed to foster professional learning and growth.” ... “the purpose of the teacher appraisal system is not [emphasis added] primarily to reward high performers and identify and eventually get rid of low performers. Rather it is to create goals for learning and improvement.” (p. 10)
- 2) Teacher appraisal is linked to professional standards of practice
- 3) Teacher appraisal can be tiered to career opportunities and compensation

At this time the link between educator evaluation and individualized and group professional development is not well demonstrated. What may be missing is the adoption of clear and widely held standards for teaching, deep understanding regarding the link between student outcome measures and teacher professional development and coherent funding that allows well-resourced individualized training experiences and coaching.

In its policy brief about how high performing systems around the world address teacher appraisal, the National Center on Education and the Economy describe the appraisal system as implemented in Finland<sup>49</sup>.

*“Teachers begin to reflect on their work and receive feedback on it during initial teacher education. In their preparation programs, teachers are introduced to a cycle of planning, action, and reflection, and are expected to engage in similar kinds of research and inquiry throughout their careers. This process underscores the notion that learning in practice does not happen on its own without opportunities for teachers to analyze their experiences, relate experiences to research, and engage in metacognitive reflection. In some ways, it models what the entire system is intended to undergo: a process of continual reflection, evaluation, and problem solving, at the level of the classroom, school, municipality, and nation.”*

The model draws on research on metacognition and systemic design and implementation of school improvement.

### **Using what we know about how adults learn, change and improve their practice**

During focus groups and reflected in white papers the question was asked whether the evaluation process reflects what is known as best practice for teaching and learning. That is, whether the standards of practice reflected in the educator evaluation process reflect standards of appraisal and feedback that a teacher would be expected to use for students; no matter whether the student is in the K-12 setting or an adult learner in the professional workforce. Where may the system be off track? Whether

---

<sup>48</sup> Policy Brief by the National Center on Education and the Economy (NCEE), “Empowered Educators: How high-performing systems shape teaching quality around the world – teacher appraisal and feedback”, 2016

<sup>49</sup> Ibid. p 2



September 30, 2017

considering a K-12 student or their teacher, learning and change take trust, purposeful professional development and coaching, monitoring, feedback, and practice.

*“If we look at teacher learning and growth in the same framework we look at student learning and growth we would be in a better place...” – Focus group participant*

*“A supportive, inquiring atmosphere allows teacher to do more than analyze, but also to learn from their own practice.” – from Ann Blais white paper*

### ***Best Practices in Action***

#### ***Harper Woods Public Schools – Harper Woods Middle School***

**Extensive training of observers, including calibration components, is featured in this district’s efforts. The district engages teaching staff in training with an emphasis on collaborative learning. Teachers have ready access to data and information regarding student growth and academic outcomes. Schools look at data to develop action steps and teachers develop action steps centered on NWEA reports.**

**Contact Person: Dave Rabbideau,  
Assistant Superintendent; Heath  
Filber, principal**

## **Summary – What We have Learned**

*“Overall the evaluation system from the state has made improvements to teaching and learning with the rubrics of what good teaching should look like, and sound like, and feel like. But the punitive nature of student growth and other parts have set back the growth of teachers – where would we be if not for this setback?” – Focus group participant*

### **What is working**

An important charge to this evaluation initiative is to identify what should change or improve to support the educator evaluation system. But before launching into change one must be clear about what currently works and why. The MCEE report and pilots have given direction to the educator evaluation system that is present today. When we look back at the MCEE principles and recommendations we see evidence of their impact on the system. We learned through focus group participants, survey respondents, practitioner white papers and interviews that educators are astutely aware of the requirements to evaluate, and are making efforts to meet requirements within the constraints of their work demands. MDE provides a number of support resources through its website, e-newsletters,



September 30, 2017

webinar, surveys, and meetings with school leaders throughout the state, along with other strategies to both share information and to listen.

### **Educator Evaluation – Impact on the System of Education**

Educator evaluation is best understood when contextualized within systems of school level, district, region, state and national stakeholders that define clientele, taxpayers, policymakers, administrators, students and parents. Districts are often simultaneously addressing student achievement, parent outreach, competition for students, fiscal distress and budget challenges, a winnowing workforce and high turnover in students and staff. In the environment of K-12 education, therefore, the evaluation system must be situated within the real world facing schools.

This review has assembled information to address the three questions posed by the Michigan Department of Education in its request for proposal regarding 1) best practices, 2) barriers, and 3) VAM and alternate assessments. We find in our data strong overlap in the three questions and their impact on the implementation of the educator evaluation models on the system of education in schools and districts. In this section we explore the overlap in data for each of the three review questions to tell the story that lies within the data assembled for this review.

Several major points surface in this review of the educator evaluation system as designed and implemented in Michigan.

- 1) Although at different stages in implementation, overall educators in Michigan are in early stages of implementation of the evaluation system as described in PA 173. As with any systemic reform effort educators are adapting the models and practices to fit their local circumstances. Much of the attention at this point seems to be compliance focused.
- 2) There is tension between the dual roles of the educator evaluation system to serve the high-stakes human resource function of rating, screening and documentation of workforce actions and the role of the evaluation system to provide feedback for professional growth and improvement.
- 3) Overall educators report little useful feedback to improve their professional practice and support their professional growth. They describe time and lack of skills in evaluation among barriers to implementation. These, too, may be barriers to providing useful feedback – creating a negative feedback loop.
- 4) Similarly, educators cite loose connections between evaluation models used as a reflection of professional standards and expectations, student learning outcomes, and information to increase student learning and educator’s professional growth. This disconnect creates a cycle that drags down the utility of the educator evaluation and can result in compliance rather than substantive implementation –continuing the cycle.
- 5) Teachers report not being consistently engaged in goal setting with their evaluator or colleagues, and not consistently adopting improvement strategies recommended by their evaluator – factors that may be related.

September 30, 2017

6) There are technical questions surrounding the efficacy of the system to serve human resource functions, most notably the psychometric and technical issues related to the actual student growth measures and the fidelity with which the evaluation system is implemented. As noted earlier, to the extent that the educator evaluation system is used for high-stakes decision making the process would benefit from modeling to determine the reliability and validity of the models as implemented in Michigan schools.

8) Educators seem to be struggling to balance issues of inter-rater reliability and fidelity of implementation with practical issues of limited time and training.

9) There was little to no evidence of a link between the educator evaluation results and district, building or individual professional development plans or school improvement plans. If the educator evaluation system holds any value whatsoever this is a resource squandered.

### Additional questions – what haven't we learned

No review of this type can credibly cover all aspects of a system in a limited time period. Indeed, educator evaluation lies in a full system of changing demands, resources, policies and expectations. In addition to its findings this review has raised questions that warrant further exploration.

**Merit pay:** Throughout the literature recommendations regarding the appropriateness of merit pay and other pay systems linked with educator appraisal are mixed. The MCEE report specifically recommended against a merit pay system related to the educator evaluation system.<sup>50</sup> During focus groups we heard some mention of merit pay systems operating in Michigan districts. Our search for best practices reveals a merit pay structure implemented in the Lenawee Intermediate school system for all bargaining units. There is much more that can be learned from merit pay linked to educator evaluation systems as they operate in Michigan and nationwide.

### Best Practices in Action

#### Lenawee ISD

**Lenawee ISD has adopted a shared responsibility model that features goal-based instruction, ongoing feedback, and common focus. Salary increases for all bargaining units, not just teachers and administrators, are linked to student achievement outcomes. The model gives all employees a stake in student growth, and increases commitment and collaboration among all staff.**

Contact Person: Dan Garno, Executive  
Director of Staff Resources

**Quality of evaluation feedback:** Our survey shows that teachers in particular and to some extent administrators question the impact that the evaluation system has on their practice. A deeper look at

<sup>50</sup> MCEE report op cit

September 30, 2017

the actual feedback that educators receive is warranted and may provide information about the types and quality of feedback given.

**Link between educator evaluation findings and professional development:**

Perhaps most important to the utility of the educator evaluation model is its connection to inform and drive the professional growth of educators. Our survey shows weak links between professional development and evaluation feedback. It is not clear how or to what extent evaluation impacts individual professional growth plans, or building and district professional development plans.

We see some evidence of innovative districts that have begun to develop ties between individual professional growth plans for their entire workforce, but for many schools and districts the ties are not yet evident. Further review of the system of professional development and educator growth from pre-service through veteran educator is needed.

September 30, 2017

## Recommendations and next steps

*“Every system is perfectly designed to get the results it gets.” W. Edwards Deming*

We noted earlier in this report the intent to provide information useful to MDE, policy makers, K-12 educators and their supporters to improve educator evaluation in Michigan. With that purpose in mind we offer the recommendations below. We have grouped recommendations by the entity with primary responsibility for implementation.

### Supportive accountability

- ❖ **What is needed from teachers:**
  - **A commitment to:**
    - **The continual improvement of practice**
    - **Focus on those things that make a difference to students**
- ❖ **What is needed from leaders:**
  - **A commitment to engineer effective learning environments for teachers by:**
    - **Creating expectations for continually improving practice**
    - **Keeping the focus on the things that make a difference to students**
    - **Providing the time, space, dispensation, and support for innovation**
    - **Supporting risk-taking**

**From Dylan Wiliam, Institute of Education, London UK, 2010**

### Recommendations for MDE

We begin by making recommendations directed toward MDE officials.

I. **Stabilize M-Step, gather multiple year data with one version of the test** – throughout our review we heard concern raised regarding stability of M-Step and its resultant appropriateness for high stakes decisions. Two factors are pertinent. First, reliable decision making takes multiple years of trend data. Second, school factors are adapted to revisions of the state assessment and the results produced. However, it takes time and resources to adopt revisions in local curriculum, instructional practices, instructional materials,

teacher training and formative assessments to serve changes in the statewide test. For schools, teachers and students to experience the full value of the educator evaluation system the measures of student growth and the state assessment specifically must be stable for multiple years.

- II. **Communicate standards for instruction** – Among best practices is standards that are well known and supported. One type of standards that impacts teaching and learning is the standards for instruction. These standards are communicated through the evaluation model selected by the local district and appear to have wide variation in rigor and fidelity in implementation across the state and within districts. Among best practices in Michigan we see evidence of school districts and ISD/RESA adopting professional learning communities to build

September 30, 2017

- common understanding of standards of instruction. MDE can advocate for these communities and support their work through professional development support.
- III. **Add to /continue to provide face-to-face opportunities for educators to give input to MDE policy and practice, especially teachers and principals in partnership** - participants in the focus groups thanked MDE for valuing their input and for staging the conversations among teachers and administrators that the focus groups provided. They indicated willingness to participate in future focus groups and asked that MDE continue this strategy for learning from those directly involved on the frontline.
  - IV. **Enhance the mindset and focus of schools toward improvement of practice over ranking** – the work of instructional improvement strengthened through collaboration is being undermined by the competitive culture of ranking educators. Research has shown that the vast majority of educators are capable of improvements in practice when focused feedback and support are given. Instructional improvement is a learning and organizational issue not a matter of sanctions. MDE can be a catalyst to a mindset that supports improvement of professional practice throughout the professional life of educators.
  - V. **With ISD and RESA convene and support statewide professional learning communities (communities of practice) for those leading improvement of instruction and professional development and use of SLOs** (see Dearborn model) – Educators from across the state could benefit from a convening of a statewide professional learning community to explore strategies and research for implementing educator evaluation within a unified system of improvement of teaching and learning.
  - VI. **Assist districts and ISDs in creating access to training video of best instructional practices “What effective instruction looks like”** – MDE can work with local educators to develop and distribute video and online depictions of what effective instruction and rigorous standards look like. MDE can help assuage the notion *that students like the ones in my classroom cannot perform in those ways*
  - VII. **Develop a calendar of required reports and document preparation for building principals at the end of the school year and work with district leaders to consolidate reporting and document preparation.** The end of a school year is an intense time that is flooded with end-of-year reports, next year grant applications, transparency reporting, testing, annual plan development, staffing, budget planning, and year end activities, both formal and informal. Each activity and report requires principal and teacher involvement. MDE and local district leaders can look for ways to consolidate, reschedule, or omit these requirements so educators can focus on teaching and learning.
  - VIII. **Encourage and support learning strategies and resource sharing with other states and nations successfully tackling educator evaluation.** Literature shows evidence of innovative strategies across the nation. Michigan may be able to learn from what works and what has not in other states and nations.
  - IX. **Identify factors that have resulted in state and national teacher shortages and identify means to improve the numbers and skills in the workforce** – with economists, researchers, university education schools, and local districts identify the factors that lead to the decline in workforce in

September 30, 2017

K-12 education in Michigan and nationwide, and identify strategies for workforce development. Responsibility for workforce development is shared with local educators and MDE

- X. **Consider leveraging Michigan’s voice to advocate for improvements in the state-approved educator evaluation models.** Establish an educator taskforce made up of teachers, principals / assistant principals and human resource professionals to develop feedback and recommendations for the developers of the approved evaluation models. Invite the models developers to Michigan to receive the recommendations and share their response.

*Recommendations for district implementation (school and district levels or ISD /RESA)*

- I. **Train evaluators and those evaluated** – to implement evaluation models, observe, record and provide effective feedback, understand the meaning and use of student growth measures to improve teaching and learning, and to schedule and manage time to conduct all required observations
- II. **Convene and facilitate professional learning communities of member districts** – for an example see Newago ISD, Dearborn and Wayne RESA models
- III. **Evaluate the quality of feedback** – emphasis on improvement / professional growth
- IV. **Strengthen the ties of evaluation to PD** – comprehensive professional development plan / model that is tied to individual needs and professional growth model (full system)
- V. **Assist local districts by convening work sessions to develop SLO-** allowing districts to share the burden and cost of the work and technical support
- VI. **Provide access to high quality national and state trainers to enhance training available at local levels** – smaller and resource strapped districts often do not have access to high quality trainers, by providing access to high quality trainers on a regional basis MDE and ISDs can enhance the quality of training provided to educators
- VII. **Enhance the quality of calibration /inter-rater reliability** – provide training to improve inter-rater reliability and calibration technical support
- VIII. **Enhance training regarding understanding student outcomes measures and results** – educators need deeper understanding of student outcome measures and results to make more informed decisions about their links to instructional planning, school improvement, and educator training and appraisal.
- IX. **Continue to emphasize and build educator awareness /knowledge of what quality teaching and learning look like** – without boundaries of class, income, culture, race, ethnicity, not reserved for the few
- X. **Address matters of cultural relevance and competency regarding curriculum, instructional methods and observations** – provide leadership with educators throughout the state to develop understanding of culturally responsive teaching.

*Recommendations regarding policy and governance issues (state and national)*

- I. **Get the balance between improvement and sanctions right** – untangle the two functions educators are being asked to serve

September 30, 2017

- II. **Identify factors that have resulted in national teacher shortages and identify means to improve the numbers and skills in the workforce** – responsibility for workforce development is shared with local educators and MDE but is also impacted by state and national policy and funding
- III. **Identify and address underfunding and resource disparities among schools and districts**
- IV. **Support MDE efforts to stabilize statewide assessment** – educators need a well-known and understood assessment system that remains in place long enough to provide multiple year trend data allowing educators to adapt and improve their practice based on results. The assessment system must meet psychometric standards sufficient for the high-stakes decisions it impacts.

### Next steps

The fact that MDE has engaged a process to review implementation of a major policy initiative that affects all K-12 educators in the state is commendable. Even more so is the openness to continually learn and improve upon its work – the very strategies asked of schools and districts as they work to improve teaching and learning, and to support educators’ professional growth. We strongly encourage MDE to continue its journey of monitoring, listening, reflection and learning to improve professional practice and learner outcomes.

This report contains numerous tables and charts representing countless hours of analysis by the review team. Nevertheless, there is a wealth of information that remains to be found in additional analysis of the data assembled for this review. We encourage further analysis of data augmented by additional practitioner focus groups, problem solving and planning sessions.

We have included in this review specific steps that MDE, districts, ISD/RESA and policy makers can take immediately. We have channeled the voice of practitioners eager to share their thoughts, experiences, insights and recommendations. We have chronicled the most recent scholarly literature and commissioned papers from two esteemed academic voices. All of this gives fuel to learning, but also builds a map for continuous improvement.

- Broadly share the findings of this review
- Conduct focus groups with educators around topics related to the system of school improvement
- Continue to seek out and share best practices
- Help districts leverage resources to bring national trainers that work with multiple districts
- Report progress on recommendations from this and other evaluation and MCEE reports

This report does not contain all of the answers. Review and discussion by MDE officials, and educators throughout the state can turn the information and data found in this report into insight, strategies, action plans, and results.