This appendix contains documents that support 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. The full evaluation report reviews best practices, barriers to implementation and use of VAM / alternative assessments in educator evaluation.

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

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

Contents

Appendix Introduction **5**

The Ray.Taylor and Associates Project Team **6**

Literature Review **10**

A Summary of Documents Reviewed **66**

Researcher White Papers **71**

Voices from the Field: Practitioner White Papers **91**

MDE Evaluation Focus Group **115**

Statewide Survey Documents **128**

Best Practice Documents **133**

September 30, 2017

Appendix Introduction

This appendix includes several original documents developed for the evaluation report. Excerpts from these documents appear in the full evaluation report.

The Ray.Taylor and Associates Project Team

The brief biographies that follow describe the Ray.Taylor and Associates team for this project.

Rossi Ray-Taylor, Ph.D. has spent the past decade providing program evaluation services, planning, school transformation management, professional development, administrative and management counsel to a national client base. The former superintendent of Ann Arbor Schools and Deputy Superintendent for Instruction for the Lansing Public Schools Dr. Ray-Taylor has built a career using data informed and systems strategies to improve schools. Dr. Ray-Taylor's work specializes in program evaluation, planning, accountability, systems design and monitoring, and executive coaching. She has recently designed the curriculum for and co-facilitated the Fiscal Leadership Collaborative for financially distressed school districts.

Rossi Ray-Taylor designed and conducted an external review of a university based charter school authorizer. The project included extensive document review, interviews and surveys of the authorization agency team, charter school leaders and stakeholders. Ray.Taylor and Associates has also designed and conducted equity audits, and Title I and Title IIa program audits. She has written countless program proposals for state and federally funded grants and foundation support for schools and associations resulting in millions of dollars in funds. She has conducted school transformation management for the School Improvement Grant for a mid-Michigan schools district and designed and conducted focus group feedback review of the Statewide System of Support for MDE.

Dr. Ray-Taylor holds Bachelor's and Master's Degrees from Michigan State University. Her Doctorate of Philosophy in Educational Psychology with a specialty in learning and cognition and in measurement and evaluation is also from Michigan State.

John E. Norwood, Ph.D. is a career educator with over forty years of experience in the field. He is a former Urban Teacher Corps Intern with the University Of Toledo and the Toledo Public Schools. He has been an elementary education teacher, an alternative education administrator, a middle school principal, a program director, and central office administrator with the Toledo Public Schools, and an adult education supervisor and director, middle school principal, an assistant superintendent, and a program director with the Saginaw Public Schools. Dr. Norwood was also a lecturer in the College Of Education, Education Leadership Program with the University of Phoenix, Southfield Michigan. Dr. Norwood has been a university administrator as the Director of the Union Graduate School / Cincinnati Public Schools Teacher Corps Project, Cincinnati, Ohio, formerly the graduate branch of the Union for Experimenting Colleges and Universities, known today as the Union Institute. Additionally, Dr.

Norwood was the Director of the Northeastern University / Boston Public Schools Teacher Corps Project in Boston, Massachusetts. Dr. Norwood is president of John Norwood Associates, and a senior associate with Ray.Taylor and Associates. John has a Bachelor of Education degree from the University Of Toledo, a Master of Education from Bowling Green State University, and a Ph.D. from the University of Michigan.

Patrick (Rick) O'Neill, M.A. has expertise in the areas of leadership coaching and training, strategic planning, group facilitation, editing, school improvement planning, teacher evaluation and coaching, teambuilding, school building staffing, and school building level administration for elementary and middle schools. Mr. O'Neill has central office and building administrative experience and skill in school redesign and start up.

Mr. O'Neill's administrative career includes experience as an elementary and middle school principal and as president of the Ann Arbor Administrators' Association. He served as Director of Capital Planning for the district, chaired the district selection committee for planning services consultants, and supervised several administrative departments. He coordinated demographics activities for the Ann Arbor Public Schools and chaired a broad-based committee for the district, resulting in a comprehensive report to the Board of Education on demographics and facilities. During his teaching career, Mr. O'Neill served as the vice-president of the Ann Arbor Education Association.

A graduate of University of Michigan, Mr. O'Neill has earned a Bachelor of Arts degree in Psychology and a Master of Arts degree in Elementary Education.

Paula Allen-Meares, Ph.D. served from 2009-2015 as Chancellor of the University of Illinois at Chicago (UIC), and Vice President of the University of Illinois. She is currently the John Corbally Presidential Professor and a Professor of Medicine at UIC. She holds certificates in management from the University of Michigan and Harvard University, an Executive Education certificate from the Women's Director of Development Program at Kellogg School of Management, and the Creating and Leading a Culture of Innovation Certificate from Northwestern University. She also holds other faculty appointments on both the Chicago and Urbana-Champaign campuses, and is Dean and Professor Emerita and the Norma Radin Collegiate Professor at the University of Michigan.

While serving as Chancellor at UIC, she oversaw a budget of $2 billion, hosting 28,000 students and over 11,515 faculty and staff. At both the University of Michigan and UIC, she created interdisciplinary research institutes and programs of local, national and global significance. The creation of intentional partnerships characterized her tenure. Dr. Allen-Meares has written over 150 articles/books/commentaries. She is an elected member of the National Academies of Science, a Trustee and Fellow of

the New York Academy of Medicine, and for almost a decade served as a Trustee of the WT Grant Foundation.

Dr. Paula Allen-Meares is the President and CEO of Paula Allen-Meares & Associates, LLC. She and her associates embrace the growing diversity of our society, and has received competitive external funding from foundations to promote better educational and health outcomes for youth, to address issues of poverty and mental illness, to advance health literacy as an intervention to achieve better outcomes for at risk inner city diverse populations. Dr. Allen-Meares has been invited to Canada, England, Korea, Germany, and Africa, etc. to deepen understanding about the intersectionality of education and human services.

Jessica Thomas is a strategic turnaround advisor who serves distressed governmental entities to improve their financial health, thereby improving the lives of the citizens who work and reside in these communities. Jessica's skill-set and background includes: municipal finance, operational development, governmental affairs, budget deviation analysis, negotiations, governmental audit, procedure development, and GASB training and presentations. She is a trained professional with over 15 years of experience in governmental audit, accounting, and corporate finance, and strategic development with a solid understanding of financial reporting, analysis, management, quality assurance, internal controls, and team building in a corporate setting. Ms. Thomas is a proficient project leader who aligns financial initiatives to achieve strategic objectives/goals and is results driven, decisive collaborator with proven success in building strong lasting relationships with managers, associates, consultants, and clients. Jessica Thomas is a skilled auditor, experienced in analyzing processes and recommending improvements and an effective negotiator who seeks positive resolution to problems through active listening and commitment to integrity. Ms. Thomas has been recognized as one of Oakland County's Elite 40 Under 40, is an elected official for Birmingham Public Schools, and has served as a keynote speaker, panelist, and panel moderator at industry events. She also volunteers her time as a presenter for the Fiscal Leadership Collaborative, a yearlong pilot program designed to help distressed districts address both the causes and the cures of fiscal instability. Using her talents as an analyst, facilitator, and communicator, Jessica assists municipalities and school districts in re-defining business objectives, identifying steps to avoid further decline and correcting deficiencies, analyzing essential services, identifying business risks, determining the financial/fiscal condition of jurisdictions, and reviewing internal controls and decision-making processes to improve efficiencies. Jessica Thomas holds a B.B.A. in Economics and Management, from Albion College, Albion, MI with a concentration in accounting and religion.

Ira J. Washington, PhD, received his Ph. D. from Michigan State University in 1992

in Educational Psychology with a specialization in measurement evaluation and research design. His cognate areas were statistics and learning and cognition. He has expertise in the areas of test construction, counseling, inferential and nonparametric statistics, survey design and statistical analysis. Dr. Washington has long term experience with: 1) the Statistical Package for the Social Sciences, 2) Microsoft Access (QBO), 3) Statistical Analysis System (mainframe, personal computer), and 4) knowledge of Apple and MS Windows based hardware and software. Dr. Washington also holds a Bachelor of Arts in psychology with teacher certification in secondary social studies as well as a Masters of Education degree with certification in secondary guidance and counseling from the University of Louisville.

Mathis Taylor is a Research and Design Associate, and Technology Manager at Ray.Taylor and Associates. Mr. Taylor's experience includes survey data compilation, analysis and summary, online research, graphic design and illustration. Mr. Taylor holds a Bachelor of Arts degree from Saginaw Valley State University, with a major in art and a minor in sociology.

Amber Tselios serves as a content writer, editor, and marketing consultant for Ray.Taylor and Associates. Her experience includes, but is not limited to, graphic design, creative and technical writing, and editing in MLA and APA formats. She holds a Bachelor of Arts degree in English from Western Michigan University, numerous Inbound Marketing certificates from Hubspot, and a certificate in Marketing from Cornell University. .

Marshall Taylor acts as media coordinator, content writer, and editor at Ray.Taylor and Associates. These duties include the creation and management of all video content produced by and for Ray.Taylor and Associates, and generating and editing text content. Marshall Taylor holds a Bachelor of Arts degree from Western Michigan University, where he dual-majored in English and Communication, as well as extensive training and certificates in Inbound Marketing from Hubspot.

Literature Review

This literature review was conducted to understand what are the best practices for the evaluation of teachers and the barriers to the implementation of those practices, as well as alternative methods/tools used to measure student learning outcomes. As stated previously, in accordance with Public Act 173, the MDE approved several approaches to teacher evaluations: 1) the Danielson Framework, 2) the Marzano Teacher Evaluation Model, 3) The Thoughtful Classroom, and 4) the 5 Dimensions of Teaching and Learning.

Executive Summary

This review draws on an in-depth review of literatures on the topic to summarize best practices in teacher evaluation. It is organized under three main questions:

1. What are the best practices in education evaluation?
2. What are the barriers to implementation?
3. 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 Teaching, the Robert Marzano Approach, the Thoughtful Classroom method, and the 5 Dimensions of Teaching and Learning method. It 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. The section of this review about 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. Lastly, 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?

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, collective bargaining, 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- and 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.

Introduction

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, Ellett, & 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 (Ellett & Teddlie, 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.

The interest in teacher performance evaluation is not unique to the USA. The Organisation for Economic Co-operation and Development (OECD) takes on a global perspective in a 2013 report which provides an analysis of how other countries evaluate their teachers and the elements of that process. In that report, as in many reports published in the USA, the rationale for evaluation are to provide accountability and to improve teaching and learning, (OECD, 2013; Akiba, 2017; Ingvarson,& Rowley, 2017).

The School as a Social Organization

In addition to its function as a bureaucracy, the school is also a social organization. It has definable roles and responsibilities for its employees, and it labors under a number of societal expectations. For example, schools are expected to perform important social functions such as transmitting knowledge, instilling an appreciation for learning, and preparing a 21st-century workforce.[1]

As a goal-oriented social system, the school is a special type of learning organization. Bowen, Rose, and Ware (2006) associate learning organizations with "a core set of conditions and processes that support the ability of an organization to value, acquire, and use information and tacit knowledge acquired from employees and stakeholders to successfully plan, implement and evaluate strategies to achieve performance goals" (pp.98-99).

Question 1: What are Best Practices in Education Evaluation?

Introduction

In this section, we review established best practices in several areas of educator evaluation. Areas reviewed include: developing teacher evaluation systems; selecting tools for conducting a teacher evaluation; implementing teacher evaluation systems; providing feedback on teacher evaluations and the use of collected data; and principal leadership style and student learning.

In teacher evaluation systems we address components of well-structured local evaluation systems. We then review several tools currently in use for conducting teacher evaluations, analyzing their

[1] This list of responsibilities in not exhaustive but is intended to capture the essence of school roles in American society.

respective methods and highlighting strengths and weaknesses. In implementing teacher evaluations, we focus on components and structures of effective evaluation measures, as well as the role the evaluator plays in obtaining an accurate evaluation. Best practices in teacher evaluation methods survey a variety of inspection and demonstration approaches, analyzing the strengths and weaknesses of each. In the use of data section, we highlight the importance of translation to usable feedback, and the need for ongoing evaluation. In exploring principal leadership style, we focus primarily on how teachers perceive principal leadership, and those principal behaviors that teachers perceive as influencing student achievement.

Best Practices: Development of Teacher Evaluations

Although recommendations for teacher evaluation vary depending on theoretical perspective, the following represent best practices and essential components, found in through a review of the literature on the evaluation of teacher performance, for developing local teacher evaluation systems. Evaluation practices focus both on district- and school-level needs, as well as on teacher- and classroom-level interests.

Regarding district- and school-level needs, a number of points are salient. At a foundational level, evaluation systems should link directly to the mission of the school district and systems must integrate evaluation procedures with curricular standards, professional development activities, targeted support, and human capital decisions (e.g., recruitment, selection/placement, induction, professional development, performance management and evaluation, and career progression) (Goe & Croft, 2009; National Education Association, 2011). To ensure the success of such a system, the district and/or school must commit to allocating adequate resources to the initiative. Further, schools and districts must develop evaluative criteria grounded in recent teaching and learning scholarship, ensuring the

validity of the criteria. Finally, when identifying criteria that contribute to good practice, they should ensure that teachers will be able to demonstrate the criteria.

Rather than being developed solely at the institutional or administrative levels, evaluation systems should involve all important stakeholders, including teachers, in selecting evaluation criteria and developing the assessment system (Goe & Croft, 2009). Key decisions about assessment and evaluation systems need to be made as close to the local level as possible and in partnership with teachers and their teacher evaluation representatives, and obtaining teacher input in determining performance and student-focused learning outcomes should be part of the evaluation process. (Darling-Hammond, 2012; National Educational Association, 2011) Moreover, evaluation developers should consider teaching contexts and how the evaluation system will accommodate them. For example, early elementary teachers cannot be evaluated with value-added models (Goe & Croft, 2009).

A number of scholars recommend including multiple evaluative pathways and avoiding reduction of teacher effectiveness to a single score obtained via an observation instrument or value-added model. Teachers' contributions to the work of their colleagues and the school as a whole should be considered. Since student learning gains are a function of teachers' collective efforts, these valuable contributions should be part of the evaluation process (Mulford, 2003). Darling-Hammond (2012) has also asserted that evidence of teachers' contributions to student learning based on classroom work and other assessments that are appropriate and valid for the curriculum and for the students being taught should be included. Davis, Ellet and Annunziata (2002) advocate for including teacher self-assessment and assessment of written planning documents, as well as a focus on learner engagement in the teaching and learning process.

Best Practices: Choosing a measurement tool for teacher evaluation

The following findings from the literature explore best practices for selecting tools to use for conducting teacher evaluations. This section also describes the methods, advantages, and disadvantages of several teaching evaluation frameworks.

According to Goe, Bell, and Little (2018), prior to selecting a method or instrument of evaluation, the purpose for teacher effectiveness should be considered; evaluations should always measure what is most important to administrators, teachers, and other education stakeholders. When evaluating, it is a best practice to use an instrument that is already valid and reliable, and adapt if necessary. However, it is important to note that validity lies not solely with the quality of the instrument or model, but also with how well the instrument measures the construct and how the instrument is used in practice (Sechrest, 2005)

When selecting evaluation instruments, developers should seek or create additional measures that capture important information about teachers' contributions beyond student achievement score gains. Education analysts now commonly accept that multiple measures of student learning and teacher effectiveness are necessary components of teacher assessment and evaluation. To be effective, a teacher assessment and evaluation system must employ a carefully constructed set of multiple measures. As measurement tools, teacher evaluations—both value-added models and standards-based observations—should be judged according to three criteria: Are they unbiased? Are they reliable? Are they valid?

In its 2013 report, the Michigan Council for Educator Effectiveness (2013) recommends a number of specific observational tools. Discussed below are the Danielson Framework for Teaching, the

Marzano Approach, and the Thoughtful Classroom Framework.

Charlotte Danielson's Framework for Teaching. Danielson's Framework for Teaching is a research-based set of components of instruction, aligned with the Interstate Teacher Assessment and Support Consortium (InTASC) standards, and grounded in a constructivist view of learning and teaching. Launched in 1996, the most recent edition was published in 2013. Teaching is divided into 22 components and 76 smaller elements that are clustered into four primary domains of teaching responsibility:

Domain 1: Planning and Preparation

Domain 2: Classroom Environment

Domain 3: Instruction

Domain 4: Professional Responsibilities

Danielson also offers a mathematics-oriented version of the Framework, reflecting her recent thought on organizing it to promote the professional growth of teachers through reflection and conversation that center on instructional practices common in mathematics classrooms.

Assessing Danielson's Framework for Teaching. Mielke (2012) points out a number of advantages to Danielson's Framework, including its coherent structure and its comprehensive, research-grounded approach to "[describing] all of teaching, in all its complexity." Regarding teachers' engagement with the Framework, it is a public approach, allowing teachers to understand what observers and evaluators are looking for. Regarding applicability and generalizability, the Framework recognizes and attempts to identify the "powerful commonalities" among unique features of diverse settings and contexts.

The Robert Marzano Approach. In addition to serving as a leading educational researcher and developing a framework for teaching effectiveness, Robert Marzano has also developed the Marzano

Observational Protocol (based on the Art and Science of Teaching Framework). This evaluation tool measures the success of highly effective teachers. He identifies a number of classroom practices strategies that allow teachers to work effectively in the classroom. According to Marzano, good teachers set goals; provide feedback; help students interact with new knowledge; provide students with simulations and low-stakes competition; establish and maintain classroom rules; maintain relationships with students; and communicate high expectations. Marzano further believes that teachers need to choose areas of improvement throughout the year and that administrators should be responsible for providing them with opportunities to observe effective strategies of other teachers. Administrators should then provide feedback and dialogue to teachers about how they can become more successful.

Marzano has also developed a framework for evaluating levels of school effectiveness, identifying five distinct levels:

Level 1: Safe and orderly environment that supports cooperation and collaboration

Level 2: Instructional framework that develops and maintains effective instruction in every classroom

Level 3: Guaranteed and viable curriculum focused on enhancing student learning

Level 4: Standards-referenced system of reporting student progress

Level 5: Competency-based system that ensures student mastery of content

A number of general guidelines, not included in the levels referenced above, enhance use of this framework. They include:

- Schools can and should work on the leading indicators for multiple levels simultaneously, especially levels 1, 2, and 3.

- The teacher evaluation system used in a school should directly support attainment of the levels in the framework.
- The school leader evaluation system used in a school should directly support attainment of the levels in the framework.

Assessing the Marzano Approach. Marzano places an emphasis on classroom instruction and behaviors while making a distinction between new knowledge, deepening knowledge, and hypothesis generation. The greater emphasis on classroom strategies and behaviors also enables those who use the Marzano protocol to be specific in identifying strategies and practices that directly affect student learning. However, training and implementation materials may be cost prohibitive (Mielke, 2012).

The Thoughtful Classroom Teacher Effectiveness Framework. The Thoughtful Classroom Teacher Effectiveness Framework is a comprehensive system for observing, evaluating, and refining classroom practice. The Framework aims to create a common language for talking about high-quality teaching and classroom practice improvement. The Framework assesses ten dimensions of teaching, outlining a set of observable teaching indicators within each dimension and relevant student behaviors associated with effective instruction. (Silver, Perini, & Boutz, 2016; Silver Strong and Associates, 2014)

Thoughtful Classroom Methods. The first component of the Thoughtful Classroom, *Four Cornerstones of Effective Teaching,* encompasses dimensions 1-4, four foundational dimensions adapted from preeminent teacher-effectiveness frameworks (Danielson, 2007; Marzano, 2007; Marzano, Frontier, & Livingston, 2011; Saphier, Haley-Speca, & Gower, 2008). These cornerstones represent the universal elements of quality instruction, whether in a kindergarten class, AP Physics

lab, or anywhere in between. Without these four cornerstones in place, student learning will be compromised. They include:

Dimension 1: Organization, Rules, and Procedures

Dimension 2: Positive Relationships

Dimension 3: Engagement and Enjoyment

Dimension 4: A Culture of Thinking and Learning

The second component, *Five Episodes of Effective Instruction*, encompasses dimensions 5-9. Developers identified five critical episodes that increase the likelihood of deep learning. In these five episodes, listed below, teachers work towards distinct instructional purposes:

Dimension 5: Preparing Students for New Learning

Dimension 6: Presenting New Learning

Dimension 7: Deepening and Reinforcing Learning

Dimension 8: Applying Learning

Dimension 9: Reflecting on and Celebrating Learning

The developers state that understanding these five episodes— and their driving purposes—is critical for both teacher and observer. Teachers use these episodes to design high-quality lessons and units. Classroom observers use these five episodes to immediately orient themselves within the instructional sequence, ensuring that teachers and observers are on the same page.

The Framework's tenth dimension focuses on professional practice, which addresses important non-instructional responsibilities, including the teacher's commitment to ongoing learning, professionalism, and the school community.

Dimension 10: Looking Beyond the Classroom to Effective Professional Practice

Assessing the Thoughtful Classroom Teacher Effectiveness Framework. Advantages of the Thoughtful Classroom framework include the presence of rubrics for developing summative evaluations; protocols to help school leaders provide meaningful feedback to teachers; and information for conducting quality pre- and post-observation conferences. However, training and implementation materials can be cost prohibitive.

5 Dimensions of Teaching and Learning™ .

Developed at the University of Washington, the 5 Dimensions of Teaching and Learning instructional framework (5D) is a tool to help schools and districts successfully implement high-quality instructional practices (University of Washington Center for Educational Leadership, 2002). The tool can be used as the "lens" for classroom observations, as a guide for teachers, and as a reference during lesson planning and staff meetings about instructional practices. The 5D instructional framework is closely aligned to the Center for Educational Leadership's (CEL) 5D+ Teacher Evaluation Rubric.

Methods. The 5D is composed of the following five dimensions: Purpose, Student Engagement, Curriculum and Pedagogy, Assessment for Student Learning, and Classroom Culture and Environment (University of Washington Center for Educational Leadership, 2002), which are then divided into 13 sub-dimensions, as follows:

Dimension 1: Purpose

- Standards
- Learning Target and Teaching Point

Dimension 2: Student Engagement

- Intellectual Work
- Engagement Strategies

Talk

Dimension 3: Curriculum and Pedagogy

Curriculum

Teaching approached and/or strategies

Scaffolds for Learning

Dimension 4: Assessment for Student Learning

Assessment

Adjustments

Dimension 5: Classroom Environment and Culture

Use of physical environment

Classroom routines and rituals

Classroom culture

Best Practices: Implementing systems to measure teacher performance

This section will discuss best practices for implementing teacher evaluation systems, compiled from a review of the literature on measuring teacher performance (Rowan, Schilling, Spain, Bhandari, Berger & Graves, 2013; Darling-Hammond, 2012; National Educational Association, 2011; Michigan Council for Educator Effectiveness, 2013).

- Sufficient time should be allotted for development and implementation to enhance understanding of evaluation criteria, processes and targeted outcomes.
- Evaluation systems should involve safe and open collaboration. When assessment of teacher practices is transparent and openly collaborative, teachers can build professional communities and learn from one another

- Data should be collected from multiple sources (use triangulation, making sure that all education-related activities are rated by the people best qualified to rate them).
- Standards-based evaluations of practice based on observations, curriculum plans, assignments, and assessments revealing teachers' classroom practice should be included
- Evaluations must be comprehensive—based on multiple indicators to provide teachers with clear and actionable feedback to enhance their practice—and must include all three of the following components
    a. Indicators of Teacher Practice demonstrating a teacher's subject matter knowledge, skill in planning and delivering instruction that engages students, ability to address issues of equity and diversity, and ability to monitor and assess student learning and adjust instruction accordingly.
    b. Indicators of Teacher Contribution and Growth demonstrating a teacher's professional growth and contribution to a school's and/or district's success.
    c. Indicators of Contribution to Student Learning and Growth demonstrating a teacher's impact on student learning and growth.

- Evaluations must be meaningful, providing all teachers with clear and actionable feedback linked to tailored professional development
- Evaluations must be fair, conducted by highly trained and objective supervisors or other evaluators as agreed to by the local affiliate, whose work is regularly reviewed to ensure the validity and reliability of evaluation results.
- Evaluators should be allowed to make reasonable judgments regarding the quality of teaching; and schools and districts must include procedures to offer intensive assistance, if needed, to teachers who are struggling to perform adequately.

  - Those making evaluative judgments must be adequately trained so their judgments are accurate, consistent, and based on evidence.
  - The documentation of teaching is only as good as the observer. Evaluator training increases the reliability and validity of observation tools and enhances administrators' ability to deliver fair, accurate, and usable feedback to teachers.
  - Training in the use of observation tools should consist not only of the 4 days of introductory training provided by vendors at the outset of the pilot, but also additional calibration training designed to improve observation scoring and reduce rater error .

- Fidelity
  - Individuals conducting classroom observations for teacher evaluations should be instructed to use the classroom observation tools in the "manner prescribed" by tool vendors. It will be especially important for principals to score mandatory items on a protocol, for missing item data can affect observation reliability and (perhaps) validity
- Number of observations
  - Data suggest that 3- 4 observations per year should be specified as the minimum number of observations per year when a teacher is in an evaluation cycle
- Steps to Correct for Rater Error.
  - A very good way for districts to correct for rater error is to "randomly" assign individuals other than the principal to conduct at least some observations alongside the principal over the course of the school year. This practice should be encouraged by the state

Incorporating multiple observations into an evaluation helps to increase reliability, particularly if the observations are unannounced

To satisfy these requirements, evaluation systems must be adequately funded and staffed, and fully developed and validated, and all teachers must be trained on the new systems before they are used to make any high-stakes employment decisions

Best practices: Methods/Approaches to Teacher Evaluation

This section is drawn from the literature on teacher evaluation and presents an analysis of advantages and disadvantages of specific teacher evaluation models. Included are both inspection and demonstration frameworks, among them classroom observations, standards-based evaluation, principal evaluation, student feedback, teacher self-report measures, teacher portfolios, and analysis of instructional artifacts.

Roeber (2011) describes two basic frameworks for educator evaluation, the *inspection*- and the *demonstration framework*. In the former, a person or persons familiar with the work of the educator conducts the evaluation. In the latter, the educator is responsible to demonstrate her or his proficiency by collecting and organizing appropriate evidence. Each carries with it a set of benefits and detriments, discussed in turn below.

In the inspection framework, each teacher is judged in a standard manner. The formal external evaluation framework requires educators to provide evidence of their competence through the actual demonstration of their instructional prowess and/or school leadership. Moreover, supervisors can be taught to judge the competence of their subordinates. However, because it is standardized, the framework may not be equally applicable to all persons in a particular job, since individuals differ. Too, criteria for conducting the inspection may not be well documented; supervisors might

operationalize the criteria differently. Those conducting the external evaluation may lack necessary training or, in larger districts, the time to evaluate every educator annually. Further, some evaluators may not bring a neutral orientation to the evaluation process.

The demonstration framework, in contrast, works equally well for teachers and school leaders, and motivates educators to demonstrate proficiency. Because this framework requires each educator to be able to show competence, she or he needs to think about types of evidence to gather, striving to answer the key question, "How can I show that I am an effective teacher, or an effective principal?" Evidence collection should be ongoing, so that the educator is considering competency demonstration throughout the school year (and summer break), not just once or twice annually when observed by a supervisor. Placing the onus for data collection on the educator also enables the building principal to limit observational visits to classrooms. This framework is not without its drawbacks. The public—parents and local and state policy makers such as local school board members or state legislators—may not trust the types of evidence provided by educators, and each educator's collection of evidence may differ from the others. Supervisors currently may not be prepared to evaluate such broad and disparate sets of evidence. Thus, educators (and others) who review the collections of evidence will need both good examples of varied proficiency levels and training to judge the collections. Of the unanswered questions regarding this method, one may ask: If an educator submits a thinly documented collection of evidence, is this a demonstration of incompetence or of inability to collect good evidence?

Methods for Evaluating Teachers.

Classroom Observations. Classroom observations, the most common form of teacher evaluation, provide a useful measure of teachers' practice but little evidence of student learning. When choosing

an observation method, the administrator should select one that is based on what he or she hopes to learn from the process. Examples of widely used observation protocols include Danielson's Framework for Teaching and University of Virginia's Classroom Assessment Scoring System (CLASS) (Pianta, La Paro, & Hamre, 2006). Less widely used protocols, suitable for limited contexts, include the Reformed Teaching Observation Protocol for mathematics and science (Piburn & Sawada, 2000) and the Quality of Mathematics in Instruction for Mathematics (Blunk, 2007).

Like all methods, observation has both advantages and disadvantages. For example, observations can be conducted either by a school administrator or an outside evaluator; can measure general teaching practices or subject specific techniques; can be scheduled formally or unannounced and occur once or several times a year; and are often viewed as credible by a variety of stakeholders.

However, valid and appropriate instruments must be used to avoid bias, and observers must be well trained to use the instrument in standard ways for fair comparison across classrooms.

Standards-based Evaluations. Standards-based evaluations build on the traditional framework of teacher observations; however, this approach goes beyond simple classroom observations. In a rigorous system, the district develops a clear set of instructional standards and a detailed rubric that explains specific levels of performance for each standard (Papay, 2012). This type of evaluation stems from a burgeoning effort to define standards of instructional practice. Efforts like the National Board for Professional Teaching Standards and Charlotte Danielson's Framework for Effective Teaching pushed this conversation forward, and districts have begun developing instructional standards based on these frameworks (Papay, 2012). In fact, standards-based evaluations of practice have been found to be significantly related to student achievement gains and to help teachers improve their practice and effectiveness (Milanowski, Kimball & White, 2004).

Methods. Such systems for observing teachers' classroom practice are based on professional teaching standards grounded in the scholarship of teaching and learning. They use systematic observation protocols, administered by trained evaluators, to examine teaching along a number of dimensions, such as classroom organization and management, planning a well-organized curriculum, making subject matter accessible to students, assessing student learning, and differentiating instruction to meet student needs (Darling-Hammond, 2012). Successful systems include multiple classroom observations by expert evaluators across the year (Darling-Hammond, 2012).

Advantages. According to Papay (2012), standards-based evaluations afford a much richer view of a teacher's instructional practice because evaluators visit classrooms several times over the course of the year. Such evaluations are also based on clear evidence and standards, rather than administrators' hunches or judgments. Evaluators must justify all assessments with the evidence that they have collected during the observation. Finally, unlike traditional observations, in which most teachers report getting little useful feedback (Weisberg, Sexton, Mulhern & Keeling, 2009), these evaluations provide rich information about instructional practices and how these practices meet the district's standards.

Disadvantages. A high-quality, standards-based evaluation system requires rigorous instructional standards with clear rubrics that define success on these standards. Each district must adapt these existing frameworks to their local context and work carefully with both administrators and teachers to develop understanding, buy-in, and trust. There should also be standards of practice for the evaluation itself, including clear expectations about the level of evidence required to make a summative assessment and the extent of feedback provided to teachers (Darling-Hammond, 2012). Further, evaluators must be well trained, knowledgeable about effective teaching practices as defined in the

standards, and able to analyze observed practices to determine how well teachers are meeting these standards (Danielson, 2010). Measurement imprecision and bias can occur, but can be reduced by using multiple raters to observe the same teacher and by conducting more classroom observations on that teacher (Rowan & Raudenbush, 2016).

Principal Evaluation. Principal evaluation can vary widely, from a formal process using validated observation instruments to an informal, unannounced visit. As an advantage, principals are most knowledgeable about the contexts of their schools and their student and teacher populations. However, observations should be conducted several times a year to ensure reliability and include announced and unannounced visits to capture a complete picture of teacher practices, which—depending on the number of teachers who require evaluation—may detract from the principal's other duties (Little, Goe & Bell, 2009) Further, principals may not be well trained in methods of evaluation, or may lack the specialized knowledge needed to make informed judgements (Yon, Burnap, & Kohut, 2002). Not least, principals may be biased in either direction in their interpretation of teaching behaviors.

Student Feedback. Student evaluations most often come in the form of a questionnaire asking students to rate teachers based on a Likert-type scale. A well-designed questionnaire is key to measuring meaningful teacher behaviors to maintain the validity of the results (Little, Goe & Bell, 2009). Advantages of this approach include:

- Student ratings are cost-efficient and time efficient, can be collected unobtrusively, and can be used to track changes over time (Worrell & Kuterbach, 2001).
- Student ratings require minimal training.
- Student feedback has considerable credibility due to reliability from the high number of raters.

- Student feedback is observed over many hours, therefore based on representative behavior.
- Student observations are made by those who have been personally affected, so high face validity (Braskamp & Ory, 1994, Cashin, 1995).

However, disadvantages of this approach may also emerge, including but not limited to:

- Students may not take evaluation systems seriously.
- Student ratings may not be comparable from one faculty member to the next.
- Extraneous influences can come into play and cause bias.
- Technical and statistical support for interpretation of results may be lacking.
- Students may not be able to comment on considerations such as curricular design, subject mastery, goals, content and organization of course design, methods used in delivery, or practices used in evaluation of student work.

Teacher Self-Report Measures. Teacher self-report measures ask teachers to report on their classroom activities and can take the form of surveys, instructional logs, or interviews. They may focus on broad and overarching aspects of teaching or focus on specific subject matter, content areas, grade levels or techniques.

Advantages to this approach are significant. Self-report data can tap into a teacher's intentions, thought processes, knowledge and beliefs and be useful for self-reflection and formative purposes. Teachers are the only ones with full knowledge of their abilities, classroom context and curricular content and can provide insights observers may not recognize. Surveys tend to be cost-efficient, unobtrusive and capable of gathering a large array of data. It is important, however, to consider potential disadvantages. For example, self-reports may be subject to bias, and may lead to over- or underreporting of practices. Moreover, teachers may unintentionally misreport their practices

believing they are correctly implementing a practice when they are not (Little, Goe & Bell, 2009).

Teacher Portfolios. Teacher Portfolios, which have gained considerable popularity within colleges and universities, have not gained the same momentum in the K-12 grades as a technique for evaluating teachers. Portfolios are a collection of materials compiled by teachers to exhibit evidence of teaching practices, activities and student progress, for the distinct purpose of evaluation (Little, Goe & Bell, 2009). While methods for assembling portfolios can vary, documentation of teaching ability and teaching contribution can be evaluated from information in the following five areas (Berkeley Division of the Academic Senate, 2015):

1) **Course Design/Class Design.**

- Statement of teaching goals and objectives
- Detailed syllabi and course goals and objectives
- Teaching materials (tests, homework, reading lists, handouts, assignments)
- Graded term papers, projects, assignments

2) **Teaching Methodologies**.

- Statement of teaching goals and objectives
- Detailed syllabi and course goals and objectives
- Teaching materials (tests, homework, reading lists, handouts, assignments)
- Graded term papers, projects, assignments

3) **Content Knowledge**.

- Evidence in teaching materials
- Record of attendance at disciplinary-based conferences resulting in presentations to faculty or application to classroom

- Record of public lectures or performances, reviews of scholarly or creative work
- Record of student research directed
- Evidence of consultations and invitations related to teaching and participation in faculty colloquia

4) **Student Learning**.

- Statement of evaluation criteria for student grades and grade distribution
- Copies of graded exams, student papers, homework or other related products
- Successful drafts of student work, along with the professor's suggestions as to how each draft might be improved
- Student learning portfolio (e.g., pre- and post-test performance, videotape of student presentations at the beginning and the end of a course)
- Statements by alumni on their learning

5) **Departmental Responsibility**.

- Record of service on teaching-related committees (e.g., curriculum, faculty development)
- Evidence of design of new courses and programs
- Evidence of involvement in student advising or career development
- Record of teaching load, class size and teaching improvement activities

As with other evaluation approaches, both advantages and disadvantages can be noted. For example, portfolios are comprehensive and can capture effective teaching that occurs both inside and outside of the classroom. Moreover, they can be specific to any grade level, subject matter, or student population. It should, though, be noted that their use for summative or high stakes assessment has not

been validated. Moreover, scoring may be difficult, including the process of obtaining reliability between scorers. Further, it is difficult to determine what materials are accurate representations of a teacher's practice. From an educator's perspective, developing portfolios and their corresponding reflections can be considered a time burden.

Analyses of Instructional Artifacts. This method uses lesson plans, teacher assignments, assessments, student work, and other artifacts to determine the quality of instruction in a classroom. Evaluators can use these artifacts to gain a better understanding of day-to-day learning opportunities for students. Artifacts may be judged on a wide variety of criteria, including rigor, authenticity, intellectual demand, alignment to standards, clarity, and comprehensiveness (Little, Goe & Bell, 2009). Like many approaches, it contains both advantages and disadvantages. For example, because artifacts have already been created by teachers they do not appear to place unreasonable burden on teachers (Borko et al., 2005). Moreover, examination of classroom artifacts may be less labor intensive and costly than full classroom observations. Among potential detriments is the need for specialized knowledge. Accurate scoring is essential for validity. Scorers must be well trained, calibrated, and may need to possess knowledge of the subject matter being evaluated.

Best Practices: Feedback/Data use

This section discusses best practices from the literature on providing feedback on teacher evaluations and the use of collected data**.** Evaluations are best viewed as a continuing process. Thus, on the whole, evaluation systems should be judged not only by the quality of measurement but also by the quality of the targeted feedback they provide and their ability to drive continued instructional improvement (Papay, 2012).

After a formal evaluation, teachers should receive clear and actionable feedback based on standards

for teaching and student growth and learning. These standards must be comprehensive and transparent and must be based on criterion-referenced assessments of teacher practice (Lok, McNaught & Young, 2016; National Educational Association, 2011). The best evaluation and administration teams give teachers opportunities to improve in the areas in which they score poorly, providing assistance in determining problem areas and planning strategies to address them (Goe & Croft, 2009). If, following such supportive coaching and appropriate subsequent evaluations and due process, teacher performance is not at least minimally acceptable, schools and districts must have in place—and must make use of—manageable procedures for termination.

Best Practices: Principal Leadership and Student Learning

This section explores research the role principal leadership style plays in student learning. Much of Leithwood's widely cited research on school leadership (Leithwood, Menzies, Jantzi, and Leithwood, 1996; Leithwood, 2002; Leithwood, Seashore, Anderson & Whalstrom, 2004) and student learning draws on industrial and organizational theory. Leithwoood's framework takes into account state and district politics, district leadership, policies, practices that interact to exert influence, classroom conditions, student and family background, school leadership, other stakeholders, school conditions, and teachers. Successful principal leadership empowers others, provides instructional guidance, develops school improvement plans, is fair or embraces "equity." The individual teacher needs basic skills, subject matter content, pedagogical skill and knowledge and classroom experience.

Leithwood's research findings are cited in a dissertation, authored by Brenda Kay Hardman, titled "Teachers Perceptions of their Principal leadership style and the effects on student achievement in improving and non-improving schools" (2011). Hardman raises several questions in this dissertation; the two that follow are most relevant to the purpose of this review.

1. How do teachers in improving and non-improving schools perceive the leadership of their principals?

2. What are the behaviors of school principals that influence student achievement (math and reading) as perceived by teachers?

Leithwood's study explored three types of leadership styles: *transformational*, *transactional,* and *passive-avoidant,* in a secondary school setting. The primary assumption was that a certain leadership style, one that promoted collaboration and support of the professional development and engagement of teachers, would create a positive learning environment for pupils that would enhance schools' capacity to improve achievements. Teachers must be engaged in decision-making and view the principal—who has frequent contact with them—as a role model. In other words, the perceived leadership style of the school principal impacts the school's culture and performance.

Leithwood found that teachers in improving and non-improving schools reported minimal differences in how they perceived their principals' leadership styles. Essentially, a place existed for all three leadership styles. Each had a significant effect on predicting student achievement, but in terms of improving and non-improving schools there was no difference. Three themes emerged from the teacher's beliefs about principal leadership style: principal role modeling, school culture, and leadership decisions.

School culture was the highest-reported theme by both improving and non-improving high schools. In improving elementary schools, 85% of teacher responses indicated *school culture* had an impact on teaching and student achievement. The author suggests that a positive school culture can be built through teacher-focused leaders who build teacher capacity.

Question 2: What are the Barriers to Implementation?

Introduction

A number of challenges, arising both within and outside of schools, hinder the implementation of teacher evaluations. In this section we identify those that are within-school specific, including cultural, technical, and political challenges. We then turn to those that are located outside the school, focusing on 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* (has systematic error, or lack of objectivity, been introduced?); *validity* (are the measures accurately measuring teacher performance?); and *reliability*, *(*if performed again, would this evaluation yield the same results?)

Barriers to implementing teacher evaluations

School-based challenges. School-based challenges can range from a lack of institutional support and leadership turnover, to administrative deficiencies and training deficits, to inadequate human resource support, limited financial allocations for training and/or merit raises, and collective bargaining barriers.

Ingram, Seashore, and Schrodeder (2004), for example, identified cultural, technical, and political barriers to establishing a school culture supportive of data-based decision-making. Often implicit, an organization's culture is grounded in deeply held values and exerts a powerful influence on how decisions are made and organizations learn, and on the data teachers find meaningful and useful. Thus, many teachers have developed personal metrics for judging their teaching effectiveness, and these often differ from external metrics (e.g., those of state accountability systems and school boards). Moreover, many teachers and administrators base decisions on experience, intuition, and anecdotal

information (professional judgment) rather than on systematically collected data, and some may disassociate their performance from outcome-oriented effectiveness, identifying only a modest relationship between their efforts and student achievement. These cultural factors are complicated by division among stakeholders regarding which student outcomes are most important and what kinds of data are meaningful. Technical and political impediments also abound. For example, data that teachers want about ‘‘really important outcomes’’ are rarely available and are usually hard to measure. For their part, schools rarely provide the time needed to collect and analyze data. Alternatively, they may be experiencing data overload: That is, they may have tons of data, but often the data are in the wrong form or teachers have to spend an inordinate amount of time to get it. Not least, data have often been used politically, leading to mistrust of data and data avoidance.

Community-based Barriers. Community-based barriers to implementing evaluation will vary depending on the community context. For example, in economically challenged communities, there may already be such a demand on limited resources that stakeholders have limited capacity to attend to, become informed about, and address education quality issues to the extent required to effect lasting change. Impoverished communities may have numerous economic challenges and social needs, while the school has limited educational staff and financial resources to meet these needs.

State/National-based Barriers. State and national-based Barriers include changes in the philosophy of education at the national and state levels of government, as well as changing trends in education. For example, over several decades, federal educational policies have gone from being very corrective in nature (e.g., holding schools to a higher level of accountability), to letting local governments determine and apply policies as they see fit.

Barriers to Implementing Specific Evaluation Design/Methods. In addition to traditional barriers

to implementing specific evaluation tools—among them bias, reliability, and validity (discussed below), personal relationships can also hinder implementation. Classroom observations provide a useful measure of teachers' practice, but they offer little evidence of student learning. In one recent study of an evaluation program, researchers found that a key limiting factor in the implementation of teacher evaluations is principals' unwillingness to identify teachers as not meeting standards: telling teachers they are not doing a good job is difficult (Johnson et al., 2009). Effective evaluators must be willing to provide tough assessments and to make judgments about the practice, not the person. Although administrators have traditionally conducted evaluations, several districts across the country have experimented quite successfully with engaging expert peers to serve as evaluators (Johnson, Fiarman, Munger, Papay & Qazilbash, 2009), thus separating evaluation of practice from personal relationship.

Bias. Standards-based evaluations that rely on classroom observations have been seen as subjective and bias- ridden (Papay, 2012). Limiting bias in standards-based observations presents challenges because such observations rely on human judgments. As mentioned above, evaluators report that it is difficult to separate what they know of the teacher, or the teacher's contributions outside of the classroom, from their judgments of the teacher's instructional practice. However, having clear standards, using highly qualified and well-trained evaluators, and using evaluators from outside of the school if possible can help to reduce subjective bias (Whitehurst, Chingos & Lindquist, 2014).

Although value-added models are based on objective test scores, not personal judgments, this alone does not eliminate bias. Possibly the largest threat that may bias value-added estimates is the extent to which value-added models can fully account for differences in student assignments. Value-added models typically account for a wide variety of student characteristics, explicitly comparing the

performance of students with similar test-score histories. If value-added measures hope to isolate a teacher's contribution to student achievement growth, they must fully account for these differences in students taught, both within and across schools (Darling-Hammond, 2012).

Reliability. Reliability is a widely reported concern with classroom observations (Little, Goe & Bell, 2009). Because high-quality observations are time consuming, evaluators must make judgments based on a relatively limited sample of instruction. A common criticism of traditional evaluation is that observations are announced, so teachers can prepare and execute an effective lesson on the day that they are observed. Different evaluators may have different standards, achieving sufficient inter-rater reliability may be difficult. It is entirely possible that two different evaluators could rate the same teacher's practice differently. Teachers should not be rewarded (or punished) simply for having an easy (or tough) evaluator.

Researchers have attempted to quantify the variability in value-added measures in several ways. First, they have examined whether teachers' value-added estimates are similar from year to year. Although the specific results depend on the dataset and model used, most studies find moderate-sized year-to-year correlations that average between 0.4 and 0.5 (e.g., McCaffrey, Sass, Lockwood, & Mihaly, 2009). These figures would represent substantial changes in teacher effectiveness from one year to the next.

Researchers have also examined value-added estimates from two different tests of the same content area in the same year (Papay, 2011), on the assumption that, if the tests measure similar material, estimates of a teacher's effectiveness using either test should be quite similar. However, even when they include multiple years of data, these correlations typically range between 0.3 and 0.5 (Corcoran, Jennings, & Beveridge, 2011; Gates Foundation, 2010; Lockwood et al., 2007; Papay, 2011). These

estimates are not sufficiently reliable to consistently classify the highest- and lowest-performing teachers.

Validity. One key challenge in assessing the validity of evaluation measures is defining what a district hopes to measure. This distinction between evaluating whether teachers promote *student learning* or whether they raise *student test scores* has important implications for assessing both standards-based observations and value-added approaches (Darling-Hammond, 2012). Currently, value-added estimates are only practical for teachers of annually tested subjects (typically mathematics and English language arts in grades 4–8). High school, early elementary school, history, science, and arts teachers are thus excluded. In recent years, several researchers have attempted to "validate" observational measures by comparing teachers' evaluation ratings to value-added estimates. Standards-based evaluations are relatively strong predictors of teachers' value-added measures (Grossman, Loeb, Cohen & Wyckoff, 2013; Hill, Kapitula, & Umland, 2011). In fact, Kane, Taylor, Tyler and Wooten (2011) found that a teacher's standards-based evaluation rating actually predicts student test performance above and beyond that teacher's value-added rating.

Question 3: How do we evaluate the Measurement of Student Growth?

Introduction

In this section, we explore two primary categories of evaluating student growth: value-added models, and the assumptions underlying them, and alternative methods. Before proceeding, it is essential to clarify just what is intended by *student growth.* In an evaluation context the term refers to change in students' knowledge and skills over time, and it is in this sense that we employ it. As it relates to educator evaluation, student growth can be one measure of educator effectiveness—if the changes

attributable to instruction that students receive can be isolated and estimated. *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 they may be deployed as evaluative tools regardless of grade or subject. 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.

Student Growth and Assessment

Educator evaluation focuses on estimating the changes that can be attributed to instruction students receive. According to the 2013 Michigan Council for Educator Effectiveness (MCEE) report, this can be done by:

1. Using changes in students' scores on two administrations of a particular assessment. (Although this alone does not provide a fair estimate of how much change is due to teaching.)
2. When standardized tests are used to measure student learning, using statistical techniques called *value-added models.* (These are designed to isolate the contribution of instruction by controlling for other factors that might impact students' growth.)
3. Providing evidence of the students' progress toward a set of articulated learning goals.

The use of student growth—a student's progress between two points in time—in educator evaluations constitutes a high-stakes use of assessment data. According to the MCEE (2013) it is important to remember there are multiple ways to measure growth. While giving students the same test at time A and time B might seem like the most sensible approach, it involves testing students on material they have never studied, which is educationally problematic. Moreover, assessments useful for measuring student growth can be provided in multiple ways: *full-service assessments* are developed,

administered, scored, and reported on centrally, using rigorous professional measurement practices; *model assessments* are developed centrally, using rigorous professional measurement practices, but administered, scored, and reported on locally; *locally developed assessments* are developed, administered, scored, and reported on locally; and *student learning objectives* (SLOs), or specific measurable academic goals that teachers and evaluators set for groups of students.

Teacher ability does not account for all of the variance in student achievement. Hattie (2003) reported on the sources of variance in student achievement. Using Hierarchical Linear Modeling, the major sources of variance were identified as: students (accounted for about 50% of achievement variance); home (5 to 10% of variance); schools (5 to 10% of variance); peer effects (5 to 10% of variance); and teachers (30% of variance). Home effects were more related to levels of expectation and encouragement, and certainly not a function of the involvement of the parents or caregivers in the management of schools. Principals were included in school variance, mainly because of their influence on the climate of the school. While peers can have a positive effect on learning, the discussion is dominated by concern for bullying and for peer reputations built on almost anything other than pride in learning.

Formative and Summative Assessment

Like evaluation of teaching practices, assessment of student growth can be either *formative* or *summative*. *Formative assessment* aims to *monitor student learning* to provide ongoing feedback instructors can use to improve their teaching and students can use to improve their learning. More specifically, formative assessments help students identify their strengths and weaknesses and target areas that need work, and help teachers recognize where students are struggling and address problems immediately. Formative assessments are generally *low stakes*, carrying low (or no) point values. A

formative assessment may ask a student to draw a concept map in class to represent their understanding of a topic; submit one or two sentences identifying the main point of a lecture; or turn in a research proposal for early feedback. By contrast, summative assessment *evaluates student learning* at the end of an instructional unit by comparing it against some standard or benchmark. Summative assessments are often *high stakes*, carrying a high point value. Summative assessments may include a midterm exam, final project, paper, or senior recital. Information from summative assessments can be formative when students or faculty use it to guide their efforts and activities in subsequent courses.

Value-Added Models

The statistical method known as value-added assessment is a way of isolating the effect of instruction on student learning. These models use value-added student achievement test scores from state or district standardized tests as a key measure of teachers' effectiveness.

Value-added approaches can separate students' annual academic growth into two parts: that which can be attributed to the student, and that which can be attributed to the classroom, school or district (Hershberg, 2005). Several different value-added models (VAM) are in use today, although use and requirements of VAM vary by state.

Assumptions of Value-added Models. Reardon and Raudenbush (2009) state that the ability of a school's (or teacher's) value-added model to provide unbiased estimates of school (or teacher) effects rests on a set of assumptions. They identified six assumptions necessary to make valid causal inferences about school effects on student learning: manipulability, lack of interference between units, an interval scale metric, homogeneity of effects, strongly ignorable assignment and functional form.

They found that "modest violations of these assumptions degrade the quality of value-added estimates, but models that explicitly account for heterogeneity of school effects are less affected by violations of the other assumptions" (p. 2).

Darling-Hammond (2015) observes that whether value-added ratings will ultimately improve or undermine teacher evaluation depends largely on whether VAM metrics can accurately identify an individual teacher's contributions to student learning and offer a credible measure of teacher "effectiveness." Darling-Hammond suggests that VAMs could accurately identify individual contributions given a set of ideal conditions. Specifically:

- Student learning must be well-measured by tests that reflect valuable learning and the actual achievement of individual students along a vertical scale representing the full range of possible achievement measured in equal interval units.
- Students must be randomly assigned to teachers within and across schools—or, conceptualized another way, the learning conditions and traits of the students assigned to one teacher do not vary substantially from those assigned to another.
- Individual teachers are the only contributors to students' learning over the period of time during which gains are measured. (Of course, none of these assumptions holds, and the degree of error in measuring learning gains and attributing them to a specific teacher depends on the extent to which they are violated, as well as the extent to which statistical methods can remedy these problems.)

Advantages. According to Papay (2012), value-added approaches are attractive to policy makers for several reasons. With the growing focus on test-based accountability, these measures directly assess student test-score growth. Because they are based on external assessments, VAMs are seen as objective. Moreover, given the accessibility of datasets, these measures can be fairly easy and

inexpensive to estimate. Because individual students, rather than cohorts, are traced over time, each student serves as his or her own "baseline" or control. This removes much of the influence of the unvarying student characteristics, such as race or socioeconomic factors (Hershberg, 2005).

In its substantive study of value-added models, McCaffrey and Hamilton (2007) identified important research questions about school effects; the comparability of instructional difficulty at different grade levels and subjects; and the quality of the tests used. It concluded that the teacher effect is real, could be quite large, and persists beyond the first year in which it is evident. McCaffrey and Hamilton also concluded that value-added models "might actually provide less-biased and more precise assessments of teacher effects" than existing test-based systems. Moreover, as "policymakers evaluate alternative models for school or teacher accountability, VAM should be given serious consideration even in light of its limitations" (Hershberg, 2005; McCaffrey et al., 2003).

Disadvantages. Despite numerous advantages, researchers highlight cautions about basing individual teacher evaluations on annual student test scores. Concerns with value-added achievement measures are varied. For example, in many states, state testing occurs only in grades 3–8; even then, such testing occurs annually only in reading and mathematics (with testing in other subjects occurring less frequently and, in many subjects, not at all). For this reason, state assessment data are typically available for only approximately 25–40% of the teachers in any given school system. As a result, these data suffer from what sampling statisticians call coverage error—a failure of the sample of teachers for which value-added estimates of teaching effectiveness can be calculated to fully represent the population (Rowan & Raudenbush, 2016).

Current tests, which focus narrowly on basic skills and use primarily multiple-choice questions, raise concerns about teaching to tests at the expense of other kinds of learning, such as writing, inquiry, and

complex problem solving (Baker et al., 2010). Other studies have found that teachers' measured effectiveness differs significantly depending on the tests used (Lockwood et al., 2007). Teachers who rate highly on VAM estimates of achievement on basic skills tests are often rated lower when evaluated more conceptually.

The American Statistical Association (2014), in its statement on VAMs, noted that: most VAM studies find that teachers account for between 1 and 14% of the variability in test scores, and that the majority of opportunities for quality improvement are found in system-level conditions. Ranking teachers by their VAM scores can have unintended consequences that reduce quality. A few non-teacher factors are measured in some VAM models; most assume that controlling for prior test scores resolves unmeasured influences on gains, but unmeasured variables that influence achievement gains become part of what Darling-Hammond (2015) terms the "teacher effect."

Rowan and Raudenbush (2016) state that value-added measures of teaching performance based on student achievement data have measurement error associated with test(s) used to construct the measures; samples of students over which the measures are constructed; and measurement instabilities. Measurement precision can be improved and potential distortions decreased if VAMs are based on data from more students, on more occasions, using achievement data from multiple tests. They also state that measuring teaching performance can lead to distortion and risk in the teacher evaluation process. When performance measures are characterized by risk and distortion, organizations will be more likely to benefit from subjective (versus objective) performance assessment. However, objective measures can provide a basis for specific narrative feedback and for setting clear, attainable improvement goals as part of the evaluation process. Value-added measures provide fewer opportunities for such improvement goals than teacher observation measures. Additionally, teachers are often unclear in their understanding of the models being used to evaluate

them, making it difficult for the assessment to lead them to improvements in teaching methods (Amrein-Beardsley, 2008). Thus, the value added method needs further validity research before wide implementation is justified (Amrein-Beardsley, Pivovarova and Geiger, 2016).

Recommendations for Value-added Models. The American Educational Research Association (2000) issued 12 recommendations for high-stakes testing. According to Amrein-Beardsley (2008), the six most relevant to VAM include:

1. High-stakes decisions should not be made on the basis of a single test score.
2. High-stakes tests must be validated for each intended use.
3. Negative side effects of a high-stakes assessment program must be fully disclosed to policy makers.
4. The accuracy of achievement levels must be established.
5. Students with disabilities must be offered appropriate accommodations.
6. Intended and unintended effects of the testing program must be continuously evaluated and disclosed.

One challenge in using additional student-learning measures in teacher evaluations is developing measurement and feedback protocols based on more effective evaluation feedback on student learning. Improved feedback will need to be richer and more frequent than current annual performance feedback produced by a complex, numerical, "value-added" score. In its present form, such feedback provides teachers little to no specific guidance for improvement (Rowan & Raudenbush, 2016).

Alternative indicators of student growth

Student Learning Objectives. Student learning objectives (SLOs) are classroom-specific growth targets chosen by individual teachers and approved by principals (Darling-Hammond, 2012; Gill, Bruch & Booker, 2013). SLOs are becoming popular as alternative measures of student growth because they can be used to evaluate teachers in any grade or subject (Gill, Bruch & Booker, 2013). According to Slotnik and Smith (2008), a rigorous, high-quality SLO has a number of key elements:

1. Clear identification of the student population

Ensuring that teachers are accountable for the academic progress of all students.

2. Specific time period.

A clear timeline within which students will reach an academic goal.

3. Assessment(s) of student progress connecting teacher, student, and expectations

The best guidance for implementing SLOs includes information on the attributes of high-quality assessments and tools to create them.

4. Rigorous, yet realistic, expected growth or achievement target to be met by students

5. Strong rationale for the expected student growth

6. Strategies for achievement

For an SLO to be an instrument of good instructional practice, rather than simply an evaluative tool, teachers need to be able to identify specific approaches they will use in the classroom to meet student growth expectations.

Methods. According to Tyler (2011), teachers follow a series of steps to set effective SLOs. At the beginning of the semester or year, teachers review available data on the students in the class, including prior-year test performance and any course pre-tests administered. Based on the data, they set a designated number of objectives, usually two, that are classroom wide and student- or subgroup-specific. (A class-level objective might be something like, “Increase the Algebra I end-of-course pass

rate by 5 percentage points over last year's 85 percent pass rate." Following identification, teachers set appropriate measures against which attainment will be judged. Following the above example, the objective attainment measure is the class pass rate on the Algebra I end-of-course exam).

Advantages. Although very little SLO literature addresses statistical properties, key findings show that SLOs have the potential to better distinguish teachers based on performance than traditional evaluation metrics do. For example, research on the Denver Public Schools District's use of SLOs found that rigorous and high-quality growth objectives were associated with higher student achievement (Community Training and Assistance Center [CTAC], 2004).

According to Gill, Bruch and Booker (2013) and CTAC (2004), a host of benefits are associated with the effective use of SLO. Not only are SLOs good instructional practice, but they enable schools and districts to make this best practice a common expectation across the teacher and principal workforce. Indeed, they are adaptable and can be used not only by individual teachers, but also with groups of teachers or the whole faculty. Because SLOs are most often developed through principal-teacher collaboration, they may reinforce the credibility of the evaluation process and build ownership for student results among teachers and principals. On a broader level, SLOs may help educators buy in to state and district evaluation systems.

Disadvantages. Student learning objectives are not without their drawbacks. They can require substantial training and technology infrastructure, and developing and monitoring them can be time-consuming for teachers and evaluators alike (Gill, Bruch & Booker, 2013; CTAC, 2004). Moreover, it can be difficult to ensure the quality of SLO and the assessments used to measure objective attainment.

Recommendations for SLO Implementation. Experienced practitioners of SLO evaluation methods have amassed a wealth of ideas for successfully implementing this method of assessment (Reform Support Network, 2015). These include:

1. Develop an online library of SLO resources.

2. Provide teachers and evaluators with thorough and rigorous professional development.

3. Provide guidance for how to choose and develop high quality assessments.

4. Regularly analyze and compare SLO data with teacher-level value-added data, observation ratings, and other predictors of future student success.

This will align, and encourage the improvement of, all performance measures over time.

5. Provide support for evaluator calibration sessions.

6. Spot check SLOs.

Allow states or districts to randomly select and review SLOs and give feedback. Allow them to examine patterns of SLO attainment, including disproportionate SLO attainment by certain schools or in certain subject areas, flagging objectives could benefit from quality review.

7. Hold administrators accountable for SLO quality.

In order for SLO to yield useful information, they must be appropriately rigorous, cover a reasonably representative portion of the curriculum, and be accompanied by tools to measure students' progress toward goals. This requires training so that teachers and administrators can create, implement, and assess student growth on SLOs in ways that ensure fair standards and high quality (MCEE, 2013). In many instances teachers must review and discuss their SLOs with their principal (and sometimes central office staff) as a part of the SLO-setting process. The principal (and central office staff when appropriate) must then approve and sign off on each teacher's SLOs. While there is little research to date on SLOs, principal and/or central office approval is likely a key part of the SLO process. For

example, one pay-for-performance pilot program in 17 Denver schools, including two middle- and two high schools, found that 89 to 93 percent of teachers in the pilot met their objectives over the four-year pilot (1999–2003). Further, the quality of teachers' SLOs increased over the four years (CTAC, 2004).

A significant amount of research remains to be done on SLOs. For example, future studies should look at the relationship between SLOs and student achievement gains, not just levels (Tyler, 2011). In addition, little is known about whether SLOs can yield ratings that correlate with other measures of teacher performance. More research is needed as states and districts roll out SLOs as teacher evaluation measures and instructional planning tools (Gill, Bruch & Booker, 2013). Finally, it is not clear at this point how effective the SLO evaluation process can be in identifying teachers who are differentially effective (Tyler, 2011).

Additional Alternative Indicators of Student Growth

Darling-Hammond (2012) examines a number of additional alternative indicators of growth. These include curriculum-based assessments, test achievement results, student work, student presentations, and project-based inquiries.

Locally Developed, Curriculum-based Assessments. Much less is known about measuring growth based on locally developed, curriculum-based assessments as indicators of student growth. Only two studies—conducted by the same research team and examining data gathered in Pittsburgh—have examined the use of value-added models incorporating locally developed curriculum-based assessments. Both reported encouraging results in the efficacy of these assessments to reliably distinguish among teachers at the end of the performance distribution (Johnson, Lipscomb, Gill,

Booker, & Bruch, 2012; Lipscomb, Teh, Gill, Chiang & Owens, 2010).

Local and District-wide Achievement Test Results. Local and district-wide achievement test results may be comprised of learning pre- and post-tests conducted by districts or schools, or even learning evidence assembled by teachers themselves. Such evidence can be drawn from classroom assessments and documentation, using instruments like the Developmental Reading Assessment. They can also draw on pre- and post-test measures of student learning in specific courses or curriculum areas, and can be developed by individual teachers, departments, school faculty, or district faculty or staff.

Student Work that Shows Evidence of Growth. Student work that shows evidence of growth is a record of student accomplishments in response to teaching activities, such as research papers, science projects, artwork, and so forth.

Teacher-generated Information about Student Growth and Goals. Some districts use evidence from teachers' careful documentation of learning of a set of diverse students over time, like that included in National Board Certification portfolios (National Board of Professional Teaching Standards, 2012). Analysis of standardized test results could be included, where appropriate, with attention to the relationship of the tests to the curriculum and their appropriateness for the students being taught.

Additional Alternative Methods

This review of alternative methods is not comprehensive. Others, about which less is known, include subject matter assessment, i.e., testing of the level of knowledge mastered by individuals in a specific topic, students' oral and written presentations to demonstrate learning; and project-based inquiry activities, an approach to teaching in which students explore real-world problems and challenges.

Evaluations of Effective Engagement, Critical Thinking, Self-efficacy or a Combination Judgment Model. Darling-Hammond (2015) calls for a more thoughtful approach to examining student learning in teacher evaluations. Under this proposal, teachers could combine evidence from multiple sources in a *judgment model*, using a matrix to combine and evaluate several pieces of student-learning data. They could then integrate the judgment model rating with those from observations and professional contributions.

Teachers receive low or high ratings when multiple indicators point in the same direction. Rather than merely tallying disparate percentages and urging administrators to align their observations with inscrutable VAM scores, this approach would identify teachers who warrant intervention, while enabling pedagogical discussions among teachers and evaluators based on evidence that connects what teachers do with how their students learn. A number of studies suggest that teachers become more effective as they receive feedback from standards-based observations and as they develop ways to evaluate their students' learning in relation to their practice (Heneman, Herbert, Milanowski, Steven & Odden, 2006; Goe, Biggers & Croft, 2012).

REFERENCES

Akiba, M. (2017). Editor's Introduction: Understanding Cross-National Differences in Globalized Teacher Reforms. *Educational Researcher*, 46 (4), 153-168.

American Educational Research Association (2000). *AERA Statement on High-stakes Testing in Pre-K-12 Education. Position Statement.* Available at https://www.aera.net/policyandprograms/?id=378

American Statistical Association. (2014). *ASA Statement on Using Value-added Models for Educational Assessment.* Alexandria, VA: Author.

Amrein-Beardsley, Pivovarova, M, & Geiger, T. (2016) Value-added model: What the experts say. Phi Delta Kappan, 98 (2), 35 - 40

Amrein-Beardsley, A. (2008) Methodological concerns about the education value-added assessment system. *Educational Researcher, 37*(2), 65-75.

Baker, E.L. Barton, P.E., Darling-Hammond, L., Haertel, E., Ladd, H.F., Linn, R L., …Shepard, L. A. (2010). Problems with the Use of Student Test Scores to Evaluate Teachers. Economic Policy Institute Briefing Paper # 278. Retrieved from http://files.eric.ed.gov/fulltext/ED516803.pdf

Berkeley Division of the Academic Senate (2015, April). Evaluation of Teaching. Report by the Committee on Teaching (COT). Retrieved from: http://academic-senate.berkeley.edu/sites/default/files/committees/cot/annual_reports/cot_teaching_evaluation_proposal_april_2015.pdf

Bill & Melinda Gates Foundation. (2010). Learning about teaching: Initial findings from the Measures of Effective Teaching Project. Seattle, WA: Author.

Blunk, M. L. (2007, April). *The QMI: results from validation and scale building.* Paper presented at the annual meeting for the American Educational Research Association, Chicago.

Borko, H., Stecher, B. M., Alonzo, A., Moncure, S., and McClam, S. (2005). Artifact packages for measuring instructional practice: A pilot study. *Educational Assessment*, 10(2), 73-104.

Bowen, G. L., Rose, R. A., & Ware, W. B. (2006). The reliability and validity of the School Success Profile Learning Organization measure. *Evaluation and Program Planning, 29*(1), 97-104.

Braskamp, Larry A. & Ory, John C. (1994). Assessing faculty work: Enhancing individual and institutional performance. San Francisco, CA: Jossey-Bass.

Cashin, W. (1995). Student ratings of teaching: The research revisited (IDEA Paper No. 32). Manhattan, KS: Kansas State University Center for Faculty Evaluation and Development.

Community Training and Assistance Center. (2004) *Catalyst for Change: Pay for Performance in Denver Final Report.* Retrieved from: http://www. ctacusa.com/PDFs/Rpt-CatalystChangeFull-2004.pdf.

Corcoran, S., Jennings, J., & Beveridge, A. (2011). Teacher effectiveness on high- and low-stakes tests. Mimeo, New York University, Institute for Education and Social Policy.

Darling-Hammond, L. (2012). *Creating a comprehensive system for evaluating and supporting effective teaching*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher, 44* (2), 132–137.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development

Danielson, C. (2010). Evaluations That Help Teachers Learn. *The Effective Educator*, 68(4), 35-39

Danielson, C. (2013). The Framework for Teaching: Evaluation instrument. Princeton, NJ: Danielson Group.

Davis, D., Ellett, C., & Annunziata, J. (2002). Teacher evaluation, leadership and learning organizations. *Journal of Personnel Evaluation in Education 16*(4), 287-301.

Ellett, C., & Teddlie, T. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of Personnel Education in Education, 17*(1), 101-128.

Gill, B., Bruch, J, & Booker, K. (2013). Using alternative student growth measures for evaluating teacher performance: what the literature says. (REL 2013–002). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from: http://ies.ed.gov/ncee/edlabs.

Goe, L., Bell, C., & Little, O. (2008, June). Approaches to evaluating teacher effectiveness: A research synthesis. Washington, DC: National Comprehensive Center for Teacher Quality.

Goe, L., Biggers, K., & Croft, A. (2012). Linking Teacher Evaluation to Professional Development: Focusing on Improving Teaching and Learning. Washington, DC: National Comprehensive Center for Teaching Quality.

Goe, L. & Croft, A. (2009). *Research-to-practice brief: Methods of evaluating teacher effectiveness.* Washington, DC: National Comprehensive Center for Teaching Quality.

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added. American Journal of Education, 119(3), 445-470.

Hardman, B. K. (2011). *Teacher's perception of their principal's leadership style and the effects on student achievement in improving and non-improving schools.* (Doctoral dissertation). Retrieved from: http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=4921&context=etd Hattie, J.A.C. (2003, October). *Teachers make a difference: What is the research evidence?* Paper presented at the Australian Council for Educational Research "Building Teacher Quality: What Does the

Research Tell Us" Research Conference, Melbourne, Australia.

Heneman III, Herbert G.; Milanowski, Anthony; Kimball, Steven; and Odden, Allan. (2006). Standards-Based Teacher Evaluation as a Foundation for Knowledge- and Skill-Based Pay. CPRE Policy Briefs. Retrieved from http://repository.upenn.edu/cpre_policybriefs/33

Hershberg, T. (February, 2005). *Value-added assessment and systemic reform: A response to America's human capital development challenge*. Presented at the Aspen Institute Congressional Institute, The Challenge of Education Reform: Standards, Accountability, Resources and Policy, Cancun, Mexico.

Hill, H. C., Kapitula, L. R., & Umland, K. L. (2011). A validity argument approach to evaluating value-added scores. *American Educational Research Journal, 48*(3), 794-831.

Ingram, D., Seashore, L. K., & Schrodeder, R. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106*(6), 1258–1287.

Ingvarson, L., & Rowley, G. (2017). Quality Assurance in Teacher Education and Outcomes: A Study of 17 Countries. *Educational Researcher*, 46 (4), 177-193.

Johnson, S. M., Fiarman, S. E., Munger, M. S., Papay, J. P., & Qazilbash, E. K. (2009). A user's guide to Peer Assistance and Review. Retrieved from http://www.gse.harvard.edu/~ngt/ par

Johnson, M., S. Lipscomb, B. Gill, K. Booker, and J. Bruch. "Value-Added Models for the Pittsburgh Public Schools." Report to the Pittsburgh Public Schools. Cambridge, MA: Mathematica Policy Research. 2012. Retrieved from: http://mathematicampr.com/publications/redirect_PubsDB.asp?strSite=PDFs/education/valueadded_pittsburgh.pdf

Kane TJ, Staiger DO. 2008. Estimating teacher impacts on student achievement: an experimental

evaluation. NBER Working Paper 14607.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practice using student achievement data. *Journal of Human Resources*, 46(3), 587–613.

Leithwood, K., Menzies, T., Jantzi, D. & Leithwood, J. (1996). School restructuring, transformational leadership and the amelioration of teacher burnout. *Anxiety, Stress, and Coping*, *9*, 199-215.

Leithwood, K., Seashore, L. K., Anderson, S., & Wahlstrom, K, (2004). How leadership influences student learning. *Report*. Commissioned by the Wallace Foundation. Retrieved from: http://www.wallacefoundation.org/knowledge-center/Documents/How-Leadership-Influences-Student-Learning.pdf

Leithwood, K., (2002). *Organizational learning and school improvement*. Greenwich, CT: JAI Press.

Lipscomb, S., Teh, B., Gill, B., & Chiang, H., & Owens, A. (2010). Teacher and Principal Value-Added: Research Findings and Implementation Practices. Mathematica Policy Research, Final Report. Retrieved from: http://files.eric.ed.gov/fulltext/ED531785.pdf

Little, O., Goe, L., & Bell, C. (2009). *A practical guide to evaluating teacher effectiveness*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from: www.tqsource.org/publications/practicalGuide.pdf

Lockwood, J. R., McCaffrey, D. F., Hamilton, L.S., Stetcher, B., Le, V. N., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement, 44*(1), 47–67.

Lok, B., McNaught, C. & Young, K. (2016). Criterion-referenced and norm-referenced assessments: compatibility and complementarity. *Assessment & Evaluation in Higher Education*, 41(3)

Marzano, R.J. (2007). The art and science of teaching. Alexandria, VA: ASCD

Marzano, R. J., Frontier, T., & Livingston, D. (2011). Effective supervision: Supporting the art and science of teaching. Alexandria VA: ASCD

Marzano, R. J. (2012). Marzano levels of school effectiveness. Bloomington, IN: Marzano Research Laboratory. Retrieved from: http://legisweb.state.wy.us/InterimCommittee/2012/Z02MarzanoLevels.pdf

McCaffrey, D. F., & Hamilton, L. S. (2007). *Value-added assessment in practice: Lessons from the Pennsylvania Value-Added Assessment System pilot project.* Santa Monica, CA: RAND Corporation.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). Evaluating value-added models for teacher accountability, (MG-158-EDU). Santa Monica, CA: RAND.

Michigan Council for Educator Effectiveness. (2013). Building an improvement-focused system of educator evaluation in Michigan: Final Recommendations. *Report*. Retrieved from: file:///C:/Users/owner/Downloads/midnightreport_july24_2013%20(2).pdf

Mielke, P. G. (2012). Investigating a systematic process to develop teacher expertise: A comparative case study. *Doctoral dissertation*. Retrieved from: http://www.marzanocenter.com/files/Paul_Mielke_Dissertation.pdf

Milanowski, A., Kimball, S. M., & White, B. (2004). *The relationship between standards-based teacher evaluation scores and student achievement.* Madison, WI: University of Wisconsin-Madison, Consortium for Policy Research in Education.

Mulford, B. (2003). School leaders: Changing roles and impact on teacher and school effectiveness. Paper commissioned by the Education and Training Policy Division (OECD). Retrieved from: https://www.oecd.org/edu/school/2635399.pdf

National Board of Professional Teaching Standards (2012). General Portfolio Instructions. Retrieved

from: http://www.nbpts.org/sites/default/files/documents/certificates/nbpts-certificate-general-portfolio-instructions.pdf

National Council on Teacher Quality. (2010). *Human capital in Boston Public Schools: Rethinking how to Attract, Develop and Retain Effective Teachers. Report*. Retrieved from: http://files.eric.ed.gov/fulltext/ED512671.pdf

National Educational Association. (2010). Teacher assessment and evaluation: The National Education Association's framework for transforming education systems to support effective teaching and improve student learning. White Paper. Retrieved from: http://www.nea.org/assets/docs/HE/TeachrAssmntWhtPaperTransform10_2.pdf

National Educational Association. (2011). *Promoting and implementing the National Education Association Policy Statement on Teacher Evaluation and Accountability: An NEA toolkit*. Retrieved from: http://www.nea.org/assets/docs/2011NEA_Teacher_Eval_Toolkit.pdf

OECD. (2013). Teachers for the 21st Century: Using Evaluation to Improve Teaching. OECD Publishing.

Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163– 193.

Papay, J. (2012). Refocusing the debate: Assessing the purpose and tools of teacher evaluation. *Harvard Educational Review*, *82*(1), 123-141

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2006). *Classroom assessment scoring system: Observation protocol manual.* Baltimore, MD: Brookes.

Piburn, M., & Sawada, D. (2000). Reformed teaching observation protocol: Reference manual (Technical Report No. IN00-3). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers. Retrieved from:

http://www.public.asu.edu/~anton1/AssessArticles/Assessments/Chemistry%20Assessments/RTOP%20Reference%20Manual.pdf

Popham, J. (2013). On Serving Two Masters: Formative and Summative Teacher Evaluation. *Principal Leadership*, 13(7), 8-22

Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy*, *4*(4), 492-519.

Reform Support Network. Targeting Growth Using Student Learning Objectives as a Measure of Educator Effectiveness. 8 pages, 2015. Retrieved from: http://education-first.com/wpcontent/uploads/2015/10/RSN-Targeting-Growth.pdf

Roeber, E. (2011). Educator evaluation: Models, parameters, issues and implementation. White Paper Commissioned by the Michigan Education Association. Retrieved from: http://www.michiganassessmentconsortium.org/sites/default/files/MAC-Whitepaper-Roeber-Evaluation-Models.pdf

Rowan, B. & Raudenbush, S.W. (2016). Teacher evaluation in American schools. In D.H. Gitomer & C.A. Bell (Eds.), Handbook of research on teaching (fifth ed.), (pp. 1159-1216). Washington, D.C., American Education Research Association.

Rowan, B., Schilling, S., Spain, A., Bhandari, P., Berger, D., & Graves, J. (2013) *Promoting high quality teacher evaluations in Michigan: Lessons from a pilot of educator effectiveness tools.* Ann Arbor: University of Michigan.

Saphier, J., Haley-Speca, M.A., & Gower, R. (2008). The Skillful teacher: Building Your Teaching Skills. 6th Edition, Research for Better Teaching, Inc., Acton, 5.

Sechrest, L. (2005). Validity of Measures Is No Simple Matter. *Health Services Research*, *40*(5 Pt 2), 1584–1604.

Silver, H. F., Perini, M. J., & Boutz, A. L. (2016). Tools for a successful school year (starting on day one): Classroom-ready techniques for building the four cornerstones of an effective classroom. Franklin Lakes, NJ: Thoughtful Education Press.

Silver Strong and Associates (2014). *The thoughtful classroom: Teacher Effectiveness Framework Basic Rubric. Toolkit.* Thoughtful Education Press. Retrieved from: https://www.thoughtfulclassroom.com/PDFs/TCTEF/TCTEF_Basic_Rubric.pdf

Slotnik, W., & Smith, M.D. (2008). Tying earning to learning (CTAC Issue Series). Retrieved from: http://www.ctacusa.com/PDFs/Rpt-TyingEarningtoLearning-2008.pdf

Taylor, E S., & Tyler, J. H (2012). The effect of evaluation on teacher performance. *American Economic Review,* 102 (7): 3628-3651.

Toch, T., and Rothman, R. (2008). Rush to judgment: teacher evaluation in public education. Education Sector Reports. Washington, DC: Education Sector, Retrieved from: www.educationsector.org/sites/default/files/ publications/RushToJudgment_ES_Jan08.pdf.

Tyler, J. H. (2011). *Designing high quality evaluation systems for high school teachers: Challenges and potential solutions*. Center for American Progress. Retrieved from: http://www.americanprogress.org/issues/2011/11/high_school_evaluation.html

University of Washington Center for Educational Leadership (2002). 5 Dimensions of teaching and Learning™ Instructional Framework, Version 4.0. Retrieved from: http://info.k-12leadership.org/5-dimensions-of-teaching-and-learning?_ga=2.137110067.220546004.1497545197-558680653.1497545197

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. Brooklyn, NY: The New Teacher Project. Retrieved from: http://widgeteffect.org/downloads/TheWidgetEffect.pdf

Whitehurst, R., Chingos, M., & Lindquist, K.M. (2014) Evaluating Teachers with Classroom Observations Lessons Learned in Four Districts. Brown Center on Education Policy. Retrieved from: https://www.brookings.edu/wp-content/uploads/2016/06/Evaluating-Teachers-with-Classroom-Observations.pdf

Wiggins, G. & McTighe, J. (2005). Understanding by Design. Expanded 2nd Edition. Upper Saddle River, NJ/Alexandria, VA: Pearson Education/Association for Supervision & Curriculum Development.

Worrell, F. C., & Kuterbach, L. D. (2001). The use of student ratings of teacher behaviors with academically talented high school students. Journal of Secondary Gifted Education, 12(4), 236–247. Retrieved from: http://eric.ed.gov/?id=EJ632563http://eric.ed.gov/?id=EJ632563

Worrell, F.C., & Kuterbach. L.D., (2001). The use of student ratings of teacher behaviors with academically talented high school students. *Journal of Secondary Gifted Education*, 14(4), 236-247.

Yon. M., Burnap, C., & Kohut, G. (2002). Evidence of effective teaching perceptions of peer reviewers. *College Teaching, 50*(3), 104-110.

Document Review

A number of documents, webinar and other artifacts were reviewed to give background information for this report.

A Summary of Documents Reviewed

Introduction: The data collection and methods section of the Proposal for Educator Evaluation Research and Evaluation Activities submitted by Ray Taylor and Associates, LLC would consist of reviewing existing documents from The Michigan Department of Education (MDE) and the Center for Educational Performance and Information (CEPI), among other resources. The following documents have been reviewed and summarized starting with documents from year 2011 through the spring of 2017.

The Education Alliance of Michigan, Developing an Educator Evaluation System: Improving Educator and Student Performance Guidelines for School Districts and Unions, 2011.

Legislation in January of 2011 required annual performance evaluation for all educators. Representatives of professional associations collaborated on developing a Framework for Michigan Educator Evaluations. Although widely supported, school districts and teacher unions felt a need to have more for their districts. The response was a report to address this need. This report was prepared with funds from a grant by the Federal Mediation and Conciliation Service. As a collaborative effort by stakeholders in the State Of Michigan, a document was developed that provided guidelines and recommendations for school systems to develop effective evaluation systems. The report also contains attachments, district models and frequently asked questions.

The Education Alliance of Michigan, Effective Evaluation, 2011.

This group of organizations in Michigan, AFT Michigan, Michigan Association of Secondary School Principals, Middle Cities Education Association and Michigan Association of Public School Academies, just to name a few, provided recommendations on effective evaluation systems in the State of Michigan. Some areas covered included Measuring Growth, Incorporating a Standards Based Evaluation Process, Collaborative Peer Review or review panels, Teacher/Administrator Self Assessments and Portfolios.

National Education Association, Promoting and Implementing the National Association Policy Statement On Teacher Evaluation and Accountability: An NEA Toolkit, 2011.

The National Education Association's 2011 Representative Assembly adopted a new Policy Statement on Teacher Evaluation and Accountability. This policy statement was written by and for teachers and takes into account the evidenced based research on teacher evaluation. The policy statement offers guidelines for developing an evaluation and accountability system that enhances the practice of teaching rather than simply identifying teachers for dismissal; emphasizes high standards and calls for robust evaluations that are based on multiple indicators, including indicators related to student learning and growth. An electronic tool kit was created by NEA leaders and staff to help members understand four key issues: teacher evaluation, peer assistance, peer assistance and review, and fair dismissal.

Michigan Council for Educator Effectiveness, October 2012

In response to Public Act 102 of 2011, The Advisory Committee to MCEE met to make recommendations about an administrative tool. In July of 2011, the committee initially submitted recommendations to MCEE that covered critical factors and suggested elements found in an administrator evaluation. Some but not all areas covered was reliance on data, incorporating get technology to assist with data collecting and managing data, nationally agreed upon dimensions of professional practice, target behaviors in a graduated approach appropriately applied to first year and veteran administrators and utilization of multiple indicators(observations, portfolios, artifacts, etc.) to identify progress.

Along with critical factors and suggested elements, the committee identified six challenges. Some but not all were fidelity for administrators, quality training of evaluators and including a student gowth model that aligned to local and state value added measures. Additionally, the committee vetted evaluation tools, national reports and participated reviewing research methodologies for research for research and development of evaluation tools.

After screening and reviewing various models, the committee recommended the following evaluation tools for piloting based on selected criteria. They were Marzano's School Leadership Evaluation Model, Masa and Michigan ASCD's School Advance Administrative Evaluation System-Principal Evaluation Instrument and District Leadership Evaluation Instruments, and New Leader For New Schools' Principal Leadership Actions

Michigan Council for Educator Effectiveness, Building an Improvement-Focused System of Educator Evaluation in Michigan: Final Recommendations, 2013.

The Michigan Council for Educator Effectiveness was established in June 2011 as part of the reform efforts of PA 102 of 2011. The Council was formed by appointments of the Governor and representatives from the Michigan State Legislature in 2011. The Superintendent of Public Instruction had a non-voting representative on the council. Funding was appropriated in December of 2011. The Council had a two year existence.

The work of the Council was realized utilizing two processes: a pilot study of educator Effectiveness tools to provide data on implementation and validity and crucial feedback for educators and meetings, consultations and research. The processes resulted in recommendations in these areas: A state evaluation tool for teachers, A student growth and assessment tool, A state evaluation tool for school administrators, Changes to the requirements for a professional teaching certificate, and a process for evaluating and approving local evaluation tools for teachers and administrators that are consistent with the state evaluation tool for teachers and administrators and the act.

Michigan Department of Educator Evaluator Landscape Day 2, October 2014, ASA, Michigan Association Of School Administrators, Michigan ASCD, and School Advance, Educator Evaluations: The Current Landscape, Your Role, and Best Practices; Designing an Educator Evaluation System For Learning, Growth, and Adaptation. A workshop for primarily administrators that covered topics under three parts; Growing your district performance evaluation system, Aligning administrator evaluations and Managing politics.

Michigan Department of Education, Educator Evaluation and Effectiveness in Michigan: An analysis of 2013-2014 educator evaluation systems and survey and educator effectiveness data, 2014.

Public Act No 102 0f 2011 provides for statewide system of educator evaluation for teachers and Administrators at traditional public schools and public school academies. The legislation allowed districts to use their formulas to rate educators as ineffective, minimally effective and highly effective.

A significant part of the evaluation was based on student growth from national, state and local assessments and other objective criteria. Twenty five percent of professional evaluations had to be based on student growth and assessment.

The data showed variation in ratings observation tools used and measures used in year end evaluations; percent of teachers rated effective or highly effective, uniformity of relationships between outcomes and percent of evaluations on growth data, percent of districts using evaluation for targeted professional development, coaching and support for teachers and administrators, superintendents, percent of districts using student growth in administrator and teachers final ratings, percent of districts using local common assessments, Michigan Educational Assessment Program ,

(MEAP), and the Michigan Merit Exam (MME). Information was also provided for teachers and administrators in public schools academy (PSA) schools and public school academies unique education providers.

Michigan Department of Education, MDE Educator Effectiveness Snap Shot, 2015.

This chart shows 2015-16 data by school district numbers and percent of teachers rated highly effective, effective or more, minimally effective and ineffective. This data may be crossed matched with other school district data.

Michigan Association of School Administrators, Michigan ASCD, School ADvance, Designing an Educator Evaluation System, 2014.

This was a session from a conference adapted from the work of Dr. Ed Roeber on educator evaluation. The session agenda covered the following topics: Legislation and MCEE Updates, Designing an Evaluation System (Perspectives and Assumptions, Six Research Supported Principles, Designing and Growing the System, Administrator Evaluation-School Advance), Tools for Evaluation, Words and Reflection and Implementation (more protocols). This session also covered the origin of MCEE, its' origin and individuals comprising the make-up of MCEE. Time was also dedicated to examining the inspection and demonstration models in the context of the school district as a system, teachers' role in instruction, issues around portfolios, and administrator training in evaluation.

Michigan Department of Education, Michigan Educator Evaluations At-A-Glance, 2015.

This is a document covering Public Act 173. It provides information for educators on the legislation behind the law, the need for the law to be implemented with fidelity, and highlights three sections, each with a different purpose. The first section emphasizes some of the purposes for efforts to improve educator evaluations. The second section is an overview of the law and some of the factors at the local state, and federal levels that contributed to the content of legislation. The third section provides a more concise and detailed analysis of the law.

Michigan Department of Education, Michigan Educator Evaluations Frequently Asked Questions, 2015.

A document that provides answers to forty one questions covering the following topics: Observation Tools, Observations, Training, Student Growth and Assessment Data, Appeals, Public Reporting, Teacher Certification and Roles and Responsibilities. Some view this document as a companion to Evaluations At-A-Glance document.

Educator Effectiveness Factors and Ratings for Teacher Evaluations per School Districts in the State of Michigan for School Year, 2014-2015.

Teacher Evaluation Factors included were Absenteeism from the Job, Classroom Management, Content Knowledge, Instructional Practices, Pedagogical Knowledge & Practice, Professional Development, Professional Responsibilities, Student Growth Measures, Student Achievement Data, Student Learning Objectives, Portfolio and/or Peer Reviews, Self-Assessments and Surveys.

Michigan Department of Education, Michigan Association of School Administrators, Mid-Winter Conference, School Growth For Education Evaluations in Michigan, 2017.

This Workshop was conducted for public school k-12 personnel affected by the current evaluation in the State of Michigan. Topic covered were the Legislation for Student Growth Within Educator Evaluations, Student Growth Models used in Michigan and other states, Student Growth Guidance, Communication, and Exemplars, Webinar Series, SLO (Student Learning Objectives)

and SGP (Student Growth Percentiles), Calculator Tools-Beta Version.

Michigan Department of Education, Student Growth Percentiles (SGPs), 2015.

Student Growth Percentiles is a way to quantify student learning from one year to the next. In The State of Michigan students will get student growth percentiles for each subject in which they Tested and there is one previous Michigan test for that student. Student Growth Percentiles will compare achievement of students to those of their peers.

Michigan Department of Education, Student Learning Objectives (SLOs), 2015.

Student Learning Objectives are another way of measuring student growth. The State of Michigan consulted with other states in implementing student learning objectives. The State of Michigan utilizes a Student Learning Objective Guidance Document, Student Learning Objectives FAQ, a Student Learning Objectives Template and A Student Learning Objectives Template.

Wayne RESA, Measuring Student Growth: A Practical Guide to Educator Evaluation, 2016.

This is a guidance document that provides school districts with more than one method of measuring student growth when conducting evaluations. Professionals from Wayne County School Districts met from the winter of 2015 to July of 2016 in developing this document. The research framework was taken from works of Stiggins, Popham, and Darling Hammond. The work of the Michigan Council for Educator Effectiveness, The Widget Effect and Standard Setting by Cizek and Bunch was also referenced in preparing this document.

Numerous Tables and Figures are included in the document along with six sections which are the following: Growth Models, Developing and Selecting Assessments of Student Growth for Use in Teacher Evaluation Systems, Measuring Student Growth: A Step by Step Process to Analyzing Data, Standard Setting For Student Growth, Student Learning Objectives: A Measure of Educator Effectiveness and Formative Assessment.

Michigan Association of School Boards, Public Comments on Proposed Rules 2016-31 ED, by Don Wotruba, Executive Director, MASA, 2016

This communication to the Office of Education Talent, Michigan Department of Education, identified three areas of concern on proposed rule 2016-31 ED. Accompanying recommendations for the rules are also included in the communication.

Proposed Rule 380.21- Definition of Efficacy focuses on differentiation of assessment rather than areas of improvement in practice and developing the individual being evaluated. The recommendation was that MDE follow the accepted definition of efficacy as “the capacity to produce a desired result: effectiveness”.

Proposed Rules 380.21, 380.22 and 380.23 regarding the creation of a scoring guide and minimum standards. Neither was included in the rules. The recommendation was for MDE to rewrite the rules detailing that evaluation tools will be scored for inclusion. MASB would also ask for a second public hearing.

The last concern was with proposed Rule 380.22- Requirements contained in on-line applications be included on the MDE list. Requirements to have evidence of reliability, validity, and efficiency creates a disadvantage for MASB due to the language of the rule. The recommendation was that MDE revise the language to include evidence or a plan to provide evidence.

Michigan Department of Education, Education Evaluation Overview Webinar Deck from MDE, February 2017.

Provides an in depth and concise presentation of the evaluation model for teachers and administrators in the State of Michigan. Covers all aspects of the evaluation model including the legislative, the the operational requirements as well as the focus and impact on student growth requirements and and recommendations.

Michigan Educator Evaluation Tool by District, March 2017

Alphabetically list by region and, district, the type of evaluation model being implemented in the State of Michigan. The models are Danielson, 5 Dimensions of Teaching and Learning, Marzano, Thoughtful Classroom, or an adopted, modified non-approved, or district developed tool.

Michigan Department of Education, Office of Education Talent, Educator Webinar, 2017.

The Office of Education Talent presented a series of webinars on the topics of educator evaluations and student growth. Links were provided for each webinar beginning with an Educator Evaluation Overview in February. The other topics covered were; What is Student Growth?, From Compliance to Excellence, Looking At Student Learning Objectives, Combining Student Growth and Performance and Data, and an Open Q&A Session.

Michigan Department of Education, Office of Education Talent, Michigan Association of School Administrators, Mid Winter Conference, Student Evaluations in Michigan, 2017.

Conference participants were given an explanation of Student Growth Models in Michigan, including Student Growth Percentiles (SGPs) and Student Learning Objectives (SLOs).

Michigan Department of Education, Office of Education Talent, Education Webinar, 2017.

This webinar on Student Learning Objectives was presented for educators interested in utilizing Student Learning Objectives (SLOs) in student achievement and in instructional methodology.

Researcher White Papers

Ray.Taylor and Associates commissioned two white papers from expert academic researchers. Each paper reflects the expertise and findings of the author.

Implementing Michigan's Educator Evaluation System – by Robert Floden, PhD

K-12 Educator Evaluation: Lessons Learned in a Decade of Policy Implementation and Research – by Suzanne Wilson, PhD

September 30, 2017

Implementing Michigan's Educator Evaluation System

Robert E. Floden
Michigan State University

April 3, 2017

The Michigan legislation mandating that districts adopt educator evaluation systems specifies that the systems should be rigorous, transparent, and fair. Those characteristics should be used in identifying best practices, as should the implicit goal that the purpose of the evaluation system is to improve the education of the district's pupils.

The evaluation system should contribute to the improvement of education in two ways. First, it should inform personnel decisions, so that educators with the most positive evaluations will be retained and rewarded, and those with the worst evaluations will be moved out of the system. Through these personnel actions the best performing educators will be encouraged to continue working in the district and the lowest performers will leave, resulting in a workforce that is better in the aggregate. These personnel actions will, however, affect a small fraction of the educators, with little effect on the quality of educators in most schools.

Second, it should inform the actions of all those in the system regarding efforts to improve the performance of all the educators working in the district. The evaluation system should provide feedback to those evaluated about what they are doing well and where they need to improve, giving them specific ideas about how they could improve. The evaluation system should also provide information that can be used to plan for professional development. For example, a building principal should be able to use information from the evaluations of the school staff to decide what professional development to offer for the school as a whole, or for individual teachers. Or a superintendent might use information from the evaluation of building principals to decide what professional development would be appropriate for them.

For both of these mechanisms of improvement, it is important that the system be rigorous, transparent, and fair. It should be rigorous in the sense that the conclusions drawn about those evaluated are valid. That is, if an educator is classified as ineffective, the educator should indeed be ineffective. Otherwise the personnel decisions will not result in improvements in the workforce. Similarly, the conclusions drawn about specific needs for professional development should be accurate. The system should be transparent, so that both those being evaluated and those reading the evaluation understand what evidence led to the evaluation. Clarity about the basis for the evaluation makes it possible for the educator to appropriately address weaknesses and build on strengths. It also makes it possible for supervisors to select professional development tailored to the strengths and needs of the educator evaluated. Finally, fairness is important for those involved in the evaluation to buy in to the system, to see that it is aimed at improving teaching and learning, and to be motivated to invest in actions driven by the evaluation.

Best Practices

These mechanisms of action and general principles suggest several practices. First, the systems used for observations of practice should have a substantial base of research supporting links to desired outcomes. A basis in research contributes to both rigor and fairness. Rigor comes from evidence that observation results have either been shown to lead to desired outcomes, or are based in professional expertise. By using an established observation protocol, the system avoids relying on idiosyncratic views of individual observers.

Several observation systems have substantial bases of support, including those recommended by the Michigan Department of Education – Charlotte Danielson's Framework for Teaching, the Marzano Teacher Evaluation Model, the Thoughtful Classroom, and the 5 Dimensions of Teaching and Learning. The resources required to develop and validate a new system are substantial, likely more that local districts would be able to invest.

Second, those conducting the observations should be trained in use of whatever system is selected, and their ability to use the system appropriately should be periodically checked. Appropriate use of an observation system is important for rigor, transparency, and fairness. Over time, it is common for observers to "drift" away from the observation system as designed, relying more on their individual, and perhaps idiosyncratic, judgments about educator quality, rather than sticking to the system's rubrics. In my own experience in developing protocols for classroom observation, I have seen how difficult it can be for observers to maintain fidelity to a rubric, because they come to the observation with their own judgments about teaching quality and wish to give scores that match their own beliefs, even if those beliefs are not in line with the rubric they have agreed to follow. If the ratings from the observations are based on individual judgments, rigor is reduced, because the ratings no longer have support from research to connect ratings to pupil learning. Transparency also suffers, because the educator being observed is not aware of the observer's personal criteria for quality. If the educator observed recognizes the discrepancy between the scores given and the officially adopted observation system, he or she will justifiably believe that the system as used is unfair.

Third, the educators being evaluated should learn about the observation system, either through being trained to use it or at a minimum having the opportunity to review materials describing the specifics of the system. Knowledge of what aspects of practice are being evaluated, and the rubrics used, is a key component of system transparency. Knowing what the observation system defines as good practice will help educators understand strengths and weaknesses in their practice, allowing them to work toward improvement.

Fourth, educators should be given detailed constructive feedback based on the observation, shortly after the observation takes place. For educators to take advantage of the observation to make improvements to their practice, they need to know details about their strengths and weakness. Making connections between written comments and events in the classroom is much easier if little time has elapsed between the observation and the written evaluation.

Fifth, educators should be engaged in decisions made about the evaluation system. Engagement allows those being evaluated, and those using the evaluation information, to make recommendations aimed at achieving rigor, transparency, and fairness. Engagement will increase the perception that the evaluation is tapping into key features of instructional and leadership quality. Engagement should include participation in decisions about how scores on the observation, student performance, and other measures will be used to decide placement in one of the performance categories.

Sixth, administrators should get professional development on how to use the evaluation results to make personnel decisions and to work with their teachers to improve performance. In the absence of such professional development, administrators may not have skills needed to use the evaluation results to improve the levels of teaching and learning in their schools.

Finally, to 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 evaluations are extremely low, for most educators, the student test data, with their statistical adjustments, are difficult to understand and give only crude indications 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 learning data, at least until they have taken steps to get teachers to improve, focusing on areas of weakness revealed through observations.

Barriers to Implementation

As with many other new education initiatives a substantial obstacle to effective implementation is the time and effort required. Four requirements that will require substantial time and effort are 1) the training in use of observation systems, 2) the conduct of the observations (including follow-up meetings to discuss what was observed), 3) the establishment of student growth measures for teachers where state assessment data are not available, and 4) the computation of adjustments to student growth measures to take account of entering characteristics of students in a teacher's class. Some formula funds will be provided to ISDs to provide training in the observation instruments and MDE will provide districts with some growth data, but those resources may not be sufficient. Resources will also be needed for periodic checks on the fidelity of use of the system, including the inter-rater agreement on observation scores. The time and effort educators will need to conduct the observations and develop additional growth measures will need to be diverted from their current instructional and administrative work. That poses a major challenge for implementation.

Securing and maintaining public support for the evaluation system may also be a barrier to implementation. The apparent logic of the system will likely be appealing: Districts should take seriously the evaluation of school employees and base personnel decisions and professional development on those evaluations. But as the new systems are put in place, members of the public may find the specific workings of the system difficult to understand. This is particularly likely for the statistical adjustments used with the student learning data. This lack of public understanding may impede use of the evaluation system if, for example, a teacher who is well liked by parents has below-average student learning results. If parents' views of an educator conflict with what comes from the evaluation system, the system loses credibility. Lack of transparency can lead to perceptions of lack of fairness.

The effective implementation of an evaluation system depends in large part on buy-in from those being evaluated. This is particularly the case for decisions about professional development that are based on evaluation results. For professional development to lead to improvements in education practice, those participating in the professional development must be motivated to learn from the experience and, even more important, must be motivated to put what they are learning into practice. If educators do not see the system as legitimately focused on improvement, they may not put in the effort needed to make improvements.

Measurement of Student Growth Using Alternative Tools and Processes

For educators in non-tested grades and subjects, district systems will require the selection or construction of alternative measures of student growth. For some subject areas and grades, this may be done by using commercially available assessments, such as the MAP offered by NWEA. Decisions about what tests to use should be made with great care, to ensure that the test is a reliable measure, and it is able to support valid conclusions about educator effectiveness. The MAP is a case in point. The assessment was developed to be used as an interim assessment to support inferences about the performance of individual students, so that

educators can adjust instruction in light of current pupil performance. NWEA tailors its item pool to correspond with topics on a given state's curriculum, but it is probably not as closely aligned with new curriculum standards as MSTEP. NWEA assessment results are also measures of student achievement, rather than growth, so if this measure were used in the evaluation system, fairness would require some adjustment for differences in student characteristics.

For some subject areas, such as art, music, and social studies, districts may not find acceptable commercially available assessments. For these subjects they may need to work with educators in these subjects to develop student learning objectives. As these objectives are developed, it will be important to include ambitious goals, in addition to goals that may be important, but are simply restatements of what has typically been achieved.

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. Wherever possible, the measures of student growth should be consistent across grade levels, and over time, so that proper consideration can be given to characteristics of students at the beginning of the school year.

Concluding Remarks

One issue that deserves more attention is how districts will use the results of the evaluation system to assign educators to the four categories specified in the legislation: highly effective, effective, minimally effective, and ineffective. Once each educator has scores from the selected observation system and the designated measure of student growth, those scores will be used to determine the category to which that educator will be assigned.

The legislation dictates the weight that must be given to the student growth measure, but does not otherwise specify how to use the scores to make assignments to categories of quality. Districts will need to use a standard setting process for establishing the cut points, which will in turn determine how many educators will fall into each category. The criteria of rigor, transparency, and fairness should be used in this standard setting procedure as well.

Achieving rigor will be a challenge, because the research underlying observations and growth measures supports the overall association with quality (e.g., higher scores on the observation measure is associated with higher quality), but does not establish break points. That is, research does not, and at present probably cannot, support the conclusion that an educator with an overall score of 85 (on some composite measure) is effective, while an educator with an overall score of 84 is minimally effective. The rigor to look for at present is that of using a standard setting process that is endorsed by the assessment community.

Transparency can be secured by making clear how the standards are set. Just as standards for other assessments are established by a systematic process of having representatives of knowledgeable and interested groups meet to consider various possible cut points, the process for setting standards assigning educators to categories should be done through a process that is clearly described.

To promote fairness of the process of assignment to categories, educators should be among the groups involved in setting these standards for assignment to category. Including educators in this process will

help to make those being evaluated to be confident that the evaluation system is aimed at the goal of improving the quality of education for all students.

An open question is what proportion of educators is expected to fall into each of the four categories. Based on the results of previous teacher evaluation systems, it may be that the vast majority of educators will be rated highly effective or effective, particularly for those with at least four or five years of experience. Some small percentage, perhaps 1-3% will be rated ineffective. It may also be the case that the distribution across categories is roughly similar across school districts. If that occurs, some commentators will conclude that the systems established by districts are not sufficiently rigorous, in light of the overall low performance of Michigan students on both NAEP and the state assessments. Others will argue that, if systems take account of the social, medical, and economic contexts of schools, as fairness dictates, it is reasonable to think that most teachers should be judged effective or highly effective, even though their students' achievement scores are low.

As noted at the beginning of the paper, the Michigan legislation calls for districts to adopt educator evaluation systems that are rigorous, transparent, and fair. Systems that meet those standards are more likely to achieve the goal of improving educator quality, because they will be based on research linking the system to student learning, will offer actionable guidance about current strengths and weaknesses, and will increase educators' motivation to put in the effort needed to improve practice.

K-12 Educator Evaluation:
Lessons Learned in a Decade of Policy Implementation and Research

Suzanne M. Wilson[2]
5 April 2017

In the last 10 years, interest in creating educator evaluation systems that help to both improve teaching and to identify under-performing teachers has risen dramatically. In some ways, the U.S. has been witnessing a natural experiment as state after state has implemented its own version of educator accountability, stimulated largely by Race to the Top which encouraged states to adopt ambitious teacher and principal evaluation systems. The purpose of this paper is to synthesize what we have learned in from the last 10 years of both policy implementation and education research concerning best practices for implementation and the challenges faced by states and districts as they installed educator evaluation systems nationwide.[3]

WHY DO WE NEED EDUCATOR EVALUATION?

The call for increased accountability for teachers was catalyzed by several forces. Perhaps centrally, holding districts, schools, and then teachers accountable for measurable metrics grew out of a general movement toward systemic and standards-based reform that arose in the 1980s, which was then accelerated by calls for choice in the educational marketplace. Research in the 1990s and early 2000s began to demonstrate that teachers are critically important to student development and achievement. From this perspective, one needs teacher evaluation so as to identify under-performing teachers.

At the same time, there was increased understanding of the central role of professional learning in teachers' on-going development. This need for support is especially acute in the early years of teaching, when one is moving from preparation to practice. As a result, considerable investment was made in strengthening teacher preparation, and providing appropriate early career (sometimes called induction) support. Given new developments in instructional methods, the explosion of knowledge in academic fields, new academic standards for students, innovative forms of assessment, new technologies, and changing demographics of the

[2] Dr. Suzanne M. Wilson is a Neag Endowed Professor of Teacher Education at the University of Connecticut where she currently serves as Professor in the Department of Curriculum and Instruction. Her undergraduate degree is in history and American Studies from Brown University; she also has a M.S. in Statistics and a Ph.D. in Psychological Studies in Education from Stanford University. She was a University Distinguished Professor in the Department of Teacher Education at Michigan State University, where she served on the faculty for 26 years. Wilson also served as the first director of the Teacher Assessment Project (PI, Lee Shulman), which developed prototype assessments for the National Board for Professional Teaching Standards. Her current work concerns exploring various measures of teaching and teachers' understanding that might be used for teacher education and education research, as well as a study of the contemporary and jurisdictional battles over who should control teacher education and licensure. She has published in *American Educator*, *American Educational Research Journal*, *Educational Researcher*, *Elementary School Journal*, *Journal of Teacher Education*, *Phi Delta Kappan, and Teaching Education*. She is author of *California Dreaming: Reforming Mathematics Education* (Yale, 2003), and editor of Lee Shulman's collection of essays, *Wisdom of practice: Essays on teaching, learning, and learning to teach* (Jossey-Bass, 2004). Wilson is a fellow of the American Educational Research Association, as well as a member of the National Research Council's Board on Science Education and the National Academy of Education.

[3] Because there is much more empirical evidence on teacher evaluation systems, the paper focuses largely on lessons learned concerning teacher evaluation systems. Although many of the insights apply equally to principal evaluation systems, this paper does not make assertions about that aspect of educator evaluation.

U.S. student population, there was also increased awareness of the need to provide teachers with focused, targeted professional development that would enhance their capacity to teach to new standards using new pedagogies to reach all students. From this perspective, one needs educator evaluation to support all teachers in their on-going learning, no matter where they are in their career.

In sum, the increased emphasis on educator evaluation can be understood as being driven both by (1) concerns for identifying teachers who need remediation or need to be released from the workforce and (2) an interest in identifying important areas of growth for all teachers to sustain their development and excellence. *One lesson learned in the last 10 years is that competing ends/goals can create a divide in how various stakeholders perceive the purpose of an evaluation system. Being explicit and clear about purposes and goals is essential to the creation and implementation of a system that multiple stakeholders trust.*

Consider the reactions to 2014-15 teacher evaluations in Michigan, which reported the vast majority of the state's teachers as "effective" or "highly effective" (56% and 42% respectively), with only 463 teachers out of 94,972 as ineffective and the remaining 1900+ as minimally effective. Moore (2015) reported that, since the passage of the teacher evaluation reforms in 2011, only 19 of the state's nearly 96,000 teachers met the official threshold for being terminated due to poor evaluations. Concerns were raised statewide that the high ratings seemed "unlikely," "unrealistic," and an inaccurate portrait of the Michigan teacher workforce competencies (Moore, 2015). 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 the 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 "weaknesses." 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.

A VISION OF EFFECTIVE TEACHING

Evaluating quality requires making explicit one's assumptions. For teaching, there are two anchoring visions: a vision of what students learn in school and a vision of what teachers do to enable that learning. Because education is under state jurisdiction in the U.S., these visions are worked out in state student and teacher standards, mandates, and values. In general, the goals of U.S. public education include children learning the academic content they encounter across their preK-12 years. Many people also want them to have high attendance, to be promoted to the next grade, to graduate in a timely fashion, to be safe, to be respected, and to get along with others. Recently, there has also been increased attention to issues concerning students' grit, and their social and emotional well-being.

Given this broad vision of what students should learn in school, policymakers and scholars have argued for clearly explicated visions of teacher effectiveness that go beyond teachers' students having high test scores. For example, Goe, Bell, and Little (2008) recommend a "five point" comprehensive definition that includes the following:

- Effective teachers have high expectations for all students and help all students learn, as measured by value-added or other test-based measures, or by alternative measures;
- Effective teachers contribute to positive academic, attitudinal, and social outcomes for students such as regular attendance, on-time promotion to the next grade, on-time graduation, self-efficacy, and cooperative behavior;
- Effective teachers use diverse resources to plan and structure engaging learning opportunities; monitor student progress formatively, adapting instruction as needed; and evaluate learning using multiple sources of evidence;

- Effective teachers contribute to the development of classrooms and schools that value diversity and civic-mindedness; and
- Effective teachers collaborate with other teachers, administrators, parents, and education professionals to ensure student success, particularly the success of students with special needs and those at high risk for failure. (p. 8)

An ambitious and comprehensive view of effective teaching anchors high quality educator evaluation systems. And while this is by no means the only example of a set of expectations about teachers (see, for example, the State of Michigan's Standards of Professional Learning for Educators), it is worth noting that most visions of effective teaching go well beyond student achievement in core academic subjects. But most educator evaluation systems place more emphasis on student achievement than on any other goals. *A second lesson learned in the previous 10 years is that disagreement or ambiguity around the expectations for effective teaching can lead to breakdowns in communication and implementation of a strong educator evaluation system.*

Michigan has a clear set of professional standards for teachers. However, while there is considerable overlap in the four observation tools for teachers that were state approved (Danielson's Framework for Teaching, Marzano's Teacher Evaluation Model, the Thoughtful Classroom, and the 5 Dimensions of Teaching and Learning); each is based on different conception of effective teaching. And districts were also allowed to use tools that are not on the state approved list, adding even more images of teaching and learning into the Michigan mix. Complicating matters still further, districts tailored, streamlined, and adapted observation tools, which leads to truncating or changing the conception of teaching unerlying the (adapted) tool. In essence, the state policy enabled ambiguity around expectations for effective teaching. This ambiguity can undermine the integrity of educator evaluation systems.

BEST PRACTICES IN EDUCATOR EVALUATION SYSTEM IMPLEMENTATION

When states began designing their own educator evaluation systems, they faced a series of questions, including:

- Who should develop the evaluation system and what is the relationship between the state and the districts?
- What measures (and how many) should be used?
- What should be specified about classroom observations? How many? Who should conduct them?
- How will the evaluation results be used? And how are they tied to employment?

Even though there is a growing body of research on educator evaluation systems, we lack a definitive evidentiary basis for answering these questions. That said, a great deal has been learned and I summarize those lessons here.

Who should develop the evaluation system and what is the relationship between the state and the districts?

Generally, states fall into three categories with regard to this question. Some states mandate an educator evaluation system; other states have a presumptive system but districts can elect to propose an alternative approach; and still others have district designed systems within state parameters. There is no research that contrasts these approaches in terms of their capacity to improve teaching and learning. Michigan opted for an approach that resembled the presumptive+alternative system since it offered districts choices in four state-adopted teacher observation protocols, two administrator protocols, and a graduated plan for the incorporation of student growth and assessment data, with considerable latitude for using "research-based" growth measures or alternative assessments, ranging from locally developed student learning objectives (SLOs) to nationally-normed assessments.

There are pros and cons for each approach. A single state level system allows for standardized data

collection and the comparison of districts, and an unambiguous single vision of teaching and learning. Such a system is perceived as fair because everyone is held accountable to the same mandate. But a single state system diminishes local flexibility and can prevent a sense of buy-in; further, local contextual issues are not acknowledged. Because there is variance in district resources, the system can be seen as unfair because lower-resourced districts do not have sufficient resources to meet all of the associated demands of the policy.

An elective state system allows for some flexibility, and data collection can still be (somewhat) standardized, so that many districts can be compared. This approach also allows for the continuance of local models that were already in place, so this approach poses challenges for state oversight and data aggregation can be compromised.

The district evaluation system increases local ownership and buy-in, allowing districts to address local priorities. This is especially relevant when thinking about issues of diversity, whether that diversity falls along lines of ethnicity, race, and culture, or along lines of rural and urban. No research has interrogated the implicit bias of these observation protocols, nor implicit bias associated with underlying visions of teaching and learning. Yet it is likely that the tools and visions underlying them are not universally representative of local values concerning teaching and learning. No instrument is color- or culture-blind. But the elective state system also makes it very difficult to compare progress across districts and to aggregate data. Reliability is vulnerable across districts and training to ensure fidelity requires more resources since training is done at the district level.[4]

Michigan's approach has been one of less central mandate and more local control. Keesler and Howe (2016) describe the difficulties of creating systematic supports for Michigan districts: Districts vary considerably in whether they have the capacity to work with metrics and evaluation systems in ways that ensure fidelity, and with little consistency across the state, it is nearly impossible to create supports for all schools or to work toward a consistent and comprehensive system.

What measures and how many should be used?

A third lesson learned is that using multiple measures provides a more accurate estimate of teacher effectiveness than reliance on a single measure. Goe and Holdheide (2011) note that multiple measures strengthen the evaluation, providing more complete evidence about teachers' practice.

When states and districts began responding to the mandates for educator evaluation systems, they encountered a busy marketplace of vendors and products. Among the available measures were tools for classroom observations, principal evaluations, the analysis of instructional artifacts, teacher portfolios, teacher surveys or interviews, student and parent surveys, and value added models that make use of student data from a range of different measures, including standardized tests and locally developed student learning outcomes.[5] For example, the American Institute for Research's Center for Great Teachers and Leaders has a website (http://www3.learningpt.org/tqsource/GEP/) that offers information on more than 75 educator evaluation tools that are currently used across the country, including information on any relevant research, the educators assessed (and their student populations), costs, contact information, and available technical support.

States use different measures and put them together in different ways. Thus, while most schools report that they have an educator evaluation system 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 observers, from a focus on struggling teachers to a focus on schoolwide continuous improvement.

[4] See Goe, Holdheide, and Miller (2014) for a complete analysis of these pros and cons.

[5] See Goe, Bell, and Little (2008) for an analysis of the evidentiary basis for these various measures.

Current state systems use a range of measures, including student data from standardized tests and other locally developed measures, classroom observations, teacher surveys, instructional artifacts, and parent and teacher surveys. For example, in New York the evaluation system includes (1) statewide student growth measures; (2) locally selected measures of student achievement; and (3) teacher observations, school visits and other ways of providing teachers with detailed and structured feedback. In Nevada, the system requires (1) statewide student achievement measures; (2) teacher evaluations that include assessments of classroom management skills, a review of lesson plans or grade books, an assessment of whether the teacher is meeting student needs, and an evaluation of whether the curriculum taught is aligned with the standards. And in Maryland, the system focuses on student growth and professional practice. The evaluation of professional practice requires assessing (1) planning and preparation, (2) instruction, (3) the classroom environment, and (4) professional responsibilities.[6]

Michigan stipulated that between 2015-2018, 25% of the annual teacher evaluation would be based on student growth and assessment data (which was then raised to 40% in 2018-19). Starting in 2018-19, 50% of student growth was to be measured using available state assessments. However, no state tests in some content domains or at some grade levels. This has led to considerable latitude concerning the measures of student growth used. Teacher performance was to also be measured with an observation tool. Moreover, the observation tools are generic and are not tailored to the specific needs of particular content areas or grade levels. In sum, the Michigan requirements were both streamlined and adaptable. That said, the ability of local districts to select or adapt measures has created considerable variation across districts, and the state is often left comparing apples and oranges. And the lack of grade level- and content-specific measures has raised questions about the validity of many measures.

No research sheds light on exactly how many measures are optimal or which ones are best. Goe, Bell, and Little (2008) propose six criteria for judging the quality of measurements, including: (1) comprehensiveness (to what extent does the measure capture all aspects of teacher effectiveness?); (2) generality (to what extent does the measure span the full range of contexts in which a teacher works?); (3) utility (how useful are the scores? And to what ends?); (4) practicality (what are the associated costs, training, and the overall use of resources implicated?); (5) reliability (how consistently do users of the measure come up with the same score/evaluation?); and (6) credibility (how believable is the measure to teachers, principals, and the public?).

The two measures that have received the most attention are student data (the focus of another paper in this series) and classroom observations. *The general consensus is that classroom observations are critical*, in part because they provide concrete data that could help teachers and their evaluators work on improvement, and in part because observations – when done by a trainer observer and using a thoughtfully designed protocol – have face validity with teachers, policymakers, and the general public. Despite the strong argument for face validity, however, several scholars have pointed out that comparatively little research has explored the problematic aspects of observational instruments, as they too have biases. For example, studies have found that observations demonstrate considerable variability by lesson, within lesson segments (e.g., the first 15 minutes of a class vs. the last 15 minutes of the same class), time of year, content of instruction, and raters (who contribute the largest source of error in generalizability studies of observation instruments).[7] Recall also the earlier point that we have very little information about how these tools might be biased along lines of race, ethnicity, socioeconomic status, and other differences that distinguish between schools and districts.

As noted earlier, the most popular observation protocols – Marzano, Danielson, CLASS, among them – differ in their underlying conceptualizations. For example, Danielson is organized around four domains:

[6] Various websites and reports attempt to synthesize the "state of the state systems," for example, see NCTQ (2011) or http://resource.tqsource.org/GEP/.
[7] See Cohen and Goldhaber (2016).

planning and preparation, classroom environment, instruction, and professional responsibilities. CLASS is organized in three: emotional support, classroom organization, and instructional support. Observational systems also different in terms of their guidance on how observations are orchestrated and how many are conducted. *In general, there is consensus that a high quality "evidence-based" evaluation consists of the following components: (1) multiple observations, (2) the use of rubrics, (3) multiple observers, (4) measures that distinguish variation in teachers' performance ratings, and (5) policies that link teacher effectiveness to student performance* (Sartain et al., p. 4).

Some research has investigated the validity of the instruments, that is, do teachers with higher scores on the observations have higher student achievement. Perhaps the best known of these studies is the Measures of Effective Teaching Study (MET, Kane et. Al, 2013), which found that teacher effectiveness is best identified by simultaneously using measures based on student achievement, classroom observations, and student surveys.[8] But no research has demonstrated that one observation system is superior to others, nor has research clarified whether announced or unannounced visits are best, and the optimal balanced of shorter and longer observations. That said, research has demonstrated that observational systems are only as good as the materials that accompany them, including protocols for when to observe, and how often, for how to conduct an observation and an interview, for how to document interview and observation data, and for how to score. Research has also demonstrated that the consistency and rigor with which these protocols and tools are used matters to the quality of the evidence produced. Research also shows that you need multiple raters to develop reliable systems, and that raters need consistent re-training.[9]

Other research suggests that the use of an evidence-based evaluation system can lead to teacher improvement. Taylor and Tyler (2012), for example, studied a sample of midcareer elementary and middle school teachers in Cincinnati who had participated in a highly-structured evaluation process between 2003-2010. Teachers who had been through the thorough evaluation (which used an adapted form of Danielson's Framework for Teaching) had student achievement that was significantly higher than students whose teachers had not been through the evaluation. The effects of the evaluation were higher for teachers who received lower teacher evaluation scores, and improvements were sustained after the year of the evaluation. However, other research has demonstrated that the kind of information teachers need in order to improve their practice is quite different than a "thumbs up" or "thumbs down" assessment that leads to a number on an effectiveness scale.[10]

In sum, emergent research suggests that, when used appropriately, measures of teachers' effectiveness can be valid, and can lead to improvement in teachers who initially receive lower evaluations. However, research also demonstrates that educator evaluation systems need to be consistently implemented and well resourced: Time and money are required for training, for providing sufficient numbers of observers to visit teachers' classrooms a sufficient number of times. And protocols need to be used at every step along the way, with an understanding by users that adaptations to both process and instruments can corrupt the tools' capacity to accurately measure teacher quality.

What should be specified about classroom observations? How many? Who should conduct them?

Given the centrality of classroom observations in the educator evaluation systems, a great deal of attention has been focused on questions like: How many observations should be conducted? Or what length? By whom? State and district policies vary considerably on the frequency and timing of the evaluations: Probationary teachers can have as many as 38 observations in a year and as few as one, ranging from 5

[8] See http://www.gatesfoundation.org/united-states/Pages/measures-of-effective-teaching-fact-sheet.aspx
[9] See multiple essays in Grissom and Youngs (2016) for details, especially Cohen and Goldhaber (2016) and Pianta and Hamre (2016).
[10] See Hill and Grossman (2013).

minutes to one hour/class period. Experienced teachers may be observed once a year, or three times every three years. Some districts/states have adopted and implemented (with fidelity) empirically-tested systems of support like the Teacher Advancement Program (TAP), which involves evaluations of 4-6 times a year by mentor teachers or principals who go through a 4-day training. After each observation, the evaluator and teacher meet and plan for on-going growth. Another program that enjoys some popularity is the Peer Assistance and Review (PAR), which includes peer reviewers (called consulting teachers) who provide intensive, individual help to a caseload of 15 teachers, most of whom are novices.

Research on TAP and PAR, as well as recent experiences of those implementing similar policies has illuminated several important issues.[11] *One lesson learned is that the quality of the data produced by the observational system depends on the quality of the system itself. Consensus and research suggests that one needs: (1) open and rigorous selection of the consulting teachers/observers; (2) explicit guidelines, both about instruction and teaching standards, and well developed rubrics; and (3) substantial training and on-going support for observers.*

Perhaps the most repeated message here is that *training matters*. Every producer of a widely-used observation protocol insists on training users; some school districts have heeded this advice and invested in strong training and (at times) re-training. For example, in their study of an evaluation system in Cincinnati, Sartain and her colleagues found that principals and teachers reported that, when using the Danielson Framework and protocol, their post observations were more reflective and objective, and that the conversation was more focused on instructional improvement. But positive attitudes were dependent on whether the principal understood the materials and instruction, and whether the principal was "on board" with using the protocol and system. This is a consistent message documented by researchers across the country.[12]

How will the evaluation results be used? And how are they tied to employment?

Yet another question concerns how the educator evaluation system results are used. Here too states vary considerably. A small number of studies have begun to explore the effects of different policies for teacher improvement and dismissal. For example, Dee and Wyckoff (2013) studied the District of Columbia's teacher evaluation system, IMPACT. In this system, teachers who were rated as highly effective received substantial increases in pay, low performing teachers were required to demonstrate improvement or were removed from the workforce. The researchers found that the threat of dismissal had a substantial effect, that is, previously low performing teachers improved their performance. They also found that the financial incentives for high performing teachers also improved their performance. Dee and Wyckoff conclude that, overall "high powered incentives linked to multiple indicators of teacher performance can substantially improve the measured performance of the teaching workforce" (p. 28).

States are experimenting with various approaches to linking performance and jobs. Florida, for example, eliminated tenure entirely and bases annual teacher contracts on their classroom performance (e.g., two years of unsatisfactory performance in a three-year period can lead to dismissal). Probationary teachers in Colorado and Delaware have to demonstrate at least two years of satisfactory performance to gain tenure. Rhode Island and Oklahoma stipulated that experienced teachers with two consecutive years of ineffective or unsatisfactory ratings are dismissed. More research is needed to shed light on whether these policies are strengthening teacher quality and leading to the student gains envisioned in various state policies concerning teacher effectiveness.

BARRIERS/CHALLENGES TO IMPLEMENTATION

[11] See, for example, Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2011) and Johnson and Fiarman (2012).

[12] See, for example, Grissom and Youngs (2016).

Thought leaders and scholars interested in educator evaluation systems have written extensively on the challenges and barriers states and districts face when implementing these systems. Here I focus on three central themes in that literature: (1) limiting compliance and maximizing genuine engagement; (2) responsibly dealing with associated costs; and (3) building a culture of continuous improvement.

Limiting Compliance, Maximizing Genuine Engagement

The goal of educator evaluation systems is to improve teaching and learning. Because teachers are central to that work, states and districts need their active engagement. A long line of scholarship across many fields demonstrates the intended and unintended consequences of mandates that are handed down within bureaucracies. *The Widget Effect*, a report that stimulated a great deal of interest in reforming teacher evaluation, offered a portrait of traditional teacher evaluation compliance: Nearly all teachers received high ratings; districts failed to recognize or reward excellence; professional development was rarely tied to results and when it was, little support was offered; new teachers were rated above satisfactory and no one was denied tenure; and poor performance was not related to teacher dismissal.[13] These findings echo some press coverage of the Michigan Educator Evaluation System (Moore, 2016).

Thus, a central challenge in implementing these new educator evaluation systems has involved securing broad-based stakeholder "buy-in" for these policies, including buy-in of teachers, principals, parents, teacher preparation programs, and the public. *There is consensus is that teachers will be more likely to genuinely engage in educator evaluation systems that they perceive as meaningful, relevant, and consequential.* Thought leaders and implementers offer several forms of advice. They urge a *strong communication plan* that allows for all concerned to learn about the system. This includes community information nights, email updates, media relations, workshops, videos, press releases, and newsletters. Some research has demonstrated that when teachers and principals do not understand the details of the adopted evaluation system, the system does not have positive effects.

Here the tension between local control and central guidance is especially relevant. Scholarship suggests that when teachers and school leaders feel like they participated in the creation and adoption of an educator evaluation system, they were more positive about its potential and effects. But local control can lead to the absence of rigor and quality, just as central control can lead to the lack of sensitivity to important local variation. One version of this concerns the generic quality of most observational systems. Elementary teaching can be quite different than high school teaching; mathematics teaching can differ considerably from social studies teaching; teaching English language learners can require specific skills; the list goes on. No research illuminates how best to balance subject- and grade level-specificity, or student- and context-specificities with the generic standards and rubrics that tend to dominate the classroom observation marketplace. Relevant here is also the fact that states (including Michigan) do not have standardized tests for several school subjects.

Further, as pointed out earlier, teachers are expected to enhance their students' civic-mindedness, ability to collaborate, and social and emotional well-being. But the reality of the current educator evaluation systems is that these additional aspects of teacher effectiveness are harder to measure, and to measure at a large scale, and thus are marginalized. Some research has also shown that aspects of teacher quality that are related to their outside-of-classroom work – building positive staff relationships, supporting other teachers' instruction, building relationships with parents and community – are highly valued by principals but not significantly related to value-added test scores of teachers. Yet when asked which teachers they would prefer to keep, principals reported that teachers who had those outside-of-the-classroom strengths were essential to

[13] See Weisberg, Sexton, Mulhern, and Keeling (2009).

their school's well-being.[14] Using tools and measures that reflect the realities faced by teachers and leaders in a specific district and school is crucial to securing buy-in from most stakeholders.

Another issue concerning minimizing compliance is creating a system that provides useful feedback. Several scholars have pointed out that the kind of information recorded in many observation protocols or documented in scoring rubrics does not provide teachers with sufficient information and guidance for how to improve. Student test scores, in addition, do not help teachers know what they should do to improve those scores. When teachers are forced to jump through a series of bureaucratic, high stakes hoops that do not help them improve, this can lead to frustration and ennui, both of which encourage compliance, a fact that clashes with the current overemphasis on achievement only.

Associated Costs

Doing something meaningful is often resource-intensive. Sartain and her colleagues noted several challenges associated with the resources required of educator evaluation systems, including: (1) a feasible observation numbers and timeline; (2) creating data systems for documentation and the associated data entry burden on principals, observers, and teachers; (3) training observers in both how to use the tools for classroom observation and how to have meaningful, constructive discussions with teachers about instruction and how to improve their practice; and (4) holding evaluators accountable.

The bottom line here is that observations take time and skill. Debriefing discussions about instruction and improvement take time and skill. Training observers takes time and oversight for quality assurance. Entering data into the bureaucracy's system takes time. There are a limited number of observers available to schools, which means that time gets squeezed and observations are less frequent and shorter. Research has demonstrated that observers' skills can also degrade over time, which requires re-retraining. Principals' time, which is often eaten up with unanticipated events that need immediate attention, is an inescapable reality. The transaction costs of the school bureaucracy are already high – for teachers and administrators. Take, for example, the time and material resources needed to administer Michigan's testing regime: Administrators need to manage the delivery, storage, and security of the tests; they need to communicate with parents and teachers about testing schedule and its importance to the school; they need to coordinate test administration with teachers and staff; and those duties and responsibilities are on top of the daily details of running their schools. Adding more to their plates – namely multiple teacher observations and debriefs, annual evaluation meetings, and the management of all of the associated data -- without taking something away is a recipe for disaster.

Thus, the success of an educator evaluation system also requires that teachers take on new roles (for instance, as mentors and observers); principals reorganize their time to leave adequate time to conduct meaningful observations; and that states and school districts consider the resources and infrastructure necessary to support the work. Here some scholars also suggest that aligning policies and practices within the system is essential. Teachers who receive one message about how to teach from a curriculum adoption and receive another, conflicting message about how to teach from the teacher evaluation system, and yet another message about how to teach from parents or testing or professional development on working with diverse learners are caught in the middle of a confusing, conflicting policy environment. This leads scholars to argue that professional development needs to be aligned with the evaluation systems, as do new curricular, instructional, and accountability mandates.

Building a Culture of Continuous Improvement

In the end, installing educator evaluation systems in schools requires that teachers and principals share a language of both instruction and its improvement, and share a common assumption that teaching and

[14] See Grissom, Loeb, and Doss (2016).

learning requires continuous improvement. The consensus view is that the improvement of teaching depends on collaborative work, including planning and the analysis of student work. But the culture of schools has long worked against this agenda, as there has been little time for teacher and leader deliberation, and little leeway for experimentation. The accountability movement, starting with NCLB, while it might have intended to lead to improvement, narrowed schools' goals on to the measured outcomes of standardized tests in literacy and mathematics and proved a conservative, rather than enabling, force for improvement. Many teachers and schools started "teaching to the test," which is not usually the same as teaching for understanding.

In some contexts, teachers were under attack, as the finger of blame was pointed at them. But as Danielson notes, this led to worrisome consequences:

> 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 . . . it is time for a major rethinking of how we structure teacher evaluation to ensure that teachers, as professionals, can benefit from numerous opportunities to continually refine their craft. (p. 1)

In one study, Kraft and Gilmore (2016) interviewed principals about their evaluations of teachers in their schools. Evaluators in one urban district reported that they thought that many teachers in their schools deserved a "below proficient" evaluation, but the principals did not consistently give teachers those low ratings. When interviewed about their reasons for inflating evaluation scores, the principals reported a range of reasons: (1) time constraints (principals needed to spend more time observing less effective teachers, time they did not have); (2) the demotivating consequences of a poor evaluation for teachers who had the potential to improve; (3) the principal's own personal discomfort in giving negative evaluations; (4) avoidance of the "long, laborious, legal, draining process" of teacher dismissal and the fear of not having any adequate replacements. As one principal noted: "The one you know is better than the one you don't."[15]

Clearly, simply mandating a change in the rigor and quality of teacher evaluations is insufficient for the substantial change such evaluations will require. And making good on Danielson's goal of using such evaluations to help teachers improve their practice will require teachers and principals alike understanding the educator evaluation systems, feeling empowered to use them and adapt them to local contexts that matter, and trust that schools in which teachers and their leaders regularly interrogate their goals and accomplishments – and take next steps to redress those – will be rewarded.

[15] See Kraft and Gilmour (forthcoming).

CONCLUSION

Based on their analysis of states' experiences creating and implementing educator evaluation systems, Goe et al (2014) propose a "design process" of crucial steps involved in the implementing such systems, including:

(1) specifying evaluation system goals;
(2) establishing standards;
(3) securing and sustaining stakeholder investment and cultivating a strategic communication plan;
(4) selecting measures;
(5) determining the structure of the evaluation system;
(6) selecting and training evaluators;
(7) ensuring data integrity and transparency;
(8) using teacher evaluation results in appropriate ways;
(9) and evaluating the system. (p. 1)

The field has a long way to go before there is solid empirical evidence to inform the micro and macro issues associated with implementing educator evaluation systems in all of these steps. That said, as NCTQ (2001) notes, "Teacher effectiveness measures don't have to be perfect to be useful" (p. 1). But using them requires critical awareness of their strengths and limitations. *The consensus among scholars is that the state systems that are implemented need to be regularly evaluated both for bias (by subject matter, measures used, student or teacher populations effected) and for their effectiveness in improving teaching and learning.*[16]

This would suggest that the process of engaging in this review of Michigan's educator evaluation system is appropriate and timely. A review of the complete set of lessons learned also highlights several issues relevant to the deliberations. I note just three here.

First, by erring on the side of choice (in observation system, in the retraining of raters, in student growth data and assessments, for example), the state has allowed for local adaptation and (perhaps) greater local buy-in but (perhaps) at the cost of

- ensuring rigor, quality, and meaningfulness in the implementation of teacher observations. The observation instruments that were adapted statewide were never "pure," but instead were derivatives of those nationally-available tools. Those derivations may have corrupted the tools' integrity.
- state capacity to provide necessary supports for building local school and district capacity in all of the ways in which these educator evaluation systems press on local districts: observer capacity, capacity to use and develop metrics, capacity to analyze a range of data, etc.

Second, by under-funding the mandate, the state has compromised district and school capacity to train observers, and provide sufficient time for leaders to observe in classrooms.

Third, by confounding the goal of supporting continuous improvement in Michigan schools with the goal of identifying weak teachers who might be counseled out of the workforce, the state has been unable to develop leaders' and teachers' buy-in and trust in a system of evaluation that identifies areas of growth for all teachers. This is reflected in the earlier resistance to a three-point system. Even though the precision of a meticulously used tool (not to mention a tool that has been unevenly used) is by far best at the tails of the scores' distribution, the desire to institute a four-point system instead encourages a ranking or comparative orientation over an improvement-centered one.

[16] See NCTQ (2011), and Goe, Holdheide, and Miller (2014).

These observations are not unique to Michigan. In a series of telephone interviews with local school districts concerning their teacher evaluation policies and practices, principals and school superintendents report anywhere from 1 to 38 annual observations of teachers, from 5-10 minutes for drops ins and 20-45 minutes for "long" observations. Some report that leaders and observers go through extensive training, others report adapting well-known observation protocols to their own needs and interests.

Overwhelmingly, what we have learned from 10 years of experimentation with educator evaluation systems reaffirms the old adage: "Garbage in, garbage out." Some studies suggest that well-implemented educator evaluation systems (ones that use well designed tools, are based on clear expectations and standards, and invest building trust among educators) can lead to improved teacher effectiveness scores when measured by student achievement scores. But we also know that the realities faced by states, districts, and schools means that the quality of the systems' implementation is regularly compromised with pragmatic short cuts that save time and money.

Further, we still know little about how to scale these reforms up in ways that maximize teachers' opportunities to continued growth and minimize the potential for deskilling teachers and increasing unnecessary paperwork. We also need to know a great deal more about how to catalyze and support the cultural changes in schools to allow teachers and their leaders the time needed to improve instruction, and how to build an educational policy system supports improvement instead of obstructing it.

REFERENCES

Cohen, J., & Goldhaber, D. (2016). Observations on evaluating teacher performance: Assessing the strengths and weaknesses of classroom observations and value-added measures. In J. A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 8-21). New York: Teachers College Press.

Danielson, C. (2016, April 18). On rethinking teacher evaluation. *Education Week.*

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E. H., & Rothstein, J. (2011). *Getting teacher evaluation right: A background paper for policy makers.* Washington, DC: AERA and NAE.

Dee, T., & Wyckoff, J. (2013). *Incentives, selection, and teacher performance: Evidence from IMPACT.* NBER Working Paper Series. Working Paper 19529. Cambridge, MA: National Bureau of Economic Research.

Goe, L., Holdheide, L., & Miller, T. (2014). *Practical guide to designing comprehensive teacher evaluation systems: A tool to assist in the development of teacher evaluation systems (revised edition).* Washington, DC: American Institutes for of Research.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis.* Washington, DC: National Comprehensive Center for Teacher Quality.

Grissom, J. A., Loeb, S., & Doss, C. (2016). The multiple dimensions of teacher quality: Does value-added capture teachers' nonachievement contributions to their schools? In J. A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 37-50). New York: Teachers College Press.

Hill, H. C., & Grossman, P. L. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 82*(1), 123-141.

Johnson, S. M., & Fiarman, S. E. (2012). The potential of peer review. *Educational Leadership, 70*(3).

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment.* Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://www.metproject.org/reports.php.

Keesler, V., & Howe, C. (2012). *Understanding educator evaluations in Michigan: Results from year 1 of implementation.* Lansing, MI: Michigan Department of Education.

Keesler, V., & Howe, C. (2016). Teacher evaluation in Michigan. In J. A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 156-168). New York: Teachers College Press.

Kraft, M., & Gilmour, A. F. (forthcoming). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher.*

Moore, L. (2015). Without guidelines for Michigan teacher evaluations, everyone is awesome. Download from http://www.mlive.com/news/index.ssf/2015/12/unlikely_allies_behind_teacher.html.

National Council on Teacher Quality. (2011). *State of the states: trends and early lessons on teacher evaluation and effectiveness policies.* Washington, DC: Author.

Pianta, R. C., & Hamre, B. K. (2016). Implementing rigorous observation of teachers: Synchronizing theory with systems for implementation and support. In J. A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 22-36). New York: Teachers College Press.

Sartain, L., Stoelinga, S. R., & Brown, E. R. (20100). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation.* Chicago, IL: Consortium on Chicago School Research.

Taylor, E. S., & Tyler, J. H. (2012). Can teacher evaluation improve teaching? *Education Next*, 12(4). Downloaded from http://educationnext.org/can-teacher-evaluation-improve-teaching/.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. Washington, D.C.: The New Teacher Project.

Voices from the Field: Practitioner White Papers[17]

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 activity questions posed by MDE and the other data gathering components. Aside from this the practitioners were free to construct their papers.

The first paper of the papers is by Crystal Wise, a current doctoral student at the U. of M. In her paper Ms Wise reflects experience gained as a teacher, principal and central office administrator.

The second paper is by Ann Blais, also a doctoral student at the U of M. She draws on her extensive experience as a high school teacher, and department supervisor in an east coast charter school and larger more generously resourced traditional school. Ms Blais discusses teacher evaluation, observations, feedback, and supervision in the real world context of schools. In this view we see the balance of evaluation to guide and improve teaching and learning, and evaluation to meet organizational and systemic needs. We also see the complexity of meeting coaching, feedback, direction and evaluation needs over the arc of a teacher's career.

Rosalind Brathwaite is the author of the third paper. In her paper we learn of her experience as a new principal implementing the educator evaluation system. She describes her expectations and needs from the system.

.

Collaboration in Teacher Evaluation - by Crystal Wise

Best Practices in Teacher Observation: One Practitioner's Perspective - by Ann Blais

A Close-Up View of K-12 Educator Evaluation - by Rosiland Brathwaite

[17] 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.

Collaboration in Teacher Evaluation

Crystal Wise
University of Michigan

Collaboration in Teacher Evaluation

Teacher evaluations can be a source of anxiety, particularly for beginning teachers. At the time of my first teacher evaluation, I did not know what to expect. That year, I had worked with my principal organizing school community events where she appreciated my dedication and I was able to begin to apprentice her leadership skills, however when it came time for my evaluation, I was nervous. The day of my scheduled pre-conference, the usual crowd of teachers had begun to congregate for the typical after school chatter. Over the course of the school year, these teachers had become comfortable offering their "pearls of wisdom". On my way to meet the principal, my unofficial mentors advised me to just put on "the show" - teach the way my teacher education program had taught. During the pre-conference, my principal briefly explained the teacher evaluation form and process. My principal asked, "Do you have any questions?" I did not. Maybe I did. I was not sure. Politely I stated, "No, not at this time." The next day the observation and post-conference occurred uneventfully. During the post-conference I scanned my teacher evaluation form. I had received all markings to suggest that I was an "effective" teacher.

Since my first year of teaching, the teacher evaluation process has evolved. This has been in part of due to school reform initiatives and policies such as No Child Left Behind and the American Recovery and Reinvestment Act that provided the Race to the Top grant (NCLB, 2002; U.S. Department of Education, 2009).

The Race to the Top grant emphasized the use of student outcomes as a means to measure teacher effectiveness. Though research has found that teacher effectiveness is associated with student outcomes (e.g., Heck, 2009), research cautions against using student outcomes as the sole means for evaluating teacher effectiveness (e.g., Darling-Hammond, Amrein-Beardsley, Haertel & Rothstein, 2012). Darling-Hammond et al. (2012) recommend that evaluation systems not only link student outcomes with teacher performance, but also that these systems (1) provide training for evaluators, (2) provide frequent evaluation and feedback to teachers as well as mentoring and professional development, (3) institute a committee of evaluators that includes both teachers and administrators that follow due process procedures to make personnel decisions. Stronger emphasis on the evaluation process as a way to improve student outcomes would have helped to eliminate my nervousness. I should have thought of the evaluation as an opportunity for professional growth rather than reprimand. I recount my experience as a first-year elementary school teacher to first consider what it would mean for teacher evaluation systems to emphasize improving student outcomes by providing frequent evaluation and feedback to teachers and second to better understand how mentorship and professional development is needed for both individual teacher growth as well as building community.

One formal evaluation per school year is not conducive for professional growth. Informal observations can also provide a way for administrators and teachers in engage in reflective and meaningful conversations (Danielson, 2012). These informal observations can be brief drop-ins or also be conducted my mentors. Regardless, of conducts the informal observation, it is important that the conversations and feedback are useful. Professional learning communities are one way to provide this support. Professional learning communities (PLCs) that provide teachers with the opportunity to work collaboratively to critically reflect on their practice have been found to have positive effects on developing teacher's professional knowledge and increasing student achievement (Vescio, Ross, & Adams, 2017). Continuous participation in PLCs would also encourage frequent observations and consistent feedback. PLCs allow the principal and teachers to consistently in engage in the process of developing a course of action that identifies particular steps to increase student achievement through feedback on how to implement informed modifications to their teaching practice. Teachers respond

positively to evaluation in schools where they receive frequent observations and feedback (Jiang, Sporte & Luppescu, 2015). Positive reactions to the evaluation process suggest that teachers would remain engaged in teacher development that can improve student outcomes.

I was fortunate to teach at a school where veteran colleagues expressed an interest in helping me to navigate my first years of teaching. To maximize the collaborative nature that existed, it would have been more beneficial had their mentorship been explicitly grounded in a standards-based teacher evaluation system and our school improvement efforts. My colleagues' mentorship initiated me into the realities of teaching, but it was not strategic in implementing research-based practices. There was a mismatch between teachers' beliefs of acceptable teaching practice and what they enacted in their day-to-day teaching practice. Developing PLCs would allow teachers and administrators to work collaboratively to define good teaching for their context and to set the expectations for acceptable and superior teaching practice. PLCs engage teachers in "making explicit decisions based on their contexts, their goals, current and new professional knowledge and the needs of their students" (Vescio, Ross, & Adams, 2008, p. 89). Within a PLC, my colleagues and I could have identified those research-based instructional practices that were suitable to our contexts. Within the PLC, the areas for growth would have been identified through evaluations and teachers could work together to share their expertise and consult teaching resources to provide better instruction. Continued participation in the PLC provides a systematic way to use research-based practices and teacher expertise to fine-tune instruction to meet their needs of students.

Implementing an effective teacher evaluation system does not come without challenges. In the schools where I taught, there was typically one administrator responsible for conducting evaluations. This is a considerable burden for one administrator to shoulder. Principals are not likely to have the time to conduct the formal evaluation process with all the teachers that require it. It's likely my principal provided such a brief description of the evaluation rubric and limited feedback because time was a factor. Given that I had access to a community of teachers that were willing to provide mentorship, it would have been beneficial to develop a PLC that included teachers and principals. With the support of the principal, the PLC could engage in professional

development that would provide opportunities for teachers to reflect on their practice in relation to the school's definition of good teaching and a standards-based teacher evaluation system. As discussed previously, these PLCs can provide teachers with the more opportunities to receive feedback and support in improving their teaching practice. To help develop PLCs, principals would need to provide teachers with time to reflect, plan, observe each other, and engage with teaching resources that support improvement. In schools, finding time is often a barrier to collaboration. Vescio, Ross, and Adams (2017) offer one way to address this issue. These authors suggest that videotaping and reviewing lessons can also foster collaboration within PLCs. Sharing videotaped lessons can reduce the time it would take to observe other teacher's practice and can be used to generate discussion in person or through other formats.

Another barrier to effective teacher evaluations is developing trust in the process. Principals must provide teachers access to trustworthy resources. The source of advice and professional development needs to be knowledgeable (Danielson, 2012). In PLCs these sources can be the principal, outsider professional development, teacher resources such as books, or successful teachers within the school community. It is important that teachers are provided with time to discuss these resources so that they may determine how best implement the practices in relation to the school's shared understanding of good teaching and also in service of meeting their students' needs. Additionally, these conversations help teachers to develop a deep understanding of the shared expectations of good teaching in their context. Therefore, when they are evaluated they would be clear of expectations and would understand that the evaluation is part of their professional development where they have a learning community that supports them.

This reflection on the teacher evaluation process only provides a few aspects of an effective teacher evaluation system. The aim of this reflection was to demonstrate that by building trust, developing a common definition of good teaching, and providing opportunities for teachers and administrators to collaborate and share the responsibility of instructional leadership, the evaluation process can foster teacher growth and development, which can also lead to increased student achievement.

References

Danielson, C. (2012). It's your evaluation - collaborating to improve teacher practice. *The Education Digest, 77*(8), 22-27.

Darling-Hammond, L. Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *The Phi Delta Kappan, 93*(6), 8-15.

Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability, 26*(1), 5-28.

Heck, R. H. (2009). Teacher effectiveness and student achievement: Investigating a multilevel cross-classified model. *Journal of Educational Administration, 47*(2), 227-249.

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher, 44*(2), 105-116.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

U.S. Department of Education. (2009). Race To The Top: Executive Summary. Retrieved from https://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Vescio, V. Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*, 80-91.

Best Practices in Teacher Observation: One Practitioner's Perspective
by Ann Blais

Introduction: my background, teaching, my evaluation experiences

I have served in a variety of school roles since 1993, but have always kept one foot in the classroom as a high school English teacher. Formally, an East coast city school system trained me in teacher evaluation for my role as department chair for English at a city high school, in 2002. The training was conducted over a number of day-long sessions, and was required for earning my administrative certification in the district. It consisted mainly of familiarity with routines and documentation, including supervision of the annual teaching portfolio. The emphasis in this training seemed to be upon how to use the multi-step system to efficiently and legally remove ineffective teachers from service. The training did not emphasize how to help teachers learn from practice.

Informally, I was prepared for my supervisory responsibilities with sharply contrasting supervisory experiences, early in my career. I began full time teaching at a start-up charter high school in New York City, in 1993. During my training the prior year, I had spent time outside of my official internships to watch the mother of a friend teaching science at a local middle school. Her connection with students and effectiveness as an instructor made me eager to work for her when she was named the head of a local New Visions charter school.

Looking back, it was not a situation in which I should reasonably have expected strong supervision: she was a new principal, I was a new teacher, and we were both busy with the mechanics of starting a brand new school. Instruction, unfortunately, was not the focus. My principal visited my classroom on occasion. Our discussions were usually focused upon my lack of classroom management skills. I knew that her advice highlighting the centrality of student engagement was sound, but at the time it wasn't specific enough to be helpful. I remember one occasion in which we discussed a student who was acting out. She asked me, pointedly, "Do you know if he can read?" I wish she were still alive today, so I could share with her that this question fueled the better part of my learning as a teacher over the next 15 years. But, in the moment, I needed concrete advice. I needed someone to come into the room, regularly. I needed someone to follow up with students who acted out. I needed a supervisor to meet with me to plan routines and lessons that would make the class more engaging. I realized also that I needed to learn more about teaching reading, which is not something secondary English teachers usually receive unless they have special education training. Instead, my principal laid my students' misbehavior at my feet: if my lessons were more engaging, if I knew how to establish routines, if I gave them more work, I wouldn't be struggling. Of course, she was right. But in the moment, struggling and frustrated, the blame was demoralizing. I needed a supervisor who had the time to actively invest in my growth. Looking back, I realize of course that with the demands she faced, what she (and the students) needed was a more experienced English teacher. I decided to quit and search for another job where I would be in a larger community of English teachers that I could learn from. It is lucky that I left only that job and not the teaching profession, entirely.

I joined a school outside Boston with a large budget, which served a broad economic and ethnic range of students but was majority white. It wasn't the population of city students I had gotten into teaching to serve, but I knew I could learn how to teach there. The school had an English department with 40+ teachers and a paid administrator devoted entirely to our department. My supervisor paired me

with a teacher of special education to co-teach an integrated class of students who needed accelerated learning to catch up to standards before the state test. She assigned me a mentor who met with me weekly to help me plan my other classes. She made sure my office was right next door to my teaching partner and other colleagues. And she met with me for an hour every week, no matter how busy she was. We discussed my unit planning, difficulties with students, successes, approaches to different texts, how to respond to student writing. She visited my classes 2-3 times each semester. Some of the observations were planned, and we would meet ahead of time to discuss and sharpen the lesson plan. Some were surprises: I remember she appeared at my door on Halloween once, and I was observed in full *Scarlet Letter* costume. I always wanted to impress her, but I never felt nervous or shamed, no matter what mistakes were made. She just expected me to keep learning. Each time she watched me teach, she left me with an amazingly complete handwritten script of everything she saw and heard in the lesson. I would study that script, and be prepared to talk with her after the next day. I took copious notes during my meetings with her; there was so much to learn. She was a wealth of information not only about students and pedagogy, but also specifically about how to teach English. I felt both challenged and supported. To this day, she is one of my wisest and most trusted counselors, a person I return to with difficult problems of practice.

When I became a department chair years later, I employed my mentor's methods. I met with my new or struggling teachers weekly, I scripted every class I watched, and I never, ever forgot how hard any of my teachers were working. I always entered their classes with a smile and with trust that they were doing the best they knew how and continually seeking improvement. My approach was modeled upon my mentor's practices, but also informed by several years of work in the interim as a professional developer for a non-profit in Washington D.C., where we conducted standards alignment reviews as well as collaborative professional development sessions based in assessment of student work. Since this experience, I have worked in two teacher education programs (Brandeis University and the University of Michigan), supervising new teachers.

Best practices in educator evaluation implementation

Best practices: individual level

Early on in the process of being a formal supervisor, I realized that without a personal level of trust, evaluation could not be a genuine learning experience. Teachers must be willing to share when things are *not* going well. When I accepted the job of department chair at an east coast city magnet school, my department was deeply distrustful of the administration and fragmented in terms of its practices. On a personal level, they were a smart, creative community that supported one another, read books together, and enjoyed one another's company. I had to find a way to accomplish seemingly contradictory goals of aligning the department's instruction to our students' assessments and somehow gaining teacher trust.

The way in was unexpected: in the midst of various personal crises, teachers reached out to me for help – to listen, to cover a class, to hear them out with a tricky situation. I responded as supportively as I could in each situation, acting in such a way that made it clear that I would help and protect them. I would help them through difficult days, I wouldn't let the central administration supplant our goals, I wouldn't let the principal impose a curriculum the teachers didn't design, and I would take responsibility as a supervisor for teachers' missteps. Because I had my teachers' backs outside of the classroom, I could begin to ask more of them inside of it.

Inside the instructional work, my goal was to get teachers to think critically about their practice and to learn from it. The evaluation practices that lead to teacher growth are ones that respect and engage the teacher's thinking. The process of seeing a lesson plan, talking about it, watching its execution, and offering feedback are not *my* processes as an observer – they should be framed as an inquiry process *the teacher* is engaged in. The observer's role is to shift the ownership of that observational evaluation cycle, through questions and genuine curiosity: to have the teacher's questions drive the lesson and reflection. A strong observer cultivates disciplined habits of mind in the teacher. Without trust and an inquiring mindset, evaluation devolves into cursory, bureaucratic exercise: a wasted opportunity for teacher learning.

The most important practice of a teacher evaluator is then to cultivate a genuinely inquiring stance: an observer should want to know what the teacher is aiming to do, and why, how he thought the lesson worked, what evidence of student learning or engagement he noticed, and how he might adjust. In this sense, in the pre-observation conference (and informally, all the time) an observer is modeling for the teacher the inquiring.

During the lesson, best observer practices are ones that provide the teacher information with which to work. After checking in to be sure it's still a good day to visit, an observer should 1) script the lesson in its entirety as objectively as possible, 2) attend to the questions or the particular students the teacher focused on prior to the lesson, and 3) listen to and engage the students, without disrupting or undermining the teacher. A good observer collects information about whether the purpose of an activity is clear to students, what their misconceptions and levels of understanding are, connections they make (or don't), and potential opportunities for engagement. In short, while the teacher is focusing on instruction, an observer can offer insight into student learning. To collect this information without undermining the teacher through her presence, an observer has to adopt a friendly stance. The observer must partner with the teacher and be genuinely curious about the students' thinking. An observer cannot forget he is a guest. Good manners matter.

After the lesson, an observer should thank to the teacher and the class for the visit, and offer quick points of praise for the teacher privately. The speed and positivity of this initial feedback is important, as it sets the stage for the teacher to process the written feedback. Before I have even left the classroom, I email the teacher with my written transcript and set a time within a 1-2 days for a follow up conversation. Having the transcript notes and a day or two to think allows a teacher access to both in-the-moment instructional decisions and later reflections. That makes it more possible for teachers to drive the inquiry process.

My focus as an observer during the follow up conference is to help teachers learn to analyze their instruction critically. A supportive, inquiring atmosphere allows teachers to do more than analyze, but also *to learn how to learn from their own practice*. Discussion can be guided by the teacher's reflections on the transcript. I try to limit myself to questions and observations that will help the teacher answer her own questions. Often, I find that the lines of inquiry I would have introduced are already covered by the teacher. And, often, I have found that questions and ideas surface that I never would have thought to ask about. In the main, aside from cultivating this inquiring mindset, I see my job as a supervisor as one of focusing teachers on student learning as the main resource, as the most effective driver for their own learning and instructional thinking.

I try to target feedback and suggestions to individual teachers' career stage. Newer teachers need

feedback on management and more help in planning, prior to instruction. Teachers past the novice stage are able to be more attentive to student thinking and can more quickly move to discussing student responses to work to refine their instructional choices. Teachers advanced in skill can be guided to articulate the tacit aspects of their knowledge so it can be shared with the department; often, conferences with master teachers become opportunities to cultivate teacher leadership. And, finally, teachers advanced in experience but not in skill need a variety of responses: some are ready to continue learning, some are not. Some need gentle guidance to exit the profession, some need a next level of challenge, for example, to take on a professional development experience. Whatever their stage, all teachers require the respect of honest feedback, especially when performance is not effective. It helps to enter these conversations with the assumption that teachers are professionals who care about improvement and will prioritize student needs over their own personal feelings.

The evaluation cycle happens formally, with mandated frequency, but I found that evaluating informally on a more regular basis normalized my presence in the classrooms and made the entire process much easier for teachers. I learned a lot watching a wide range of styles and practices, usually enjoyed my visits, and tried to communicate this to teachers. Frequency was important to ensuring that the questions and observations I shared would be received, and that teachers would feel comfortable sharing their struggles with me honestly.

Across the levels of teacher experience, I have found that one aspect of best practice in supervision is to keep the focus on the students in relation to the particular instructional content. Teachers needed to talk about how to teach English, specifically: how to introduce or sequence a particular text, what reading strategy seemed most appropriate for a particular group of learners with a particular text, how to represent a complex idea for a particular grade level, what the teachable concepts of a particular text might be for a particular book, how to assess an assignment efficiently, how to know if a class read the book or comprehended it or not. These conversations built pedagogical content knowledge for the teachers, for myself, and for the department as a whole. I encouraged teachers to capture and preserve these ideas for their team curriculum planning efforts.

Finally, it is important to note that the practice of individual supervision is best accomplished in the context of support and investment in each teacher's professional success. Outside of the supervisory cycle, I continually followed up with difficult students, facilitated meetings with parents, created opportunities for collaborative sharing of knowledge, and promoted teacher leadership. I continually and conspicuously spread positive gossip about the great things I saw going on in my teachers' classes -- with them, with the administration, with students and parents. Official classroom observation works best as an occasional, unsurprising formalization of the daily work of positive, collaborative reflection.

Best practices: community of practice level

Although my individual work with teachers was fruitful and important, the work of supervising teachers is much more about building community. It is about leveraging teachers' affection for one another and their shared enthusiasm for books and students into a community of shared practice. Supervision is most effective when it occurs within the context of building communal knowledge. A shared goal of collaborating and reaching more students is larger than any individual ego resistance or conflict.

As supervisor, my aim was to foster a vibrant, collaborative community of practice. But, to be clear: that aim was a means to my primary goal of ensuring effective instruction for students. This meant

that I had to be willing to articulate a high standard of practice and to let teachers go as needed. That stance sometimes felt at odds with the important work of building collaborative community, but, to be frank, the teachers who were most invested did not seem to mind that I had to challenge colleagues who (everyone knew) were not pulling their weight. They would let me know, quietly, that they shared and appreciated the high standards. But because of the department's bonds of friendship, this was difficult work. The growing pains were real. Within the first two years of work, seven out of fifteen teachers left – five of their own accord. But those remaining were committed, and actively invested in the learning of the new faculty.

To pull the department together, I had to work hard outside of formal observation to fund release time and payment for curriculum teams to assess student work, design curriculum, and be trained together. I covered classes so that they could observe one another teach. I utilized my department meeting time for focusing on student work, training on reading strategy instruction, and making real curricular decisions. I involved department members deeply in the hiring and mentoring of new teachers, which surfaced a group discourse about what kind of teaching we thought was important for our students.

My teachers were a wonderfully contentious, brilliant bunch. It feels good to see that 10+ years later, the curriculum and personnel have evolved but the department remains cohesive; many of our hires remain, and there still seems to be a culture of collaboration. I hope that I helped to foster that. Whatever I may have contributed, I am sure of one thing: that in watching them teach, listening to them think, and facilitating their work, I learned more than I taught.

In summary, strong observers know that they are not "the" critical resource and that individual evaluation is not meaningful outside of the context of deep collaboration. A strong observer works to facilitate a community of collaborative practice, to create a culture in which struggle is normalized and learning is shared, and to send the message that this work is too hard for any of us to do alone.

Barriers to implementation, ways to mitigate

A significant barrier to orienting the evaluation process towards teacher learning is the content-neutrality of many observation processes: the tools are not content specific, and the observers do not consistently provide content-area expertise. An observer who does not know the subject can offer meaningful feedback on management and student engagement, but only limited information on the nuances of student understanding. For example, he or she can share neither common student misperceptions about a text or topic, nor help the teacher to enlarge their content-specific repertoire of ways to represent the content and engage students critically. I have found that while new teachers benefit from feedback on clarity and engagement that a content-neutral observer can provide, all teachers need content-specific feedback in order to become more effective. Especially, teachers beyond the novice stage need content-specific supervision if they are to go from "good" to "great." That's their level of challenge. It is important for schools to provide differentiated learning for teachers at different stages of their careers. This kind of growth in pedagogical content knowledge requires content-specific feedback.

Another significant barrier for implementation of effective evaluation is administrator time. New teachers need more time and assistance than a pre-observation conference typically provides. Sometimes a supervisor can place a new teacher on a strong instructional team or with an experienced co-teacher to facilitate this planning assistance. But absent this kind of teamwork, a content-knowledgeable supervisor needs to be available on a regular basis for significant periods of time to assist with planning for strong

student engagement. When prior planning assistance is provided, classroom engagement is stronger, and new teachers can attend to student learning earlier in their own learning curve.

The measurement of student growth using alternative methods/tools and processes

I have not worked in a school system where the student growth measures on norm-referenced tests had been incorporated into teacher evaluation; policy had not mandated that practice when I began my career, and as yet, Massachusetts' various policy levels are still locked in disagreement about how to implement this Race to the Top requirement. When I worked as a department chair, each teacher had to prepare a portfolio to demonstrate student learning, annually. We worked as hard as we could to ensure that the portfolio goals and evidence were part of the teachers' learning curves and netted useful conversations, but mainly the portfolio process seemed an onerous add-on process to the inquiry that was already driving our teamwork. The portfolio was an external demand that was not strongly enough integrated into daily planning and teaching to be a priority for teachers. Most teachers did feel the portfolios were useful for propping open the windows on hot days, however. In my district in Massachusetts, curriculum teams of teachers divided into Professional Learning Communities (PLCs) designed team goals and collected data to measure our progress. In this way, we were engaged in data driven inquiry that focused our practice.

If I were to take on a supervisory role now, in a system where student-level data is incorporated into teacher evaluations, I would want to use the data formatively. The data would need to be available to teachers on a timely basis, and teachers and observers would need enough training to interpret it. If these conditions were met, we could use the data to inform teacher learning and growth. In keeping with the goal of using evaluation for teacher learning, I would want to assess teachers on their *use* of student-level data to inform instruction, not just as an end-measure of successful teaching.

Cultural competency in evaluation models

The Massachusetts teacher evaluation framework does include an indicator that describes "reaching all learners," but the practices that accomplish this are not specified. Similarly, the Danielson framework is intended to be aligned throughout to "culturally responsive" practices. However, I have never worked in a district that was able to systematically ensure, for example, the delivery of Culturally Relevant Pedagogy (Ladson-Billings, 1995). I have seen teachers and teams of teachers work to connect school learning to student interest, leverage students' cultural "funds of knowledge" (Moll et al, 2008), and ensure the success of all – but such efforts usually seem comprised of a small group pushing against the normal systems of curriculum and instruction. Systems of formal evaluation could much more effectively foreground culturally relevant pedagogies by incorporating the practices and mindset suggested into the evaluation rubrics, directly.

Recommendations

1. Shift the focus of evaluation policy from individual *teacher* quality to systemic quality of *teaching;* require districts to build in time and systems for teacher collaboration and learning
2. Shift the focus of teacher evaluation from summative to formative, for teacher learning
3. Empower and pay evaluators as licensed administrators
4. Ensure that content specific evaluation is offered
5. Ensure the knowledge and competency of evaluators by:

   a. Ongoing training on how to engage teacher reflection, individually and collaboratively
   b. Training for content specific evaluation, and
   c. Providing practice-based supervision of evaluators' supervisory practice
6. Ensure that evaluators have sufficient time for best practices in evaluation, including increased frequency
7. Ensure that teacher evaluation rubrics are specific to:
   a. Content area
   b. Practices and mindsets for Culturally Relevant Pedagogy
   c. Different career stages
8. Where data is incorporated into teacher evaluation:
   a. Give teachers and evaluators data use training
   b. Provide data in a timely manner
   c. Provide the opportunity to use data formatively for instruction
   d. Balance evaluation of teachers' *use* of data with value added outcome measures

"A Close-Up View of K-12 Educator Evaluation"

By
Rosalind Brathwaite

Background and Introduction

Teacher evaluation has become a major focus in educational policy debates and research efforts. Ask yourself what is the purpose of evaluation? Is it compliance, routine, to create anxiety or job security? In my opinion it's none of the aforementioned. The purpose of teacher evaluation is to accelerate professional development that leads to instructional improvement to increase student achievement. This paper is my reflection on the educator evaluation process based on my experience as a secondary school principal.

In 2013, I was hired as a principal in a new school district. On arrival I found that the teacher evaluation system was developed collaboratively by a combination of district and school administrators, and was first implemented during the 2011-2012 school year. I was not involved in that process. Using the requirements provided by the Michigan Department of Education, the evaluation program consisted of four major components that included multiple classroom observations and feedback, student achievement and growth data, professional growth plans, and professional responsibilities that are expected of all staff members. After two years of implementation, the district decided that the instrument was not robust, lacking significant emphasis on student achievement. After receiving notice of the school's status of a priority school in July of 2013, the principal of the school, as well as 50% of the staff were immediately replaced, prior to the start of the 2013-2014

school year. I came aboard August 2013. Due to the fact that only a minor weight of the previous teacher evaluation instrument was connected to student achievement growth, this instrument led to inflated teacher effectiveness ratings. After reviewing the school's results on the state assessment in 2013-2014, the School Board requested that a greater percentage of the teacher evaluation instrument be tied directly to student achievement growth. The district leadership met to review the evaluation process and instrument. The process for evaluating this instrument was based on the low performance of the school compared to quantity of teachers who received effectiveness ratings of "highly effective" and "effective." The district leadership made the decision to create a new evaluation process, and to dedicate a much more significant weight to the student achievement growth component. Other components of the evaluation process include assessments of the teacher's planning and preparation, classroom environment, instructional practice, and professionalism.

During the 2013-2014 school year a new teacher evaluation instrument was developed. In September 2013 the staff was surveyed regarding the components that should make up the evaluation. This survey communicated to the staff that a weight of 40% of the overall teacher evaluation will be directly tied to student achievement in 2013-2014, and 50% in 2014-2015. The weighting of student achievement was a non-negotiable.

Barriers to Implementation

The evaluation included aspects related to professional learning, planning and preparation, professionalism, and formal observation ratings. These factors are in alignment with the responses of the staff on the survey. The initial process of training was conducted at the district and school leadership levels, focusing on Charlotte Danielson's Four Domains of Effective Teaching. School administrators were trained on each of the domains and the multiple components within each. District and school leaders took this knowledge and applied it to the classroom setting in a series of co-walk

observations, including debriefing discussions about what evidence was observed in the classroom. This was done in an effort to guide evaluators away from the concept of utilizing opinion and interpretation of what was taking place during instruction, and moving in the direction of citing evidence of what was observed in the classroom. Co-walks consisted of a group or team of administrators that focused on particular areas of the Danielson's Domains 2 and 3. Discussions took place prior to entering the class. Upon completion of the walk-through the team shared their findings amongst each other and provided immediate feedback to the teacher. I thought this was an effective practice. This contributed to the development and implementation of common "look-fors" and "listen fors" which were implemented across the district in observations. This user-friendly guide aided the evaluator(s) with quick tips for each of the Domains to improve teaching and learning. Although the guide was user-friendly and the co-walks effective, I do not believe adequate time was devoted to training teachers for proper implementation.

Once the leadership team completed their training they were responsible for training their instructional staff members on the Danielson's Domains. This gave teachers the understanding of the evaluation's observation component expectations. Initial implementation for teachers became difficult because it wasn't their way of teaching. It took my staff countless walk-throughs, co-walks, coaching and providing effective and meaningful feedback on implementing the framework.

A representative from TeachScape provided training for administrators. TeachScape is an online platform that assists educators in developing skills in observing, collecting and analyzing evidence of practice. Each teacher follows the following process:

- Pre-Observation
- Formal Observation
- Review of data observed
- Post-Observation
- Self-Assessment

Charlotte Danielson's Framework is a research-based set of components of instruction grounded in a constructivist view of teaching and learning. In this framework, the teaching is divided into 22 components, 76 smaller elements and clustered into 4 domains of teaching:

- Planning and Preparation (Domain 1)
- Classroom Environment (Domain 2)
- Instruction (Domain 3)
- Professional Responsibilities (Domain 4)

The framework defines levels of performance that are intended to define the teaching, not the educator. They are not intended to be used as checklists; rather, the purpose of the rubrics is to give educators specific and meaningful feedback to improve practice through observations and self-reflection in each domain. The levels of performance range from distinguished to unsatisfactory in each domain rubric. The levels of performance are defined as follows:

- **Distinguished (4.0)** performance and behavior is clearly and consistently outstanding and exceptional.
- **Proficient (3.0)** performance and behavior is of high quality, exemplary, commendable, skillful, and dependable.
- **Basic (2.0)** performance and/or behavior at a minimal competency level and in need of improvement.
- **Unsatisfactory (1.0)** performance is poor and/or behavior is clearly below acceptable levels.

The training provided to the administrative team was about 4 hours. In my opinion this was not sufficient. The online system was not user-friendly. As an administrator trying to navigate through the TeachScape system of walk-throughs and formal evaluations the process became quite complicated due to the lack of training. Although we practiced inputting fictitious information after creating accounts, the system looked totally different when my team and I began actual co-walks. The school did not purchase *Learn*, which is a part of the TeachScape system. *Learn* is a learning management system that provides preK-12 teachers professional videos that will assist them in areas where they

score an unsatisfactory or basic rating.

Surface tablets were purchased for administrators to use as they entered each classroom. Often there was no internet connection in rooms to complete the walk-through and provide immediate feedback. This raised concern for me. I began to utilize my personal cellular device to log onto the TeachScape webpage to complete walk-throughs or evaluations.

The first year with this system was extremely complicated. There were less challenges the second year as I found ways to assist teachers with practicing each of the Domains, thinking with the end in mind. I coached them on taking a Domain and focusing on that Domain and all the attributes that accompanied it. This model was a success for my staff. Teachers began taking responsibility for growing professionally. They also began implementing practices that improved student growth as well as provided students with the opportunity to take ownership of their learning and others in the class. I saw teachers become better with the process and some were used as “experts” in some Domains.

The Process

While the evaluation instrument does not require approval from the School Board, the evaluation process was presented at a School Board meeting, which was open to students, parents, and the public for informational purposes. Within the evaluation instrument, there were clearly-defined criteria or "look-fors" and “listen fors” that will determine each teacher's rating as it relates to the components of Danielson's Domain 2 -- Classroom Environment and Domain 3 -- Instruction. The tentative timeline for evaluation exercises included classroom walkthroughs throughout the school year. The first formal observation occurred in January 2014, the second formal observation taking place in May 2014, and final evaluation meetings taking place in June 2014. The development of the new teacher evaluation tool also included an embedded professional learning library, provided in each teacher's TeachScape account. It consisted of videos and courses focused on Danielson's Domain

Components during the second year of implementation. Teachers who were performing at low levels in any of the domain components were assigned professional learning activities, and also received direct mentoring from an instructional coach, as well as the building principal, as part of their support and continued growth as professionals.

In addition to the process above the following are additional processes in place for the evaluation system. Professional development, coaching and mentoring are assigned to staff in areas of concern (basic/2.0 and unsatisfactory/1.0 ratings). Teachers receiving high ratings are used as mentors in their high rating areas. Staff meetings may be used to review the framework as well as peer coaches. At times, a teacher may model a lesson, part of the staff may act as students and others act as the evaluators. Once complete we discuss scoring and provide evidence "why" a particular score was chosen. This process seemed to help teachers become better at implementing a particular Domain. Teachers appreciate the fact they could see the Domain and the attributes in action and then score what was observed.

An issue that arises as a barrier is implementation with fidelity. This simply means that for effective implementation we must be intentional in our process that we put in place. Barriers include but may not be limited to providing access to an electronic device for response for feedback, no internet use, teachers and administrators not being open-minded about the process, mind-set of staff including administration, and definitely the lack of training for administrators to familiarize themselves with the process of the framework to ensure effective and quality training and support for teachers.

Feedback is critical to the teacher evaluation process. Teachers must receive meaningful feedback about their performance so that relevant professional development can be provided. Upon the

completion of the evaluation, teachers participate in a Post Observation. Prior to the scheduled meeting, teachers would have completed a Self-Evaluation where they score themselves and answer a series of questions related to the evaluation. This meeting is critical to the planning for next steps for teachers. In my opinion this meeting provides honest, accurate and sometimes data/information that some may not want to hear. However, they do understand the need for this crucial conversation.

As I reviewed the evaluation tool, and considered highly effective teachers versus low performing staff, I found that there was a direct relationship between student growth and highly effective teachers that have outstanding evaluations. Low performing teachers that lack classroom management and are unable to build relationships, and who lack procedures, routines and who fail to actively engage students are more likely to score poorly on local district assessments.

Public Act 173

Public Act 1973 requires that the board of a school district, Intermediate School District or the board of directors of a Public School Academy adopt a rigorous, transparent and fair evaluation system for teachers and administrators. For that reason, we have adopted the Charlotte Danielson's Framework for Teaching. Teaches are evaluated annually and student growth is also a significant component for teacher effectiveness. Some schools use common assessments implementing a pre/post assessment. Again, with pre/post assessments you are able to measure growth from the beginning of a unit to the end of a unit. However, these are teacher made assessments and one cannot prove validity or reliability. In short, I do not believe that Mstep should be the sole method of measuring student growth to the degree that it effects a teacher's evaluation due to the lack of performance on student's part. Assessments should not be the determining factor a teacher's evaluation. I believe that peer

reviews/observations may be an effective method used in schools. My teachers conduct peer observations to improve their own learning. Once a peer review is conducted I discuss the findings with the teacher. Some teachers ask colleagues to visit their classrooms to assist them in the implementation of Charlotte Danielson's framework. Another possible method I believe may be effective is teachers' self-evaluation. A rubric as well as standards should be developed with this process prior to implementing a system.

In closing, implementing a teacher evaluation program with fidelity, providing meaningful teacher feedback, collaborating with teachers, and providing differentiated support will definitely increase student achievement. When results from teacher evaluations are obtained they provide data to improve instruction which will in turn improve student achievement. As teachers improve instructional practices through observations, coaching and professional development, student achievement will increase. If inaccurate data or feedback is provided to a teacher he/she will not grow professionally, which will hinder student learning.

Focus Group Documents

This section includes the documents developed by the Ray.Taylor and Associates team to implement the focus group sessions.

Invitation to Focus Group Participants

Description of Focus Group Sessions

Focus Group Discussion Prompt Questions Focus Group Agenda

Focus Group Feedback Survey

September 30, 2017

Description of the MDE Educator Evaluation Project: Focus Group Component

The Michigan Department of Education has awarded a contract to Ray.Taylor and Associates (RTA) to conduct an evaluation of the professional evaluation system in place in Michigan school districts. The project design includes focus groups of educators impacted by the evaluation system. The process is designed to address three evaluation major activities defined by the MDE.

The three questions of interest are:
Activity 1: Identify best practices in educator evaluation implementation to inform ongoing and future MDE supports to, and programming by, districts.

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.

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

The focus group findings will be used by RTA to inform the overall evaluation project, to help design follow up interviews, and to serve as a resource for analysis of MDE documents and artifacts.
Focus group participants will receive an honorarium for their involvement.

Background:

On January 5, 2017, MDE engaged teams of evaluators to complete evaluations of its system of professional educator evaluation. Final reports will be submitted by September 30, 2017. Additional components of the full evaluation design include white papers by selected researchers and practitioners, document and artifact review, literature review, and targeted interviews and educator surveys. The evaluation will also identify and validate evidence of schools / districts exhibiting best practices among Michigan schools.

Timeline:

Focus group sessions:

Kent ISD – 2930 Knapp St. N.W., Grand Rapids MI 49525 - Thursday, April 20, 9 a.m. until noon
COOR ISD – 11051 N. Cut Road, Roscommon 48653 - Tuesday, April 25, 9 a.m. until noon
Wayne RESA – 33500 Van Born Road, Wayne 48184 - Thursday, April 27, 9 a.m. until noon
Marquette-Alger RESA – 321 E. Ohio, Marquette 49855 – Tuesday, May 9, 9 a.m. until noon

September 30, 2017 Final evaluation report submitted to MDE

Questions or further information contact:

Rossi Ray-Taylor, PhD: *Evaluation project contractor / principal investigator*
rossi@raytaylorandassoc.org
(P) 734 975 1963
President and CEO
Ray.Taylor and Associates
Ann Arbor MI, 48104

Questions to Prompt Discussion

Best Practice: What are examples of best practice consistently implemented in your experience? What factors have made this possible? What are examples of best practice that are aspired to but not yet attained? Why not? Are evaluators / evaluatees well prepared for their roles in the process? Why / why not?

Barriers: What are significant barriers you have experienced in implementing your evaluation system with fidelity? What resources are needed to fulfill with integrity all of the expectations of the evaluation system you are using? How could you overcome these barriers to improve the system?

Student Growth Measures: How do you use student growth measures in your evaluations? What standards do you use for student growth measures in areas not assessed by the state? What alternative assessments are you using? How is it working?

September 30, 2017

MDE Evaluation Focus Group
April 20, 2017
Kent County ISD

AGENDA

* * * Participants, please sign in * * *

Welcome

Purpose

Gather information based on experiences of teachers and administrators to assist in evaluating the effectiveness of the educator evaluation systems in Michigan school districts.

Outcome

Shared narratives, themes, examples, personal experiences related to educator evaluation

Participation Expectations

Everyone speaks – one conversation
Stick to the topic – AND think outside the boundaries
eManners – silence cell phone, no texting, emailing, checking in
Frank and professional interactions
Confidentiality – nothing identifiable by name or district, recording for RTA's notes only
ELMO
Clock – start and end at appointed times, mid-morning break

Who is here

Teachers, principals, administrators

Focus Topics: Educator Evaluation in Michigan

- **Best Practice**
- **Barriers**
- **Student Growth Measures**

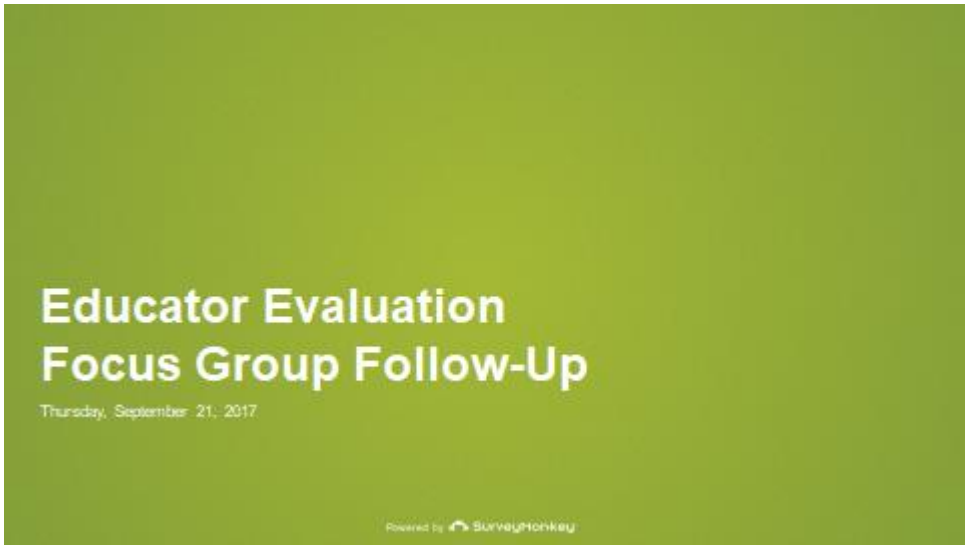
Summary and Final Comments

Feedback

Gogebic-Ontonagon Intermediate School District
LEADING • SUPPORTING • ENRICHING
.....Our Schools and Community

Ray·Taylor AND ASSOCIATES

MICHIGAN Department of Education



41

Total Responses

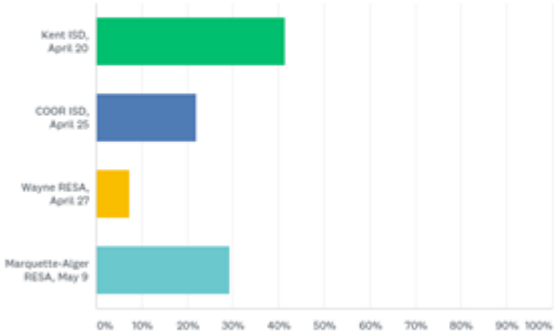
Date Created: Tuesday, April 18, 2017

Complete Responses: 41

Powered by SurveyMonkey

Q1: Please identify the Educator Evaluation Focus Group that you attended

Answered: 41 Skipped: 0



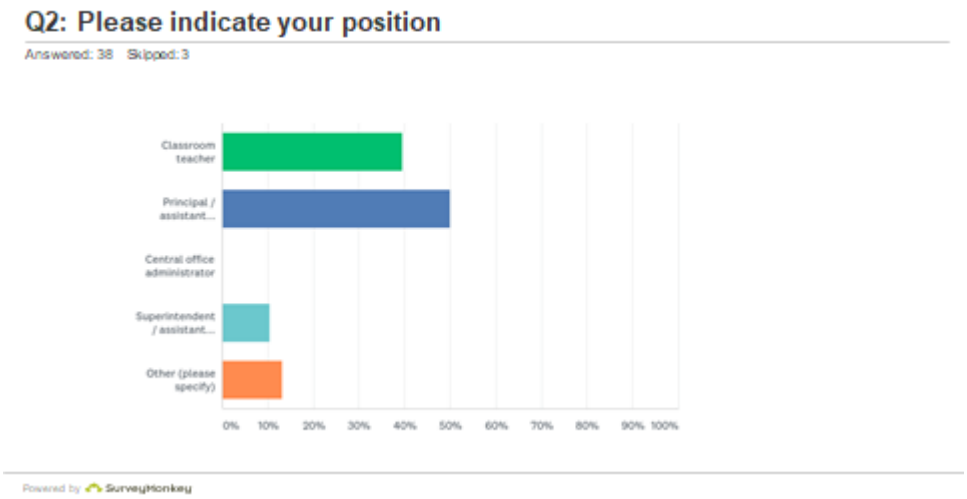
Powered by SurveyMonkey

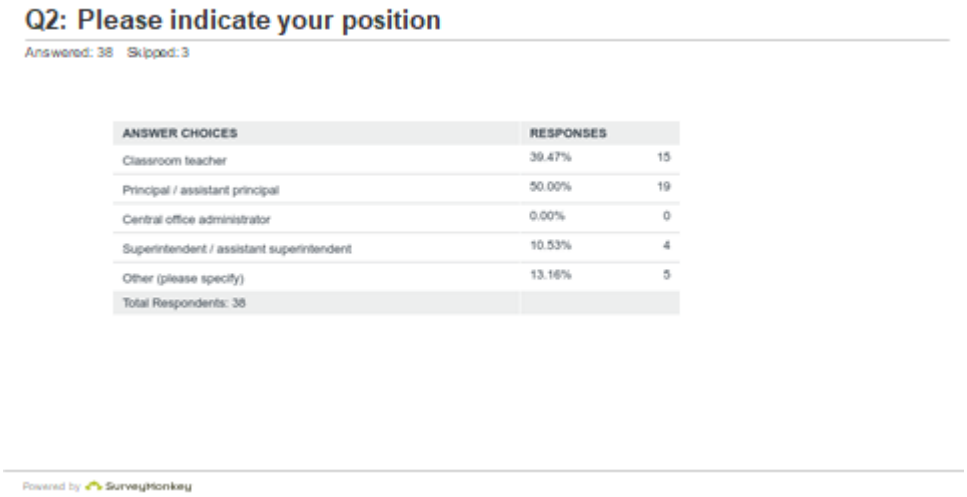
Q1: Please identify the Educator Evaluation Focus Group that you attended

Answered: 41 Skipped: 0

ANSWER CHOICES	RESPONSES	
Kent ISD, April 20	41.46%	17
COOR ISD, April 25	21.95%	9
Wayne RESA, April 27	7.32%	3
Marquette-Alger RESA, May 9	29.27%	12
TOTAL		41

Powered by SurveyMonkey





Unedited Responses to Focus Group Feedback Survey Questions 3-6

Q3 Decisions about the system to evaluate educators are made at the state and local levels. What are your recommendations to decision makers to improve the educator evaluation system? N = 40

1 I think there needs to be more local control. I also think that politicians should be listening more to educators. This shouldn't be run as a business model.

2 - Include Educators in process or retiring the current system - Remove categories Min effective, effective, etc. Set target and allow for growth of profession. - Move growth to be a part of expectations but not a weighted. Use this data to inform instruction which should be reflective in state assessments

3 - Not use Data (Student growth) as a separate requirement - Not use Labels- Let us use growth models, not evaluations models

4 There needs to be a system that focuses on educators growth and helping to develop our craft.

5 Reduce the ranking to two or 3 remove student growth or incorp. into Rather than a separate piece

6 Need more reflection time. rethink labels.

7 More Teacher input during legislative decisions.

8 Implement non-standardized testing measures

9 Embed students growth- not separate it weight planning, reflection, professional responsibilities coach instruction goal setting- plans- activities vs rubrics

10 Do not raise the growth %.

11 More local control. Listen to the people in education to make the decisions

12 Eliminate student growth and the number of evals required by effective + highly effective teachers

13 -Eliminate the Labels -Promote growth w/ Our Staff

4 Do not increase percentage of student achievement until multiple means (Measures) are consistent for all teachers. which can't be done. Because- changes in state assessment, difference in what teachers teach- from core areas to PE, Music, Technology, etc.

15 Allow administrators to make decisions regarding which teachers on staff are coachable & those who are not. Most will be!! Those that aren't shouldn't be working with our children.

16 trust locals allow time to work with teachers who need help not high will/skill

17 Increase local control; Solicit feedback from educators themselves

18 Less Evals lower state testing % 5/15/2017

19 - less evals for tenured teachers - redesign student growth element to be more individualized

20 Identify 1 test system for each of the 4 covers @ all grade -levels

21 Keep it consistent to reinforce relevance

22 To look at the/and change student growth measurements

23 Less state control and more focus on student interest

24 Allow local districts flexibility. Best practices should be reflected in a teacher evaluation system.

25 - Focus on etc - Eval to measure pbl

26 Please listen to the educator stake holder groups when making decisions.

27 I feel there needs to be more control at the district level with the evaluations. Also, an effective teacher should not have to be evaluated every year. Our resources and time should be spent on these teachers that need the support.

28 Make them more realistic, more flexible for individual districts, teachers are critical enough of themselves without feeling like their flaws are being pointed out more

29 remove the "Highly Effective" level + the "Exceeded" label for student growth

30 Focus More on walk throughs, informal classroom visits, and frequent drop-ins rather than one formal evaluation.

31 Maximizing flexibility. Clarify the role it will play - if phased in (as recommended by all models)- how can it be comparable

32 The current system is top down. A system that creates an incentive for teachers to a high bar and take more risks will do more to improve learning than a top-down eval system. More local control please!

33 My recommendation is to keep actively involving educators (both teachers & administrators) working collaboratively on impacting decisions made at the state level regarding the field in which we are professionals.

34 Include educators and administrators in the decisions (Not top down)

35 The Local and regional (ISD) leadership is very much in touch with the needs of evaluation. I would recommend paying close attention to their expertise.

36 Student growth- changing to 40% is too much.

37 More flexibility at the local level is needed.

38 Create criteria for state and local data that can be used for student growth

39 Rotational evaluations...Not all teachers need to evaluated formally every year. Also, remove the highly effective determination.

40 Involve teachers in the decision making process. It seems like no one ever gets their opinion when making decisions even at the local level.

Q4 Based on your experience what are the barriers to implementation of your district's educator evaluation system? N = 41

1 I think there is a problem. I also think administrators need more knowledge to help teachers grow.

2 There is only one way to measure all. Kids learn and grow at different ways + times. Allow for flexibility to meet all student needs.

3 Valid + Reliable data amount of documentation required for accountability/compliance 4 Time

5 Time!

6 - Assessment knowledge/Training as we tackle SLO's - Training for Teachers

7 Time- for PD, Conversations. etc quality data

8 Consistency + understanding of what is expected.

9 Time

10 Student growth instructional vs evaluation coaching

11 Lack of time. 5/15/2017 12:44 PM

12 Time, trust, political changes, training

13 Time. More staff is needed to adequately evaluate teachers.

14 - Student Growth Measures - More time needed to coach Staff W/ Feedback

15 Time for Administrators + Teachers- Balance of other responsibilities, reports (SI) (ED Yes) (Grants) IEP, Department meeting – Title 1, Reading etc. Models are designed for growth- not as evaluation tool- a number

16 Time: to Coach + Develop teachers for teaching/ learning. "Labels are limiting."

17 Time. Is it growth or is it got-ya! Trust

18 Time, training, Applicability

19 No onus for students

20 Time, cost, the human element

21 State Testing

22 Teacher fear of eval

23 Time/Training

24 Time

25 We construct 4-5 goals on a rubric that includes too many performance makers.

26 Lack of training for teachers

27 Time

28 There is not enough time for teachers and administrators to meet and accomplish all that needs to be done

29 Human error or change from one evaluator to another. Time commitment of teachers and evaluators.

30 Time + trust

31 -time effectiveness

32 Time Avalanche of competing initiatives demanding attention and PD time

33 Trust, Time.

34 Time and understanding the process at a deeper level.

35 initiative fatigue, training, resources, and time

36 1. Trust 2. Too much focus on test scores and punitive.

37 Time- we keep adding more + more to admin's job- but there is nothing taken away!!

38 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 in well.

39 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 efficacy.

40 Time, teacher training, negative cultural and political climate

41 How to fairly measure student growth. At the secondary level there isn't a fair way to track state testing back to an individual teacher and locally developed assessments are not valid and reliable. There isn't enough time for teachers to prepare lessons and class sizes are too large. Principals are assigned too many people to evaluate.

Q5 Schools and districts are required to consider measures of student growth in educator evaluation. In your experience how effective is student growth data in educator evaluations? N = 41

1 I think that there are too many variables with students that affect their learning. I think that don't want to be evaluated on students who are rarely in my class. I want to help students grow as much as they can, but I also think that you have to look at where they start.

2 This is not an effective measure. The focus is on ELA, Math, Science + Social Studies. This is difficult in other areas-- art, PE, Language, Etc. This is also difficult for special populations.

3 Not effective- We don't have good/reliable data

4 Student growth should not be a large component of teacher evaluations. The assessments used to measure student growth should be used as a tool to guide a teacher's instruction.

5 Not very Effective as growth is Not Defined.

6 Very Beginning of implementation, but I think it will get better

7 If it is quality data it can help drive instruction. The staff needs time to look at the data multiple time in a year. It is only one time a year, it is not as helpful.

8 Should only be used to dictate teaching methods, not teacher effectiveness. Teachers need it to help their teaching.

9 Not effective or reflects back to on future teaching

10 ineffective. too much variation across areas and levels

11 Little reliability validity. Becomes a barrier to an effective evaluation system when weighted separately.

12 Not all students test well. There should be multiple measures that applicable to the education classroom.

13 Student learning includes too many variables to be effectively included in evaluations

14 - If I was "Evaluating" It - It would be marked ineffective.

15 M-Step + MEAP not effective since it has changed. Multiple measures are important but districts need flexibility- It is hard for non-core teachers compared to core teachers to establish growth measures using consistent, reliable date.

16 I Sign "evaluation" Should Not evaluate/ Rank a very diverse populations of students.

17 It isn't

18 Not effective at all. We can't have Standardized tests in CTE to use!

19 Students data growth is more accurate than state tests

20 standardized tests are not effective

21 - NWEA is Relatively decent. - State testing is all over & highlighted by ambiguity. -Pre/post tests are Effective (Outside of box) States needs a Pre/Post

22 Not at all

23 Not very effective- Schools are not using student data to really measure what they know based on the evaluation

24 Ineffective

25 It is necessary, but I think teachers should be trusted to develop their own measures specific to students' academic level and content understandings.

26 Lessen student growth Measures

27 It's effective but a work in progress.

28 I feel it is important to show student growth, however, it is also important to look at each student's unique situation. Sometimes, we cannot always expect the same for each student.

29 You cannot quantify the measure teachers have on students and student growth.

30 Important but remove the "Exceeded" label. You either met it or you didn't.

31 It is only one of many, many aspects that determine effectiveness 5/15/2017

32 NWEA does show growth and does so in a relative/comparable manner SLO's do not, too loose!!

33 It is limited to the data used and how that data is used.

34 I think this is the tricky part right now. So many schools are using such a variety of methods for measuring growth, that it's hard to compare achievements.

35 not effective; let us worry about teaching not the amount of growth happening

36 Minimally effective, because the growth data we use keeps changing.

37 I understand the need to show growth- but there are so many other things teachers do that make positive impact.

38 Somewhat effective. There are numerous variables that factor into student growth data that are not able to be measured

39 Growth data must be described in different modes than achievement data. Growth in skills should be emphasized in SLOs, drawn from local student outcomes.

40 from the state level it is very ineffective. Data should be considered using local measures. Classroom data and how the data is used to drive instruction to meet the need of all learners is much more important and reflective of a teacher's performance than the results of a once a year test.

41 Not effective. They are being measured by poorly written assessments. Districts are not using pretests to look at growth of individual students. They are using it to compare teachers.

Q6 Please add any additional comments. N = 32

1 Thank you for getting the outlook of educators. Too often our voices and experiences are overlooked.

2 I appreciate the format and to hear others thoughtful perspectives. There was consensus amongst the group which spoke volumes. There is a solid understanding of better practices but that knowledge is not applied to adults. Let folks know about honorarium as they can check school policy and come prepared

3 I really appreciate how streamlined this was! I would love to have even more time to dialogue with the people in the room to learn about how others are navigating the evaluation systems within their buildings- Love ELMO !! :) Great Questions - Appreciate the facilitation of this meeting.

4 It was a good conversation. A lot of great insight and experiences

5 Well Run Organized focus group- It could have easily been 5 or 6 hours.

6 Session feedback- Ideas- Allow for reflection time when you give a topic- have them write - then share that way if you run out of time you can collect those & #'s 3-5 you have data and our feedback is we didn't get time/opportunity to share!

7 he Session was very orderly and ran smooth. It was nice to hear others ideas, concerns and issues. the ELMO was a great idea.

8 This session was very insightful. It shows that many people are feeling the same way about the education models.

9 session transitioned well from subject to subject - Adding a web based component

10 More of these sessions would be great. Maybe look at having smaller groups talk, record and then share out, instead of one at a time.

11 Great facilitators. I appreciate the thoughtful dialogue.

12 Would be interested in participating in a follow-up "brainstorm" on the solutions since we have identified the problem.

13 The Session was really well facilitated. Excellent ability to get all voices in the room. Great job. 5/15/2017 12:40 PM

14 I was asked to attend by our ISD. I was not looking forward to this focus group Bc of all the (Testing/School Improvement/Eval) work that needs to be completed. With that said, This group was insightful & allowed me time to reflect on what we do well & what we need to improve on.

15 Session was: * Nicely organized, valued thoughts + input from all *Thank you for setting time limits + keeping to your Agenda - Elmo Great! I would love to see results from your other focus groups, researchers, and what will be shared with MDE. Thanks!

16 Very organized. Respectful + powerful. Thank you for providing this forum for input. Hoping positive change will take place.

17 All we want to grow as a system we have to be able to trust our people. The ways things were intended + how they were implemented are balanced.

18 Thank you! Very good workshop well- structured

19 None

20 =)

21 The States should do a pre-post fall students every year- potentially a mid-term. 22 Today was Great, Facilitators did a great Job Summarizing & probing questions 23 Thank You! I love the small group format!

24 the discussion was wonderful + it was led very well.

25 Thank You.

26 I believe these changes would allow administrators to have ongoing meaningful conversations/ evaluations that focus on improvement rather than a label that unequally impacts their perception and creates morale issues. (Below) Thank you for taking the time to come to the U.P.! WELL organized and presented.

27 The pressures of highly effective and minuscule and variable nature of the models make layoffs based on point differential RISKY

28 Thank you for this opportunity.

29 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.

30 Thank you for letting me have a voice!

31 Thank you for the invitation.

32 I was disappointed more teachers were not involved in the focus group. I think its important to hear their voice also.

Statewide Survey Documents

This section includes the statewide survey distribution memo, the methods used for analyzing the survey results, and the survey developed by the Ray.Taylor and Associates team.

MDE would like your input.

Dear Michigan Educator,

The Michigan Department of Education is in the process of assessing the educator evaluation system in the state of Michigan. The intent of this effort is to provide research-based and field-based information to support a continuous improvement orientation to implementation of PA 173 of 2015.

Ray.Taylor and Associates (RTA) is one of the firms contracted to assist with this initiative.

Your input is an important part of this effort so we are asking for your insight and recommendations through the survey found at: https://www.surveymonkey.com/r/V6ZSJBS. We estimate that the survey will take approximately 15-20 minutes to complete.

If you have questions or need additional information, please contact Rossi Ray-Taylor, project director, at 734-975-1963, or email rossi@raytaylorandassoc.org. A description of Ray.Taylor and Associates' full project plan is attached.

We thank you for your input and assistance with this important effort.

Regards,

Rossi Ray-Taylor
President and CEO, Ray.Taylor and Associates

Joshua Roesner
Educator Talent Analyst
Office of Educator Talent @MDE

Roesnerj1@michigan.gov
517-335-5236

Description of the MDE Educator Evaluation Project: Survey Component

The Michigan Department of Education has awarded a contract to Ray.Taylor and Associates (RTA) to conduct an evaluation of the professional evaluation system in place in Michigan school districts. The project design includes several features including a survey of Michigan K-12 educators. The survey is designed to address three evaluation major activities defined by the MDE.

The three questions of interest are:

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

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.

Activity 3: Evaluate the measurement of student growth using alternative methods/tools and processes specified in PA 173 of 2015.
The survey findings will be used by RTA to inform the overall evaluation project, along with findings from focus groups, targeted interviews, and to serve as a resource for analysis of MDE documents and artifacts.

Background:

On January 5, 2017, MDE engaged teams of evaluators to complete evaluations of its system of professional educator evaluation. Final reports will be submitted by September 30, 2017. Additional components of the full evaluation design include white papers by selected researchers and practitioners, document and artifact review, literature review, and targeted interviews and focus groups. The evaluation will also identify and validate evidence of schools / districts exhibiting best practices among Michigan schools.

Timeline:

Final evaluation report submitted to MDE September 30, 2017

Questions or further information contact:

Rossi Ray-Taylor, PhD: Evaluation project contractor / principal investigator
rossi@raytaylorandassoc.org
(P) 734 975 1963
President and CEO
Ray.Taylor and Associates
Ann Arbor MI, 48104

See what's happening on our social sites

MDE Survey Analyses

Methodology

The following details the survey data analysis process:

- Data downloaded from Survey Monkey

- Data imported into Statistical Package for the Social Sciences then cross validated with Survey Monkey data
    - Currently the analysis consists of 9000 records with 317 variables
    - Teacher respondents 8404, administrator respondents 596

- Computation of multi response variables as directed by the principal investigator
    - Example Student Growth Measures both teachers and administrators
- Reconciliation of school district variable for both teachers and administrators in respondent file

    - Recode respondent school districts to match DOE district codes. Merge district information from DOE back to respondent file.

- Cross tabulation of variables related to 1) best practices in evaluation, 2) evaluation methods, 3) barriers to implementation of best practices, and 4) flagging cells in tables with observed values significantly different from expected values for tables with non-random cell values (adjusted standardized residuals)

- Computed scale mean variables for two sets of items related to best practices for teachers

- Created clustered bar charts for contingency tables with table cell percentages SPSS

- Recreating stacked bar charts for table row percentages as directed by the principal investigator

Best Practice Documents

This section includes the documents developed to identify and collect information about implementation of best practices for education evaluation in Michigan. Profiles of the best practices identified are found at the end of this section.

Best Practice Nomination Request

Best Practice Nomination Form

Follow-up Best Practice Interview Questions

Full Description of ***Best Practices in Action***

Description of the MDE Educator Evaluation Project: Best Practice Validation Component

The Michigan Department of Education has awarded a contract to Ray.Taylor and Associates (RTA) to conduct an evaluation of the professional evaluation system in place in Michigan school districts. The project design includes several features including identification of districts implementing best practices for educator evaluation in Michigan.

The three questions of interest to the MDE review are:

Activity 1: Identify best practices in educator evaluation implementation to inform ongoing and future MDE supports to, and programming by, districts.
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.
Activity 3: Evaluate the measurement of student growth using alternative methods/tools and processes specified in PA 173 of 2015.

The schools and districts identified as implementing best practices will be used by RTA to inform the overall evaluation project, along with findings from focus groups, surveys, targeted interviews, and to serve as a resource for analysis of MDE documents and artifacts.

Background:

On January 5, 2017, MDE engaged teams of evaluators to complete evaluations of its system of professional educator evaluation. Final reports will be submitted by September 30, 2017. Additional components of the full evaluation design include white papers by selected researchers and practitioners, document and artifact review, literature review, statewide survey of educators, and targeted interviews and focus groups. The evaluation will also identify and validate evidence of schools / districts exhibiting best practices among Michigan schools.

Timeline:

Final evaluation report submitted to MDE September 30, 2017

Questions or further information contact:

Rossi Ray-Taylor, PhD: *Evaluation project contractor / principal investigator*
rossi@raytaylorandassoc.org (P) 734
975 1963
President and CEO Ray.Taylor and
Associates Ann Arbor MI, 48104

The MDE is identifying schools and districts that serve as examples of implementing best practices in educator evaluation in Michigan. To nominate your school or district as an example of one or more of the best practices for implementation of educator evaluation, please complete this brief form. Feel free to add information as needed to describe your program.Your example will help improve the practice of educator evaluation statewide.

* 1. Name of school

* 2. Grade levels served by school

* 3. Nominating principal: Name and contact information - email, telephone

* 4. School district

* 5. Nominating superintendent: Name and contact information - email, telephone

6. How are evaluators prepared to implement evaluation? Check all that apply.

- [ ] Training by district staff
- [ ] Training by vendors / publishers
- [ ] Training by ISD staff
- [ ] Training by consultants
- [ ] Video or electronic training tools e.g., webinar, TeachScape
- [ ] Other (please specify)

7. Approximately how much time is spent annually training to prepare evaluators to conduct evaluations?

- ◯ Less than 2 hours
- ◯ Half day (three hours)
- ◯ Full day (6 hours)
- ◯ More than 6 hours

Comment

8. How are those being evaluated prepared to participate in the process? Check all that apply.

- ☐ Training by district staff
- ☐ Training by vendors / publishers
- ☐ Training by ISD staff
- ☐ Training by consultants
- ☐ Video or electronic training tools e.g., webinar, TeachScape

Other (please specify)

9. Approximately how much time is spent annually to prepare those being evaluated?

- ◯ Less than 2 hours
- ◯ Half day (3 hours)
- ◯ Full Day (6 hours)
- ◯ More than 6 hours

Comment

10. Does your district provide the following: Check all that apply

- ◯ Evaluation Handbook or Guidelines Manual (developed by district or ISD / RESA)
- ◯ Board adopted policy regarding educator evaluation

Other tools for public transparency and evaluation training (please specify)

11. Which observation tool is currently used in your school? Please check all that apply.

- [ ] Charlotte Danielson's Framework for Teaching
- [ ] the Marzano Teacher Evaluation Model
- [ ] the Thoughtful Classroom
- [ ] the 5 Dimensions of Teaching and Learning
- [ ] Other (please specify)

12. Which evaluation tool does your district use for administrators? Please check all that apply.

- ◯ MASA's School Advance Administrator Evaluation Instrument
- ◯ Reeves Leadership Performance Rubric
- ◯ Other (please specify)

13. From the list below please describe the area(s) of best practices for which your school / district should be recognized and provide a brief description of your practices. You may describe more than one best practice used in your school / district for which it should be recognized.

Professional development and mentoring for teachers and administrators that is aligned with their individual educator areas / results

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

Administrator evaluation in general. and specifically for school based administrators as compared to district-level administrator

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

14. Please describe the area(s) of measurement of student growth used in your school / district that represent use of best practice and provide a brief description of your practices. You may describe more than one area in which you implement best practice in applying measures of student growth. Please provide a description for each area for which your school / district should be recognized.

Student Learning Objectives (SLOs)

M-Step / MiAccess

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

15. Describe how implementation of best practices for educator evaluation has impacted learning outcomes for educators and students in your school / district.

16. Please describe in general terms how your district and schools are implementing best practices in the educator evaluation system.

Thank you for this nomination of best practices in educator evaluation. Please submit your nomination by July 7. Our evaluation team will review the nomination and will contact you within 2 weeks of receipt of your application.

Format and Questions for Phone Interview:

- Introductions and brief overview of MDE Educator Evaluation project
- Why did you choose to nominate your district as an example of best practice in educator evaluation?
- Focusing on your response to our invitation to nominate:
    - How long has this practice been in place?
    - What have been the results/outcomes of this practice?
    - What are your planned next steps related to this practice?
    - What are your plans for sustainability of this practice over time?
- What significant barrier(s) have you overcome in developing this best practice? How?
- How are you addressing the requirement for student growth indicators in your process?
- What else would you like us to know about your district's best practices in educator evaluation?

Thank you again for sharing your positive efforts.

The RTA Team

September 30, 2017

Best Practices in Action

District/School: Newaygo County RESA

Teacher Observation/Evaluation Tool: Thoughtful Classroom, Marzano

Best Practice Categories: Training, Feedback, Inter-Rater Reliability

Description of the best practice:

Newaygo County Public Schools emphasizes the use of its evaluation tools for purposes of staff growth and support. All administrators at four designated schools have participated in Leverage Leadership Training, focused on providing timely and effective feedback. To promote a growth mind set, administrators conduct multiple classroom walkthroughs. Local superintendents open their buildings to allow administrators across the county to practice classroom observations and give feedback to teachers. Different administrators have an opportunity to observe the same teacher, review their experience together, and develop consistent content for feedback. Teachers benefit from the timely feedback, which is provided outside of their evaluation process and builds a culture of trust and rapport. The training and implementation, while laying a foundation for an evaluation process, emphasizes the administrative role in instructional leadership that promotes growth and improvement in the practice of teaching. The process provides a safe environment for both administrators and teachers to learn together. Response has been positive from administrators, who have become more comfortable and eager to get into classrooms and engage in the process.

Implementation of Leverage Leadership has been in place in Newaygo County for three years. It was initially supported by a grant from MDE, and included working with a knowledgeable trainer, an important factor for its success. A five-day training program is followed by two half-days throughout the year, designated for group walkthroughs. A financial commitment is required for ongoing implementation, which can be an inhibiting factor for some districts with limited resources.

Successful implementation of this practice at a county level requires having a positive relationship with local districts. Where there is a lack of common calendar, the opportunity is available for administrators in one district to visit classrooms of a school in another that is in session. Contract language in individual districts can be a challenge if there are restrictions on outside visitors to classrooms.

This practice is sustained by Newaygo County schools by establishing a school visit calendar where half-days are reserved for cross-district classroom visits. One of the hosting schools provides breakfast and then administrators go to designated schools and classrooms for their observations. Administrators are expected to provide feedback to teachers within 24 hours of the walk-throughs. Next steps are planned to involve teacher leaders in the Leverage Leadership training. There is a planning committee that includes teachers and principals.

Contact Person: Nicole Gasper, Chief Instructional Officer

Best Practices in Action

District/School: Harper Woods Public Schools, Harper Woods Middle School

Teacher Observation/Evaluation Tool: Danielson

Best Practice Categories: Training, Collaboration

Description of the best practice:

Harper Woods Public Schools engages in extensive training of observers, including a calibration component. The middle school principal has shown leadership in implementing the district's redesigned evaluation process. The previous system had little focus, consisting of checklists and very broad commentary on teacher performance. In adopting the Danielson model, the district promoted engagement of teaching staff in training and an emphasis on collaborative learning of the process. During the year prior to formal adoption last year, the district provided Danielson training to all district administrators during the first semester, followed by classroom observations and practice with the tool during the second semester. The district uses Frontline, the vendor-provided management system, for continuous training and professional development tied to the evaluation model. Teachers have ready access to data and information, and are becoming more familiar with the framework. The district will send a group of teachers to Danielson training at the ISD this year. The district has seen increased collaboration in the evaluation process, including union support in sharing information with teachers.

Each year, Harper Woods selects a key item within each of the five domains in order to focus coaching conversations and concentrate on specific areas of feedback. Although all domains and elements are addressed, the district narrows its attention to the selected areas for emphasis during a particular year. This process has led to identifying district-wide goals related to the district and building school improvement plans.

As a result of implementation, Harper Woods reports an increase in data-driven dialogue. Buildings look at data to develop action steps, and teachers develop action steps following late-start PD days centered on use of student assessment (NWEA) reports. An additional outcome has been a shift to a coaching model that promotes conversations about effective instruction. The district reports an increase in intentional, authentic conversations by students and teachers focused on growth.

Next steps to ensure sustainability include professional development tied to Frontline. When hiring new teachers, the district considers skills and strengths that match needs indicated by analysis of evaluation data. When assessing student achievement, the district has moved away from pre-post activities to a proficiency model using multiple student growth measures. Plans are underway to publish a testing handbook to clarify testing environment expectations.

Contact Person: Dave Rabbideau, Assistant Superintendent; Heath Filber, Principal

Best Practices in Action

District/School: Grosse Pointe Public Schools

Teacher Observation/Evaluation Tool: Danielson

Best Practice Category: Multiple Pathways, Teacher Rating

Description of the best practice:

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 personal growth and professional practice criteria 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. Teachers receive the same merit pay for either effective or highly effective ratings, so there is no financial incentive or differential linked to HE ratings. This practice requires an awareness of the effort and intensity involved in achieving a rating of "highly effective," as well as acceptance of the high standards of a rating of "effective." In this district, the rating of effective indicates fine teaching performance. In addition to clarifying the different expectations for the two ratings, the process eliminates the anxiety and uncertainty experienced when teachers must wait until the end of the school year to learn which level of rating they will receive. It is known at the beginning of the school year which path will be taken. On average, about 20-25 percent of teachers elect the HE goal.

The Grosse Pointe district uses a holistic approach to educator evaluation. While some districts take a mathematical averaging approach to determine effectiveness, Grosse Pointe uses multiple measures of student growth. Teachers select NWEA and one district-provided measure each year. In an effort to utilize meaningful assessments, pre-post measures are generally avoided. Teachers work with administrators to develop student growth measures if the content area does not have appropriate indicators.

Next steps for the district include finding exemplar video examples of effective teaching to use with administrators to improve inter-rater reliability. The district is also seeking a management system that is congruent with its holistic approach, and not based on traditional checklists and scoring formulas.

Contact Person: Jon Dean, Deputy Superintendent

Best Practices in Action

District/School: Lenawee ISD

Teacher Observation/Evaluation Tool: Danielson

Best Practice Categories: Collaboration, Training

Description of the best practice:

The Lenawee County ISD espouses an evaluation philosophy of shared responsibility, goal-based instruction, ongoing feedback, and common focus. The district emphasizes professional growth and development as the key purpose of evaluation, and promotes an atmosphere of positive energy and excitement rather than anxiety related to the process. Common goals are tied to the evaluation tool for all staff. The district conducted a broad-based review of evaluation tools and standards over time, resulting in district-wide agreements on the tool, process, and procedures. The district used grant money to fund vendor-based training.

A significant practice for LISD is linking staff salary increases to student achievement. All bargaining units include student achievement as a feature of annual performance reviews, and there is a portion of every employee evaluation directed to student growth. LISD reports that this practice, consistent throughout the district, has resulted in a sense of working together regardless of role in the district. The commitment of all staff promotes collaboration at all levels. Bus drivers, teachers, support staff, and other personnel all line up to the same standards and feel that they have a stake in student growth and achievement. This shift to link salary increases to common goals occurred over a period of years, and included open meetings and discussions along the way. LISD observes that there is a continually growing momentum to maintain and sustain this practice. LISD has developed assurances within district policy and collective bargaining agreements to meet legal requirements in the evaluation process.

Contact Person: Dan Garno, Executive Director of Staff Resources

Best Practices in Action

District/School: Reading Community Schools, Reynolds Elementary School

Teacher Observation/Evaluation Tool: Five Dimensions

Best Practice Category: Collaboration, Feedback

Description of the best practice:

The Reading Public School District provides initial administrator training for implementation of the 5D evaluation model through consultants, and then follows up with district training at administrative meetings throughout the year. District staff then conducts training for teachers. The 5D model was adopted as a replacement to a previous one-page evaluation document that was perceived as weak and ineffective in providing direction and specific feedback for improvement. According to the district superintendent, the 5D model is much more powerful. It provides a road map and guidelines for teachers and administrators. The district engaged in training for a full year before full implementation of the new model, which has been in place for four years.

The district reports that best practices in its evaluation process include provision of quality feedback to teachers and administrators throughout the school year. The elementary principal meets regularly with each teacher and creates a collaborative process with each. In particular, the elementary school staff has responded to the leadership of the principal, embraced the 5D model, and generated a productive team atmosphere in its implementation. The staff has been open-minded and receptive to the change, leading to a successful implementation. The superintendent shared his perception that in the elementary school, classroom instruction has improved, student growth measures have increased and discipline referrals have decreased since implementation. The high school is beginning a similar process. The district is working on determining the most appropriate assessments for non-core classes and applying rubrics to special education, counselors, and elective teachers at the secondary level.

Contact Person: Chuck North, Superintendent; Dennis Irelan, Principal

Best Practices in Action

District/School: Dearborn Public Schools

Teacher Observation/Evaluation Tool: Danielson

Best Practice Categories: Feedback, Training, Instructional Focus

Description of the best practice:

The Dearborn school district uses a highly collaborative process to mitigate negative perceptions of evaluation. There has been a shift away from judgment and toward learning and improvement as the primary thrust of evaluation. The district puts significant effort into obtaining buy-in from staff, resulting in a decrease in the apprehension and vulnerability felt by staff being evaluated. The district convenes committees on both teacher and administrator evaluation. The groups meet once a month, agendas are shared electronically, and discussions are open and transparent. The committees are relatively large, allowing for the teacher voice and perspective to be involved in decisions and direction. In addition, district-wide conversations through the evaluation committee structure allows for increased teacher commitment to the rubric and understanding its direct connection to classroom teaching. As teachers interact with the rubric, they see that it really is all about their teaching. Because the rubric is comprehensive, the district recognizes the challenges of the learning required to integrate the various elements of the evaluation system.

In order to increase administrators' confidence and capacity to conduct evaluations with fidelity and provide accurate feedback, the district requires administrators to successfully complete a challenging vendor-provided evaluation assessment. Ongoing calibration exercises are embedded in district administrator professional development throughout the school year. Training is also provided to develop skills in providing accurate, calibrated feedback and initiating the hard conversations that are sometimes necessary in the evaluation process.

Dearborn district leadership reports anecdotal evidence that teachers and principals are reflecting more on their practice, and that the evaluation process now brings greater focus to classroom instructional practices. Next steps will include clarification of student growth measures and how they can be appropriately used in the evaluation process.

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

Best Practices in Action

District/School: Oscoda Public Schools, Oscoda Area High School

Teacher Observation/Evaluation Tool: Danielson

Best Practice Categories: Use of Data, Multiple Measures

Description of the best practice:

The Oscoda district approaches the evaluation process with a strong emphasis on developing and maintaining trust. Trust allows educators to reflect on where they are and make adjustments in practice. The district actively seeks teacher input on teaching assignments, provides immediate feedback following classroom walk-throughs, and has collaboratively examined the Danielson model to focus on the most important elements.

The district has experienced success in meeting with its OEA bargaining unit and reviewing student data. Measures of student growth include locally developed pre and post-tests in every subject, and local and NWEA assessments. The district works with the ISD to determine a simple average growth formula by which to assess student growth. In general, evaluation ratings of teachers and measures of student growth tend to be in alignment. Both teachers and administrators are seeking accurate representation of what the tests reveal, and the district is committed to a continuous process to refine the use of multiple measures of student growth. In the current year, Oscoda will ease into use of student learning objectives as a key element of student growth measures.

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

Best Practices in Action

District/School: Kalamazoo RESA, WoodsEdge Learning Center

Teacher Observation/Evaluation Tool: Five Dimensions

Best Practice Category: Alternative Measures of Student

Growth Description of the best practice:

The WoodsEdge Learning Center serves a center-based population of students with special needs, including autism, cognitive impairment, severe cognitive impairment, and multiple impairments. Over a period of one year, the Center staff worked to adapt the 5D evaluation rubric to its special education population. The 5D tool is not generally geared toward the more restrictive environments in special education. The model for student growth is more often focused on student activity for typically developing students, and is not applicable to the alternative curriculum standards and smaller growth increments experienced with students with significant impairments. The school examined and discussed each rubric, and identified examples of how it could be modified to be more reflective of student success in a center-based classroom. The examples apply primarily to the Proficient and Distinguished categories of the evaluation model.

In the past, the Center used IEP goals and objectives, selecting specific items to track data. However, that approach tended to result in writing and focusing primarily on goals that could easily demonstrate student growth. The current process uses a data portfolio that includes academics, behavior, and all IEP goals. The examples for 5D rubrics that are more appropriate for the WoodsEdge population are more concrete and allow for more meaningful conversations and feedback during the evaluation process.

Next steps for WoodsEdge are to maintain a focus group to work on continuous improvement and sustainability, to continue meeting with teacher groups, to look for additional areas to modify the 5D rubric, and to adjust rubrics to accommodate new curriculum, technology, and teaching devices.

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