Michigan Teachers' and Administrators' Perceptions of
the Teacher Evaluation Process

OCTOBER 31, 2017

Abby Cameron-Standerford, EdD
Derek Anderson, EdD
Bethney Bergh, PhD

JoeyLynn Selling, MA
Sharon Bohjanen, PhD
Anthony Barnes, MA

NORTHERN MICHIGAN UNIVERSITY

THANK YOU

MICHIGAN
Department of Education

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

TABLE OF CONTENTS

1 NMU-MDE TEACHER EVALUATION RESEARCH EXECUTIVE REPORT

Purpose of the Research

The purpose of this research project was to identify barriers to implementation in local educator evaluation systems and to make recommendations for ways to mitigate barriers and inform ongoing and future supports to, and programming by, districts.

Methods

A stratified convenience sample of 1,746 public school educators (teachers; n = 1274 and administrators; n = 474) from Michigan's 56 Intermediate School Districts were surveyed using a researcher-constructed survey tool that included categorical sliding scales and extended response questions. An additional stratified sample 128 teachers and 48 administrators from the original participants were selected for phone or electronic interviews, which consisted of extension questions from the survey.

Demographics

Participant Demographics

- 70% of the participants were female
- 98.6% of the teachers received highly effective or effective ratings in Michigan in 2017
- 20% of the schools compensate teachers for highly effective ratings

Evaluation Tools

- 50% Danielson
- 29% Five Dimensions of Teaching and Learning

Student Growth Measures on Teacher Evaluations

- 22% of schools use NWEA
- 19% of schools use pre/post tests
- 6% of schools use MSTEP

Results / Conclusions

In this section, we present results from the mixed-methods study of 1,746 Michigan educators who completed our questionnaire or interview in response to teacher effectiveness ratings, evaluation tools, and data measures to determine student growth.

What *is* the purpose of teacher evaluation in Michigan?

- **Administrators**
  19% of the administrators cited compliance and 67% of the administrators cited purposes related to improving teaching and learning

- **Teachers**
  30% of the teachers cited compliance and 42% cited purposes related to improving teaching and learning

What *should* be the purpose of teacher evaluation in Michigan?

- **Administrators**
  96% of administrators citing student growth and learning, professional development, improved instruction, or coaching as purposes

- **Teachers**
  97% of teachers cited student growth and learning, professional development, improved instruction, or coaching as purposes

For the quantitative analysis, constructs were created by statistically grouping similar items from the survey to address three broader research questions: (1) Do educators perceive that teacher evaluation processes provide critical feedback and professional development to improve teacher practices? (Teacher Improvement Critical Feedback Construct), (2) Do educators perceive teacher evaluation processes support teacher well-being? (Teacher Well-being Construct), and (3) Do educators perceive teacher evaluation processes support student learning? (Student Learning Construct). These constructs were then used as initial themes in analyzing qualitative data.

Teacher Improvement, Critical Feedback and Professional Development

- Administrators perceived that the teacher evaluation process increased critical feedback significantly more than teachers perceived.

- Administrators overwhelmingly indicated that the educator evaluation process resulted in positive professional growth and that the process drove professional development decisions; although, administrators also noted that the tools were not personalized enough.

> Administrators perceived that the teacher evaluation process increased critical feedback significantly more than teachers perceived.

- The majority of teachers indicated that the system, including the process used, the rubric, and the required data components were not flexible enough to be personalized. Teachers stated that the educator evaluation tools seemed to reflect a focus on the general education setting making it challenging for specials teachers (art, music, physical education, etc.), early childhood/kindergarten teachers, school counselors, vocational/career and technical education teachers, interventionists and special education teachers (categorical, resource room, teacher consultants, RESA program teachers) to receive effective and critical feedback from the evaluation process.

Teacher Well-being

- Administrators perceived that the teacher evaluation process increased teacher well-being significantly more than teachers perceived.

- Teachers perceived
  - Stress, which is negative to teacher well-being
  - Competition, which is negative to collaboration
  - Questions regarding fairness, which impacts job satisfaction and school culture

- Administrators perceived
  - Increase in teacher stress, which can be positive to teacher well-being
  - Increase in teacher collaboration, often due to administrative leadership
  - Increase in competition, which can be positive to teacher well-being
  - Criticism of lazy or oppositional teachers

Student Learning

- Administrators perceived that the teacher evaluation process increased student learning significantly more than teachers perceived.

- Teachers perceived
  - Improved instructional practices, which included research-based strategies, sharing learning outcomes with students, and creating positive learning environments
  - Concern that the evaluation system focused on compliance and resulted in a loss of autonomy

- Administrators perceived
  - Improved instructional practices including consistency across different teachers' classrooms
  - Improved student test scores
  - Unfamiliarity with the teacher evaluation system and an inability to recognize a relationship with student learning

Barriers to Effective Teacher Evaluation as Identified by Teachers and Administrators:

- Lack of time
- Inconsistent or inauthentic process
- Unclear and inadequate evaluation criteria
- Overreliance on standardized test scores
- Teacher resistance and mistrust
- Unfamiliarity with content and/or discipline specific practices (e.g., special education)

Recommendations for Improving the Teacher Evaluation Process

- Administrators need to identify and communicate the purpose of teacher evaluation system in their schools.
  - If the teacher evaluation system is to be used to "inform staffing decisions," principals need to communicate this process explicitly to teachers, adhere to the evaluation outcomes when making staffing decisions, and recognize/address the ramifications on school culture. (See Teacher Well-being Construct.)

- If the teacher evaluation system is to be used to coordinate professional development for the purpose of improving instructional practices, teachers need to be empowered to recognize what is working well in their teaching repertoire and be encouraged to identify individualized professional needs. (See Teacher Improvement Critical Feedback and Professional Development Construct.)

- If the teacher evaluation system is to affect student growth and learning and improve student outcomes, teachers need to be provided time and space to modify instructional practices based on feedback provided through the evaluation process.

- Administrators need to prioritize teacher evaluations within principal duties and responsibilities.
  - Allocate time and streamline the evaluation process.
  - Hire additional staff.

- Administrators need to honor the teacher evaluation process as a way to provide reliable, valid and personalized feedback to impact the professional growth of teacher and student learning.

- Administrators need to perform authentic walkthroughs and classroom visits.
  - Provide a more collaborative process that involves teachers.
  - Provide additional teacher training and resources.

2 PURPOSE OF THE STUDY

In 2016, the Michigan Department of Education (MDE) adopted the ambitious and ambiguous vision of becoming a "top 10 education state in the next 10 years" (Michigan Department of Education, 2016). In order to meet one of the four goals (develop, support, and sustain a high-quality, prepared, and collaborative education workforce) of this vision, the MDE identified "implementing Michigan's educator evaluation law with fidelity [as] a key strategy" (Michigan Department of Education, 2015, p. 4). To justify the recently passed law (PA 173 of 2015) mandating new procedures and requirements for evaluating teachers and administrators, the MDE asserted:

> High quality educator evaluations support both student learning as well as educator well-being. High quality evaluations provide teachers with critical feedback on how they can improve their own practice to impact the lives of students. In addition to facilitating educators' personal pursuits of excellence, systematic improvements to educator evaluations in schools and districts play an essential role in providing targeted professional development responsive to the needs of educators. (p. 4)

The purpose of this study was to determine the extent to which Michigan teachers and administrators perceived the above assertion to be true. The research questions guiding this study included:

1. Do educators - administrators and teachers - perceive that the teacher evaluation process supports student learning?
    a. Do teachers perceive that the educator evaluation process supports student learning?
    b. Do administrators perceive that the educator evaluation process supports student learning?
2. Do educators perceive that the teacher evaluation process supports teacher well-being?
    a. Do teachers perceive that the educator evaluation process supports educator well-being?
    b. Do administrators perceive that the educator evaluation process supports educator well-being?

3. Do educators perceive that the teacher evaluation process provides teachers with critical feedback and professional development on how they can improve their own practice?
    a. Do teachers perceive that the educator evaluation process provides educators with critical feedback on how they can improve their own practice to impact the lives of students?
    b. Do administrators perceive that the educator evaluation process provides educators with critical feedback on how they can improve their own practice to impact the lives of students?
    c. Do teachers perceive that the educator evaluation process plays an essential role in providing targeted professional development responsive to the needs of educators?
    d. Do administrators perceive that the educator evaluation process plays an essential role in providing targeted professional development responsive to the needs of educators?
4. What are educator perceptions of the barriers and solutions to effective teacher evaluation?
    a. What are teacher perceptions of the most common barriers in Michigan districts to implementing high quality educator evaluation systems?
    b. What are administrator perceptions of the most common barriers in Michigan districts to implementing high quality educator evaluation systems?
    c. What are the teachers' recommendations for ways these barriers can be mitigated or eliminated by the district and/or supports provided by MDE?
    d. What are the administrators' recommendations for ways these barriers can be mitigated or eliminated by the district and/or supports provided by MDE?

3 LITERATURE REVIEW
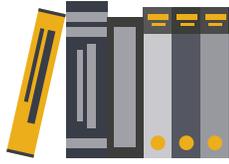
The following section reviews research related to the research questions stated above. Specifically, this chapter reviews research on teacher evaluation historically, teacher evaluation and student learning, teacher well-being, and professional development.

TEACHER EVALUATION

Historical overview. The idea of teacher evaluation is not new, for even in the early beginnings of schooling in the United States, teachers were held to some set of standards. In the mid-1600s, community members evaluated teachers based on a teacher's moral character; her community and religious mores; and the extent to which she maintained discipline, adhered to the community-set curriculum, and kept the school premises clean (Jewell, 2017).

The Industrial Revolution brought more residents to the United States and a shift in the purpose and structure of formal education. More people meant more schools, and more diversity in employment options meant curricula based solely on community mores were no longer sufficient. The community-defined curricula gave way to more academic curricula, and thus, a need for more academically adept teachers and "principal" teachers to guide, supervise, and direct curricula (Jewell, 2017; Marzano, Frontier, & Livingston, 2011).

In 1835, Horace Mann was elected to the Massachusetts senate and spearheaded the creation of the first state Board of Education as well as the first normal school to train young women to be teachers. As the century came to a close, more states had adopted compulsory education laws and created state boards of education. As those board members sought more power, teachers, who had been trained at normal schools, sought more protection from those powers and more autonomy for their teaching (Peterson, 2010).

In 1904, Edward Thorndike published *Introduction to the Theory of Mental and Social Measurement*, which "made him a leader in the test and measurement movement" (Karier, 1986, p. 95) and inspired Ellwood Cubberly to apply Thorndike's scientific principles of measurement to school management in his 1916 book, *Public School Administration* (Marzano et al., 2011). In later editions of the book, Cubberly described how administrators could apply scientific approaches to classroom observations and then grade teachers on the observed lessons.

In the late 1920s and early 1930s, Elton Mayo studied working conditions at the Western Electric factory in Hawthorne,

Illinois. His work inspired a shift from scientific measures to concerns about cooperative work and collective improvement. In the schools, "the Hawthorne model suggested that if teachers were treated as valued partners in the educational process, improved teaching quality would automatically result" (Jewell, 2017, p. 380). Administrators took a coaching approach and schools became concerned with holistic improvements.

Following World War II, record numbers of students enrolled in college under the G. I. Bill. Previously, only children of wealthy families attended college. Now, however, with the option available to more students, schools had to rethink course offerings from general education to those that prepared students for the demands of college as well as the workforce. The heightened attention on achievement was exacerbated by Russia's launch of Sputnik in 1957, which prompted the National Defense Education Act and the federal government's involvement in public education. While the national rally for better mathematics and science achievement was not unequivocally the trigger, by the 1960s, the Hawthorne approach to evaluation was replaced with those of clinical supervision (Jewell, 2017; Marzano et al., 2011).

Clinical supervision was modeled after Morris Cogan's work with teacher candidates at Harvard. He had developed a systematic approach for working with teachers. His work was taken up in Goldhammer's 1969 book, *Clinical Supervision: Special Methods for the Supervision of Teachers*, which detailed a five-step process for clinical supervision of teachers. The process included many of the steps schools are accustomed to today, pre-observation conference, observation using an objective framework, and post-observation conference (Jewell, 2017; Marzano et al., 2011).

While this era in evaluation focused on collaboration between teachers and evaluators, the pendulum was beginning to swing heavily in favor of objective measures, in part due to the passing of the Elementary and Secondary Education Assistance Act of 1965 (ESEA). ESEA allocated federal money to schools serving minority and low-socioeconomic communities. Nevertheless, the money came with strings attached; in exchange for the money, schools had to agree

to participate in standardized tests (Jewell, 2017). The pendulum had an additional push with the release of *A Nation At Risk* in 1983 (United States National Commission on Excellence in Education, Department of Education [USNCEE]).

At the same time Madeline Hunter (1980, 1984) published her seven-step lesson plan framework and the RAND group released its report, *Teacher Evaluation: A Study of Effective Practices* (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). While both emphasized reflection and collaboration as essential components of improvement, both were also used to construct objective measures. Hunter's model provided a structure for observations, and the RAND report highlighted areas of necessity in the evaluation process including, for instance, evaluator competence (Jewell, 2017; Marzano et al., 2011; Wise et al., 1984).

The 1990s saw a continued "quest for improved educational quality…[and] a corresponding increase in individual school accountability levels, along with federal government proposals to increase academic standards" (Jewell, 2017, p. 386). In 1996, Charlotte Danielson published her seminal work, *Enhancing Professional Practice: A Framework for Teaching*, which became, and still is, a standard for evaluating teacher effectiveness. The Danielson Model established a grading system for teachers: unsatisfactory, basic, proficient, and distinguished (Danielson, 1996; Jewell, 2017; Marzano et al., 2011). This model along with others are now commonplace in schools, as the turn of the century brought more legislation (i.e., No Child Left Behind, 2001; Race to the Top, 2009; Every Student Succeeds Act, 2015) that tied federal resources to student achievement scores, which in turn tied student achievement to educator effectiveness.

Efforts to quantify teacher effectiveness resulted in the widespread acceptance of Value-Added Measures (VAM), which are "measures that employ mathematical algorithms in an attempt to isolate an individual teacher's contribution to student learning from all other factors that can influence academic achievement and progress" (Value-added measures, 2013). While VAM originated in the 1980s, they came to the forefront in 2010 when the Bill and Melinda Gates Foundation founded the Measures of Effective Teaching (MET) project to "investigate better ways to identify and develop effective teaching" (Bill and Melinda Gates Foundation, 2013). VAM have continued to grow in popularity and now most states place high weight on student test scores as part of teacher evaluations. Michigan is one of those states.

Evaluation tools. In Michigan, currently 25% of teacher evaluation is based on student achievement and assessment data. Beginning in the 2018-2019 school year, this percentage jumps to 40 percent. In addition to outlining the weight of student performance, Public Act 173 (Michigan Department of Education, 2015) also listed approved evaluation tools for use by Michigan schools: Charlotte Danielson Framework for Teaching, Marzano Teacher Evaluation Model, The Thoughtful Classroom, and/or 5 Dimensions of Teaching and Learning (or other as approved).

Charlotte Danielson Framework. Charlotte Danielson (1996, 2007) conceived of the framework during her work on the development of the Praxis III: Classroom Assessments training programs at Educational Testing Service (ETS). In the process, Danielson (2007) found the "power of structured conversations" (Preface to the First Edition, para. 6) with a common language to be essential to improving practice and to establishing, what effective practice looks like. The framework, according to Danielson, establishes "definitions of expertise and procedures to certify novice and advanced practitioners" and serves as the "public's guarantee" (Why Have a Framework?, para. 1) that teachers are holding themselves to particular standards.

The Charlotte Danielson Framework for Teaching is organized around four domains: Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibility. Each domain is further divided into components, which define aspects of the domain, and elements, which detail the components. Teachers are scored on a four-level scale: ineffective, developing, effective, and highly effective.

Marzano Teacher Evaluation Framework. The Marzano Teacher Evaluation Framework has four domains and each

> In Michigan, currently 25% of teacher evaluation is based on student achievement and assessment data. Beginning in the 2018-2019 school year, this percentage jumps to 40 percent.

domain is subdivided into subcategories, which also have subcategories. The first domain, Classroom Strategies and Behaviors, which has 41 of the model's 60 elements, focuses on the teacher's employment of routines, presentation of content, and management of flexible situations. Domain 2, Planning and Preparing, revolves around the planning and preparation of lessons, materials, and special needs. The third domain details ways in which teachers reflect on their own teaching and develop plans for personal growth. Finally, the fourth domain, Collegiality and Professionalism, considers a teacher's contributions to positive working environments, professional conversations, and school development. Each of the domains has an associated rubric that rates teacher performance on a five-point scale: innovating (4), applying (3), developing (2), beginning (1), and not using (0) (Marzano, Frontier, & Livingston, 2011). A recent publication by the Learning Sciences Marzano Center (2016) reported a correlation between the model and student achievement and between the model and state VAM when the model is implemented with fidelity and enforced.

Thoughtful Classroom Teacher Effectiveness Framework. Another option for teacher evaluation in Michigan is the Thoughtful Classroom Teacher Effectiveness Framework (TCTEF). Unlike the Danielson model and the Marzano framework, which were originally designed to assist teachers and then evolved into evaluation models, TCTEF was created by teachers and administrators for the purpose of evaluating teacher practice. TCTEF is organized into three domains: Cornerstones of Effective Classrooms, Instructional Design and Delivery, and Professional Practice. Cornerstones of Effective Classrooms describes four universal components that are foundational to effective teaching regardless of population or content. Instructional Design and Delivery includes five "critical episodes that maximize learning and motivate all students to do their best" (Silver Strong & Associates, n.d., p. 4). Each of these five episodes provides evaluators an essential question to guide observations, lists of observable instructional practices and student behaviors, an evaluation rubric, and a framework for feedback. Evaluators rate teachers on a four-point scale of novice (1), developing (2), proficient (3), and expert (4).

5 Dimensions of Teaching and Learning. The 5 Dimensions of Teaching and Learning and 5D+ Teacher Evaluation Rubric were developed at the Center for Educational Leadership (CEL), a non-profit organization at the University of Washington. CEL has several national partners, including The Bill and Melinda Gates Foundation and the Michigan Association of Secondary School Principals ([www.k-12lead-](ership.org)

ership.org). The 5 Dimensions of Teaching and Learning Instructional Framework's intent is to provide a "common language and shared vision of high quality instruction" with the mission of "eliminating the achievement gap" (Ripmaster, 2014). The five dimensions - Purpose, Student Engagement, Curriculum & Pedagogy, Assessment for Student Learning, and Classroom Environment & Culture - are divided into 13 sub-dimensions, which are further described by vision statements and guiding questions. The framework is basis for the 5D+ Teacher Evaluation Rubric, which has an additional dimension, Professional Collaboration and Communication. The rubric has 37 indicators of teacher performance and is at the center of conversations during a four-stage inquiry cycle that begins with a teacher self-assessment and ends with the evaluator and teacher analyzing the teacher's impact on his/her professional goals. The rubric has four performance levels: unsatisfactory, basic, proficient, and distinguished.

Research on teacher evaluation. Teacher evaluation is a construct of much interest and research. In Michigan, this research was propelled by the formation of the Michigan Council for Educator Effectiveness (MCEE) in 2011, a committee created to advise the governor, state board of education, and state legislators on the implementation of educator evaluation legislation in the state. The council wanted to pilot the available evaluation tools, so it commissioned the Institute for Social Research (ISR) at the University of Michigan to pilot the Charlotte Danielson Framework for Teaching, 5 Dimensions of Teaching and Learning, Marzano Teacher Evaluation Model, and the Thoughtful Classroom Framework in thirteen districts in lower Michigan during the 2012-2013 school year. Through surveys of and interviews with both administrators and teachers, the ISR team learned much about perceptions of teacher evaluations and the barriers to effective implementation of the evaluation tools (Michigan Center for Educator Effectiveness, 2013).

In this section, ISR's findings are used as a frame for presenting additional research findings on teacher evaluation across the nation and in Michigan (Rowan et al., 2013). MCEE contracted with the four tool vendors to provide district administrators with four days of training on using the assigned tool. Principals felt the training provided sufficient information on the conceptual underpinnings of the tools but not enough practice with scoring evidence, using the scoring rubrics, conducting conferences, or using the tool's software. Only 39% of principals felt they could score accurately and confidently. This lack of practice became evident during observations, as many principals left some rubric elements unscored even when vendors identified the items

as mandatory. Teachers at participating schools were also unsure of the tools, as only 40% reported understanding the procedures that were being used to measure student growth. The most striking findings were those that contrasted the principals' perceptions with those of the teachers'. For instance, 69% of principals felt they had the necessary content knowledge for expertly observing in those content areas; yet, only 44% of teachers thought their principals had the content expertise to evaluate their practice. Differences of opinions were also found with regard to the intent of evaluation. Whereas 84% of principals felt the evaluation tool lead to improvements in teaching performance, only 28% of teachers felt likewise (Rowan et al., 2013).

The ISR team found three main issues with the teacher evaluation process: (a) principals did not score every item on the evaluation tool; (b) low inter-rater reliability; and (c) concerns about the amount of time necessary for evaluations. To better understand the issues, ISR hired and trained a team of former educators to sit alongside principals or alongside one another during observations. These evaluators received the same training the principals did and six additional one-hour training sessions focused on scoring. With the addition of the ISR evaluators, the research team could compare observational data from and between the two types of evaluators.

ISR evaluators and principals had different opinions on when to score non-mandatory items. The median agreement rate between the ISR evaluators and principals varied from 40%-50%, with agreement about 10% higher among ISR evaluators. When looking at just items that were scored, there was low agreement on the score assigned to the teacher for the observed lesson. This finding is especially important because as Rowan et al. (2013) explained, "Small rater errors...can have strong implications for teacher evaluations" (p. 20). Thus, IRS sought to remedy this issue. The team found that reliability is affected by the number of completed observations for an individual teacher. Observation scores were more reliable and more precise when more observations were conducted on an individual teacher. The biggest gains in reliability and precision occurred when observations jumped from one observation of an individual teacher to four observations of that teacher. Additionally, observations were more reliable and more precise measures of instructional quality when more items were scored. Finally, observation scores were more reliable when an individual teacher was observed by more than one observer.

Rowan and colleagues' (2013) MCEE-commissioned study highlighted many of the barriers to effective teacher evaluations in Michigan. First, training is often insufficient in teaching principals to reliably use the scoring rubrics. Second, teachers and principals differ on their perceptions of the purposes and procedures of the evaluation process. Finally, reliable and precise scores require principals to allot more resources and time for conducting observations. These barriers and others will be further explored.

Barriers. In Illinois, Lavigne and Chamberlain (2017) surveyed 606 K-12 school leaders responsible for teacher evaluation during the first year the state required evaluators to be certified in using the Charlotte Danielson Framework. School leaders were trained through computer modules and accompanying materials. Despite 63% of administrators spending more than 41 hours with the training modules and 18% spending more than 61 hours, 61.2% failed one or more of the four modules that had an end-of-module assessment. As in Michigan, leaders felt the most comfortable with and were most successful with the module that provided overview of the tool and state requirements. Leaders were least successful with the module that required them to use the tool's rubrics to score videos of classroom instruction. When asked about the trainings, 71% of leaders felt the trainings improved their observation abilities, but less than half of respondents felt confident in using student growth data as a measure of teacher effectiveness.

Principals felt the evaluations were valuable and resulted in richer conversations, better assessments of instruction, improved instruction, and stronger evidence for dismissal; they also felt these returns were at a high cost. More than half of participating principals felt the new evaluation process somewhat limited (37%) or limited (55%) the time they could allot to other tasks. The time crunch was particularly felt in

84% of principals felt the evaluation tool lead to improvements in teaching performance, **28% of teachers** felt likewise.

(Rowan et al., 2013)
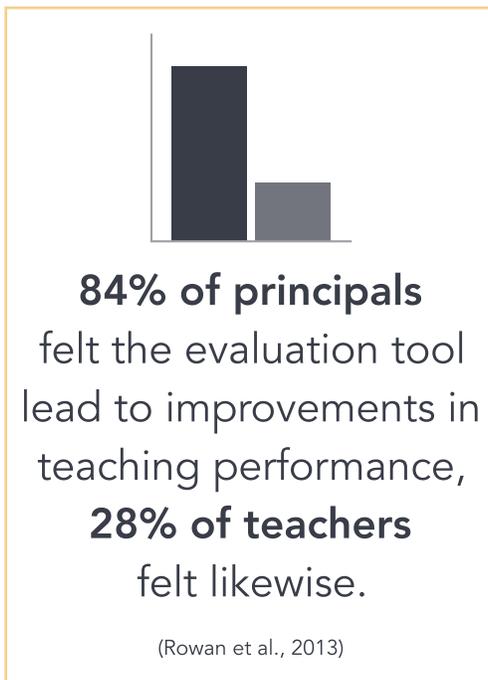
rural schools, as described by one respondent, "In small, rural schools where the principal is the evaluator, curriculum director, bus facilitator, bus disciplinarian, parent liaison, etc., it is difficult to get it all in" (Lavigne & Chamberlain, 2017, p. 194). Principals lamented having less time with students during lunch and recess and with teachers who were not being observed. Principals coped by delegating tasks to others and extending their working hours.

Time was an issue in Ramirez, Clouse, and Davis's (2014) multi-year study on teacher evaluation in Colorado. Through mixed methods and analyses of focus groups, surveys, and face-to-face interviews with representative samples of Colorado's teachers (n = 108), principals (n = 402), and human resources administrators (n = 46), the researchers found, "the most prevalent and concerning critique in the survey was lack of time and assistance for evaluations" (pp. 48-49). Respondents felt there was not sufficient time for completing evaluations and the requisite paperwork. Also of note in this study, respondents indicated the "one-size-fits-all" format of the evaluations was inadequate given the variety of teacher experiences.

Hill and Grossman (2013) suggested that one-size-fits-all rubrics are inadequate. They argued generic observation protocols ignore grade-level appropriateness and "ask us to believe that teaching kindergarten requires the same set of practices and knowledge needed to teach high school algebra" (p. 374). Further, they maintained such evaluation tools defy much of the purpose of the Common Core State Standards and now, the Next Generation Science Standards. Whereas these standards promote disciplinary competency and instruction, evaluation tools assume all teaching can be assessed using a generic, blanket instrument. Hill and Grossman (2013) illustrated the fallacy in that premise: "Engaging students in high-quality investigations in science or using multiple primary-source documents to craft a historical argument, may be entirely missing from generic protocols, since these practices are not easily generalizable" (p. 374).

Norris and colleagues (2017) studied of physical education (PE) teachers' perceptions of teacher evaluation systems in an urban district in the western United States. The team analyzed 22 completed surveys and 10 formal, semi-structured interviews. The PE teachers valued the purposes of teacher evaluation, but only 54.5% of survey respondents were confident in their evaluator's ability to determine their effectiveness as PE teachers. The PE teachers who were interviewed elaborated on that lack of confidence, explaining that they did not think their evaluators have adequate knowledge of PE pedagogy to make evaluative decisions.

Principals, however, are not immune to recognizing the faults of the teacher evaluation process. David Reid (2017) conducted an exploratory case study of six Michigan principals. All six principals worked in districts that recently, because of state mandate, adopted the Charlotte Danielson Framework for Teaching, and all six felt they were working with a flawed system. The six principals did not put much weight on student assessment data, citing the differences in demographics across their districts. Secondly, the principals unanimously agreed that scoring teachers was a biased act because the rubrics are ambiguous and as evaluators, they are not completely objective. Because of this, principals erred on the side of higher scores. Reid (2017) summarized, "In short, because principals knew these evaluation scores would be used for future teacher employment, and because they were not completely confident in the system, [the principals] did not feel comfortable critically evaluating teachers" (p. 1463).

According to Louwes (2017), "Principals are loathe to give teachers bad ratings." One reason principals refrain from low scores is poor performance evaluations require additional documentation and effort. A principal in Kraft and Gilmour's (2017) study explained selective assignment of low scores: "I did it because I couldn't tackle that many teachers at the same time as far as writing prescriptions and then following through on the work that I would need to do" (para. 5). Other principals did not like to give low scores to new teachers because they felt new teachers were still learning and should have a chance to reach potential. Another reason principals shy away from low ratings is the discomfort they have with discussing low ratings, which principals know could lead to dismissal of the very teachers with whom they try to build professional relationships. Principals also worry that if a negative evaluation led to teacher removal, they might be faced with a limited hiring pool and/or less-effective prospective employee. For these reasons and others, Kraft and Gilmour (2017) found that very few teachers across the nation receive ratings other than "effective" or "highly effective." In Michigan, only 2.4% of teachers receive either developing/needs improvement or ineffective/unsatisfactory (Kraft & Gilmour, 2017).

Nonetheless, high ratings do not mean more satisfied teachers. Ford, Van Sickle, Clark, Fazio-Brunson, and Schween (2017) interviewed 37 teachers from across Louisiana. The semi-structured interviews were conducted prior to teachers receiving their evaluation scores as well as after. Ford et al. (2017) found that despite receiving "effective" or "highly

effective" ratings, the teachers expressed unsettling dispositions toward the evaluation process. Teachers felt a loss of autonomy and self-efficacy, sensing distrust from their superiors and a "loss of joy" (p. 232) in their work. In addition, they were dubious of the process for calculating ratings, suggesting that bias plays a part in evaluations, which leads to meaningless ratings. As one teacher explained,

I think I definitely deserved the grade I got, but at the same time, I think, well, what about that football coach that won state championship last year? And he sits on his computer and orders football equipment all day long during social studies class, but gets a 3.8? You can't tell me that the work I put in is the same. (Ford et al., 2017, p. 225)

Moreover, the participating teachers felt the evaluation process was part of a bureaucratic exercise that did not capture the extent of their teaching practice.

Despite the barriers, teacher evaluation has a long tenure in U. S. schools. The developers of evaluation tools and the states that adopt them attest that regardless of the barriers of time, biases, and broadness, teacher evaluations are worthy of implementation. Evaluations are thought of as the key to higher student achievement and successful schools.

TEACHER EVALUATION AND STUDENT ACHIEVEMENT

This section explores the political and societal contexts that led the movement for teacher evaluation based on student performance. In addition, this section describes research that has highlighted the complexity of trying to definitively attribute student achievement (or lack thereof) to teacher (in)effectiveness.

The Coleman Report. The Civil Rights Act of 1964 mandated that the U. S. Commissioner of Education survey the state of educational opportunity in the U. S. and report to the president within two years. Thus, in the fall of 1965, the National Center for Education Statistics (NCES) contracted with consultants, most notably James Coleman from John Hopkins University who became eponymous with the report, and contractors including Educational Testing Service, to survey a sample of 5% of U. S. schools, focusing more heavily on schools that serve minority communities. The goal was to describe educational opportunity in terms of equality and to determine which school characteristics contribute to student achievement. The survey found that several characteristics influence student achievement, including the socioeconomic background of the community, the range of facilities at the school, and the academic aspi-

rations of a student's peers, but the dominant characteristic was the quality of teachers: "The quality of teachers shows a stronger relationship to pupil achievement. Furthermore, it is progressively greater at higher grades, indicating a cumulative impact of the qualities of teachers in a school on the pupil's achievement" (Coleman et al., 1966, p. 22). At the time, teacher quality was determined by teachers' scores on a verbal skills test and their educational background, including the education levels of the students' parents.

While Coleman's seminal study did not focus on teacher practices or individual teacher's contributions to student achievement, it was just one example of the continuing attempt to increase rigor and accountability in U. S. schools since the return of WWII veterans and the Soviet's launch of *Sputnik*.

Process-product. Against that background, research on teaching in the 1960s and 1970s supposed teaching to be a process-product profession: *Processes*, or the teachers' behaviors, determined the extent of student learning, the products. This conception of education thrust researchers from laboratories and desks into actual classrooms, asking the question: *How do a teacher's practices relate to student achievement?* Shulman (1992), attested to the moral underpinnings of such a stance:

While it did not particularly emphasize moral behavior on the part of teachers or moral outcomes as its student products, it unambiguously rested on a moral claim: teaching ought to be understood and valued primarily through its effects on student learning. The purpose of teaching is the amelioration of ignorance, and studies of teaching that make claims of excellence for some kinds of pedagogy must buttress those claims with evidence of the impact of the teaching on the lives and capacities of students. (p. 20)

The call for evidence became even louder with the outcry of *A Nation at Risk* (USNCEE, 1983). The report was intended to serve as a national bill of health for public education, and the prognosis was bleak. The report argued that modern students were not scoring as high on placement tests as were previous generations and that the deteriorating scores meant U. S. students were falling behind students of other leading nations. In short, the report said U.S. public schools were producing mediocre products.

The report presented a prescription for remedying the destitute state of public education: set higher standards by increasing graduation requirements and college entrance scores, mandate the length of the school day and the num-

ber of days in a school year, make teaching a lucrative profession so as to hire only the most adept, and allocate resources for targeting specific populations of students. More importantly, this report was a catalyst in the movement toward school evaluation based on standardized scores.

Value-added measures. Meanwhile, in Tennessee, researchers had been collecting longitudinal data on the effect of individual teachers on student progress achievement. In 1984, Dr. William Sanders and Dr. Robert McLean (as cited in Sanders & Horn, 1994) published a study on the feasibility of using student achievement data to evaluate teacher effectiveness, a model that became known as the Tennessee Value-Added Assessment System (TVAAS). In 1991, the state adopted legislation that made the "*product* of educational experience rather than the *process* by which it was to be achieved" (Sanders & Horn, 1994, p. 300) the focus of its accountability movement, which made the TVAAS integral to teacher evaluation.

Through the Tennessee Value-Added Assessment System, researchers collected over three million records of the state's entire student population of Grade 2 to Grade 8 students from 1990-1996. In particular, Sanders and Rivers (1996) analyzed teacher effects on the group of students that was in Grade 3 during the 1991-1992 school year through that cohort's fifth grade year, 1994-1995. They found the mathematics achievement of a student who had teachers with low effective ratings for those three school years differed by 52-54 percentage points from the mathematics achievement of a student who had teachers with high effective ratings during the same grade spans. What is more, if a student had a relatively ineffective teacher one year and an effective teacher the next, the student's mathematics achievement, despite making gains, was still stunted by the previous year's ineffective teacher. Thus, they concluded teacher effects on student growth achievement are residual and cumulative.

Interest in VAM grew. By 2010, large school districts, including those in New York City, Washington D. C., Dallas, and Los Angeles had adopted policies to evaluate teachers based on student growth models (Johnson, 2010). Also, in 2010, the first grants were awarded as part of President Obama's Race to the Top funds. According to the U. S. Department of Education (2016) website, one goal of Race to the Top was to advance reforms in "recruiting, developing, rewarding, and retaining effective teachers and principals" (Program Description, para. 1). VAM had become part of the national scene, but not without concern.

Teacher behavior concerns. In his book, *Measuring Up:*

What Educational Testing Really Tells Us, Daniel Koretz (2008) described seven different behaviors that raise achievement scores: *working more effectively, teaching more, working harder, reallocation, alignment, coaching,* and *cheating*. The first three are exactly the types of behaviors touted by proponents of VAM. By attaching student achievement scores to teacher evaluation, teachers are motivated to *work more effectively, teach more,* and *work harder*. Of course, when merit pay and job retention are attached to achievement scores, as they are in VAM, teachers can become incentivized to magnify these three behaviors to the extreme. For example, only two states, including Michigan, mandate recess time for elementary students. In order to *teach more,* schools eliminate recess, and some estimate the elimination adds two weeks of instruction to the year (Schachter, 2005).

Reallocation refers to the redistribution of time. For example, in Michigan, the Michigan Student Test of Educational Progress, or M-STEP, begins yearly assessments in Grade 3. However, social studies is not assessed until Grade 5. Teachers in Grades 3 and 4 might be inclined to shift instructional time away from social studies and spend more time on mathematics and English language arts (and science for Grade 4) because those are the content areas for which their students will be assessed.

Alignment is the intent to align standards and instruction to tests. While this is not necessarily problematic, alignment assumes the tests are assessing valuable skills and are representative of the larger domain. For example, it is difficult to assess processing skills and easy to assess facts.

Coaching and *cheating* refer to behaviors that inflate student test scores but do not reflect student learning. Teaching students test-taking strategies is an example of *coaching. Cheating* includes behaviors such as offering students hints, correct responses, or unapproved extended time. All of these behaviors have been of question in Michigan schools (e.g., Wells, 2016).

Other concerns. According to Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2012), when VAM attribute a student's measured achievement gains to a teacher's effectiveness, that attribution "assumes that student learning is measured well by a given test, is influenced by the teacher alone, and is independent from the growth of classmates and other aspects of the classroom context" (p. 8). Hill, Kapitula, and Umland (2011) illustrated the error in the first two assumptions. The researchers collected extensive observational data from a small set of middle school mathematics teachers and calculated value-added scores for each

teacher in the large, urban district. Student outcomes were based on the state mathematics assessment, which is administered for every student in Grades 3-8. In addition to the classroom observations, which were coded using an observational protocol designed specifically to attend to the quality of mathematical instruction, teachers completed surveys to measure their mathematical knowledge for teaching (MKT). Researchers found "strong positive correlations between teacher resources, measured by the MKT survey, and the mathematical quality of instruction" (p. 809). However, further analyses of case studies also found teachers with high value-added scores but very low MKT scores. In fact, one of the teachers even struggled to answer student problems presented in course materials. Hill et al. (2011) could not find obvious teaching characteristics of the low MKT teachers that contributed to high student scores, concluding "it seems possible either forces external to these teachers - their teaching assignments or the compensatory parental inputs as suggested in Ishii and Rivkin (2009) - or stochastic variation is responsible for their scores" (p. 824). Hill et al. also considered the possibility that "extensive test preparation activities" could also be at play.

The third assumption relies on consistency for a plethora of classroom characteristics. These include class size, curriculum choices, and the availability of expert and school resources, such as specialists and libraries. Another aspect of the classroom is the student population. VAM assume that students are randomly assigned to classrooms; yet such is rarely the case, and "statistical models can't fully adjust for the fact that some teachers will have a disproportionate number of students who have greater challenges" (Darling-Hammond et al., 2012, p. 10). This means, according to Berliner and Glass (2014), that teachers, who desire merit pay and job retention, are incentivized to stack the deck, so to speak, in terms of assigning students to their classrooms. For example, Amrein-Beardsley and Collins (2012) found that teachers in the Houston Independent School District were afraid to teach in the fourth grade, the year when English Language Learners (ELLs) transition to main-

stream English-only instruction. Having fewer ELL students meant having better chances at higher test scores. Borman and Kimball (2005) found correlations between student achievement and teachers who were rated effective, but noted that "better teachers may be assigned, and seek out assignments, to classrooms with more advantaged, nonminority, and higher-achieving students" (p. 17).

Measures of Effective Teaching (MET). Accounting for student differences was one of the challenges the MET project set out to tackle. The Bill and Melinda Gates Foundation (2013) funded the MET research partnership involving 3,000 teachers and representatives from universities, research institutes, and the Danielson Group with the goal of creating reliable procedures for identifying effective teaching. The MET project is a VAM model that incorporates several measures, including classroom observations by more than one evaluator, student surveys, teacher surveys of working conditions, and Content Knowledge for Teaching (CKT) tests. Research has revolved around finding reliable weights for each measure, determining frequency and quality of observations, and revising the student growth equations to control for individual student differences and prior test scores. The result was a model that accurately predicted student achievement based on teacher evaluations: "The research confirmed that, as a group, teachers previously identified as more effective caused students to learn more. Groups of teachers who had been identified as less effective caused students to learn less" (Cantrell & Kane, 2013, pp. 6-7).

Teacher effectiveness and student achievement. While the MET Project has been able to engineer a teacher evaluation process that positively links teacher effectiveness and student achievement, the majority of VAM models are rife in faulty assumptions. These assumptions include, but are not limited to prior student achievement, classroom and school resources, student population in a given class, invalid or unreliable tests, and quality observations.

> "The research confirmed that, as a group, teachers previously identified as more effective caused students to learn more. Groups of teachers who had been identified as less effective caused students to learn less"
>
> (Cantrell & Kane, 2013, pp. 6-7).

TEACHER WELL-BEING

The Michigan Department of Education (MDE) posits that educator evaluations support educator well-being, which could be described as the basic physiological need we all have (Ryan & Deci, 2000), and facilitates "educators' personal pursuits of excellence," or professional growth. This section explores well-being from the lens of self-determination theory and professionalism. In addition, connections are made between well-being and teacher retention.

Self-determination theory. According to Ryan and Deci (2000), well-being includes autonomy, competence, and relatedness. Autonomy refers to "the feeling of volition" (p. 74). Self-determination theory (SDT) is a theory of motivation that proposes a continuum from controlled motivation to autonomous motivation. SDT is not merely a dichotomous distinction between extrinsic and intrinsic motivation; rather SDT identifies a spectrum of conditions necessary for motivating individuals (i.e., educators).

Strong and Yoshida (2014) studied teacher autonomy in Michigan's three largest public school districts. The researchers surveyed elementary and secondary special education and classroom teachers using the Teacher Work-Autonomy Scale (Friedman, 1999). Analyses revealed that the respondents (N = 437) perceived the highest level of autonomy with regards to classroom management and lowest level of perceived autonomy with decisions about professional development (Strong & Yoshida, 2014). Pearson and Moomaw (2005) studied teacher autonomy in three districts in Florida and found that as perceptions of general autonomy increased, so did perceptions of empowerment and professionalism. Additionally, as perceptions of empowerment and professionalism increased, on-the-job stress decreased, and job satisfaction increased. Thus, the researchers concluded, "teachers who perceive themselves as empowered also view their occupation as a true profession" (p. 48).

Feelings of competence are also a nutriment to well-being (Gangé & Deci, 2005). Firestone and Pennell (1993) list four conditions that lead to teachers' feelings of competence: administrative support, adequate physical facilities, adequate instructional materials, and manageable workloads (as cited in Firestone, 2014). Ingersoll (1999) links competence to the learning environment, noting that if teachers do not have formal background training for the position for which they are assigned, the teachers' lack of feelings of competence "decreases teachers' morale and commitment" (p. 29).

Finally, teacher well-being is influenced by the degree to which teachers feel connected to others in their professional context. According to Gangé and Deci (2005), "satisfaction of the needs to be connected to others and to be effective in the social world support people's tendency to internalize the values and regulatory processes that are ambient in their world" (p. 337). Collaboration and support are essential to teacher well-being and to student success (Schleicher, 2017).

Professionalism. The Organisation for Economic Cooperation and Development (OECD) defines professionalism as "the level of autonomy and internal regulation exercised by members of an occupation providing services to society" (Schleicher, 2017). The OECD asserts that teacher professionalism is comprised of three components: autonomy, or teachers' decision-making power, knowledge base for teaching, and peer networks. When examining data from the Programme for International Student Assessment (PISA), to which 28 million 15-year-olds, 89,000 parents, 93,000 teachers, and 17,500 principals in 72 countries responded, the OECD found that high levels of professionalism within an educational system were correlated with high satisfaction with the status of the teaching profession as well as teachers' satisfaction with the profession, their work environment and their sense of self-efficacy.

Feedback. According to Firestone (2014), self-efficacy "enhances intrinsic motivation most when the individual gets feedback on performance" (p. 4). Thus, an imperative part of teacher evaluation is professional feedback, and equally as important, is what teachers do with the feedback. Tuytens and Devos (2017) described three uses of feedback as found in literature on the subject (e.g., Rossi & Freeman, 1993; Visscher & Coe, 2003): instrumental, conceptual, and symbolic. Instrumental use of feedback results in the teacher taking immediate action toward corrective practices or professional development. Conceptual use does not necessarily reflect a change in practice but rather a change in teacher beliefs. Symbolic use refers to feedback that confirms a teacher's existing practices or beliefs.

Teacher retention. Tuytens and Devos (2017) found that daily feedback was a cornerstone of teacher evaluation in the eight participating schools in their study, and this feedback included the school leaders expressing appreciation for teachers and their work. Appreciation, while seemingly minor, actually plays a large role in teacher job retention. Using data from the National Center for Education Statistics' (NCES) Schools and Staffing Survey (SASS), Ingersoll (2001) found that among teachers who left the teaching profession, lack of administrator support was one of the top two reasons for leaving. The other was salary.

In this section, connections can be made among teacher well-being, student achievement, and teacher retention. While teachers seek autonomy, they also desire administrative support. Support in the form of feedback is essential to teachers' feelings of competency, particularly when teachers receive daily feedback and appreciation. What is more, teachers feel more competent when they have the resources and facilities to support their practice. Finally, teacher connectedness strengthens the professionalism of the teaching profession and student achievement.

PROFESSIONAL DEVELOPMENT

This section defines professional development and describes the components of effective professional development. Additionally, this section provides a brief look at professional development in the United States and in Michigan.

Definition. According to the *Every Student Succeeds Act*, signed into law in 2015, defines *professional development* involves activities that

(A) Are an integral part of school and local educational agency strategies for providing educators (including teachers, principals, other school leaders, specialized instructional support personnel, paraprofessionals, and, as applicable, early childhood educators) with the knowledge and skills necessary to enable students to succeed in a well-rounded education and to meet the challenging State academic standards; and

(B) Are sustained (not stand-alone, 1-day, or short-term workshops), intensive, collaborative, job-embedded, data-driven, and classroom-focused. (Learningforward, 2017)

Successful professional development. Eun (2008) posits that teacher professional development should be grounded in a theoretical frame, in particular, Vygotsky's sociocultural theories of development. Therefore, professional development must hinge on social interactions. During these interactions, the more competent leader must be able to accurately assess teachers' current and potential levels of development with regard to the professional development focus and must invoke a mechanism to bridge present developmental needs with the future goals of the professional development process (p. 146).

Eun (2008) elaborated further by explaining that professional development, grounded in Vygotksy's theory, is likely to be the most successful when conducted in-school: "Social interaction would be most facilitated when people are interacting with others they know very well. In addition, this type of professional development would be most sensitive to the needs and goals of the participating teachers" (p. 146). Another necessity for successful professional development is time, as internalization through social interaction takes times. Finally, Eun asserted that continuous follow-up support is necessary for teachers to be successful with the new skills or practices. Eun's position has the support of other scholars who have found these components to be foundational for effective professional development (see Wei, Darling-Hammond, & Adamson, 2010).

Professional development in the United States. Wei, Darling-Hammond, and Adamson (2010) analyzed data from the 2008 SASS, 2009 Met Life Survey, and the first Teaching and Learning International Survey (TALIS). The review found that in 2008, U. S. teachers received less intensive professional development, meaning fewer hours on a given topic, than they received in 2004. This finding is of great concern because Yoon et al. (2007, as cited in Wei et al., 2010) found that professional development on a single topic requires an average of 49 hours over a time span of six to 12 months in order to be effective.

Wei and colleagues (2010) also rated states using a Professional Development Access Index, comprised of 11 indicators of effective professional development. The index ratings revealed that :at least 80% of new teachers report participating in new teacher induction programs; at least 80% of beginning teachers report having a mentor; at least 51% of new teachers receive four other induction supports;

> Professional development on a single topic requires an average of 49 hours over a time span of six to 12 months in order to be effective.
>
> (2007, as cited in Wei et al., 2010)

at least 80% of teachers received 17 or more hours of professional development in their assigned content areas; at least 67% of teachers received professional development on technological integration instruction; at least 67% of teachers participated in professional development on reading instruction; at least 67% of teachers participated in professional development focused on classroom management; at least 51% of teachers participated in professional development for teaching students with limited English proficiencies; and an average of 50% of teachers received at least 50 total hours on any combination of the previously listed six areas of professional development across the past 12 months (p. 35). Unfortunately, in 2008, Michigan fell short, receiving only two points out of the possible 11. Michigan received those two points for meeting the standards for professional development in a teacher's assigned content area and for professional development on technology-integrated instruction.

Tucker (2011) described professional development in high-performing countries, as determined by PISA. In Japan, "instructional development and professional development are merged and professional development becomes an integral part of the process of improving instruction in the school, informed by the latest and best research" (p. 20). In Singapore, professional development is a way to "nurture" (p. 21) the current talent. Both of these countries use professional development to strengthen the work of practicing teachers, helping teachers to develop in relation to current research and to refine their craft. In the U. S., however, "we do not have well-defined career paths for teachers who want to advance their careers, but stay in teaching" (p. 21). If teacher evaluation intends to help teachers improve and help students achieve, teachers and administrators will need sustainable professional development that focuses on current best practices and continuous support and feedback for improvement.

4 CONTEXT FOR THE RESEARCH STUDY ON TEACHER AND ADMINISTRATOR PERCEPTIONS OF THE TEACHER EVALUATION PROCESS IN MICHIGAN

This research study stemmed from a grant from the Michigan Department of Education, which challenged applicants to **"articulate the design, development, implementation and post implementation of a comprehensive research and/or evaluation system addressing one or more of the identified activity areas listed below:**

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."

Our application addressed sub-questions related to Activity 1 and Activity 2, and was granted funding in January 2017.

5 TIMELINE FOR THE RESEARCH STUDY ON TEACHER AND ADMINISTRATOR PERCEPTIONS OF THE TEACHER EVALUATION PROCESS IN MICHIGAN

January 2017	Notice of Michigan Department of Education grant award
February 2017	Literature review
March 2017	IRB approval
March 2017	Questionnaire design and pilot study validation
April 2017	Interview design and pilot study validation
June 2017	Distribution of questionnaire after teacher received their ratings for 2017-2018 school year
July 2017	Interviews (phone and electronic)
August /September 2017	Data analysis
October 2017	Research Report Submitted to MDE
December 2017	Manuscript #1 submitted for journal publication
January 2018	Research presented at International Education Conference
March 2018	Manuscript #2 submitted for journal publication
April 2018	Research presented at American Educational Research Conference (pending acceptance)

6 METHODS

A convergent parallel mixed-methods design was utilized to collect both quantitative and qualitative data simultaneously through the use of a survey and structured interviews/extended responses (Leech & Onwuegbuzie, 2009). Combining quantitative data from the participants' responses to anchored sliding-scale questions with qualitative data from their responses to the open-ended questions allowed for clarification and enhancement of the teacher evaluation process, as well as for convergence and corroboration (Onwuegbuzie & Leech, 2004).

Data Collection

A stratified convenience sample of 1,746 public school educators (teachers; n = 1,274 and administrators; n = 474) from Michigan's 56 Intermediate School Districts were surveyed using a researcher-constructed survey tool that included categorical sliding scales and extended response questions. An additional stratified sample of 128 teachers and 48 administrators from the original participants was selected for phone or electronic interviews, which consisted of extension questions from the survey.

Participant Demographics

Consistent with national statistics on educators (Allegretto & Mishel, 2016), more than 70% of the participants were female. Female teachers far outnumbered males; whereas, the administrator participants were more evenly gendered (see Table 1).

Table 1

Participant Gender

Participant:	Number Female	Percent Female	Number Male	Percent Male	Total
Teachers	1008	56.5%	264	14.8%	1272
Administrators	265	14.9%	209	11.7%	474
Total	1273	71.4%	473	26.5%	1746

The teacher participants' years of classroom experience were evenly distributed with nearly one in five in their first five years in the profession. Nearly half of the administrators taught for 6-15 years before they moved into administrative positions (see Table 2).

Table 2

Years of Classroom Experience

Response	ADMINISTRATORS		TEACHERS	
	Percentage	Total	Percentage	Total
5 years or fewer	7.7%	38	19.4%	250
6-10 years	25.3%	124	22.9%	296
11 - 15 years	23.4%	115	19.2%	248
16 - 20 years	18.5%	91	18.6%	240
21 - 25 years	12.0%	59	10.5%	136
26 - 30 years	7.9%	39	6.2%	80
31 - 35 years	2.7%	13	1.9%	25
36 - 40 years	1.8%	9	1.1%	14
41 - 45 years	0.4%	2	0.1%	1
46 - 50 years	0.2%	1	0.2%	2
Total	100%	491	100%	1292

Data Analysis

For the quantitative questions, we used exploratory factor analysis using principle components on the items asked of both teachers and administrators to create subscales based on identified components with reliability statistics on a subset of the data as preliminary to analysis. Subscales were then used on all data in predictor models to address these research questions.

Teachers and administrators responded to 38 of the same items contained in the survey. We used principal components analysis with a varimax rotation to identify three (3) groups of items using SPSS (2017). Inter-item reliability coefficients were calculated for items within each subgroup. Seven items were not included in the final groupings to increase reliability of the measures. Coefficient alpha (Cronbach, 1951) was calculated for each group of items The identified groups and reliability coefficients for responses were as follows: teacher improvement (16 items), α = .89, (used to answer the research question, "Do educators perceive teacher evaluation processes provide critical feedback and professionally development to improve teacher practices?"), teacher well-being (11 items), α = .92, (used to answer the research question, "Do educators perceive teacher evaluation processes support teacher well-being?"), and student learning (4 items), α = .82 (used to answer the research question "Do educators perceive teacher evaluation processes support student learning?) (Table 3). Coefficient alpha for the total of grouped responses was α = .94. Inter-item correlations ranged from .24 to .76 for the teaching improvement subgroup (Appendix A). Inter-item correlations ranged from .38 to .86 for the teacher well-being subgroup (Appendix B). Inter-item correlations ranged from .27 to .81 (Appendix C).

Table 3

Coefficient Survey Item Group to Answer the Research Questions

Survey Item Group	Number of Items	M	SD	α
Teacher Improvement	16	2.11	3.65	.89
Teacher Wellbeing	11	.26	3.89	.92
Student Learning	4	.11	4.59	.82
Total	31	.83	3.72	.94

We used nonparametric tests to determine if there were differences in teacher responses and administrator responses. We used Kruskal-Wallace one-way analysis of variance (KWANOVA) for comparison as sample distributions were not normally distributed and because there were more teacher responses than administrator responses.

We analyzed participants' responses to the open-ended questions from the survey and electronic interviews using a multi-step coding process. First, we used a priori codes based on our research questions to categorize the responses using Dedoose data management system. Next, we applied collaborative open coding to sub-categorize the participant's responses. Next, we combined categories and recategorized until the data were saturated. Finally, we identified emergent themes from the data, from which we excerpted quotes to illustrate each theme in the Results section.

7 RESULTS

In this section, we present results from the mixed-methods study of 1,746 Michigan educators who completed our questionnaire or interview in response to teacher effectiveness ratings, evaluation tools, and data measures to determine student growth.

TEACHER RATINGS

Nearly 99% of the teacher participants were rated Effective or Highly Effective during the 2016-2017 school year (see Table 4) .

Table 4

Teacher Effectiveness Ratings

	Highly Effective		Effective		Minimally Effective		Total
Teachers	559	44.0%	693	54.6%	18	1.4%	1270

COMPENSATION

Nearly 20% of the participants' schools compensate teachers for receiving a highly effective rating (see Table 5).

Table 5

Teacher Compensation: Are Teachers in Your School Compensated or Incentivized for Receiving a Highly Effective Rating?

	Yes	Percent	No	Percent
Teacher	236	18.6%	1035	81.4%
Administrator	109	22.9%	366	77.1%

Of those who reported that highly qualified teachers receive compensation, approximately half reported that teachers are compensated but do not specify an amount. Earning less than $249 appears to be the most common reward (see Table 6).

Table 6

Highly Qualified Teacher Compensation: How are Teachers with Highly Effective Ratings Compensated or Incentivized?

Response	ADMINISTRATORS		TEACHERS	
	Percentage	Total	Percentage	Total
Unspecified Stipend	51.9%	56	47.9%	104
249 or Less	22.2%	24	22.1%	48
250 or More	9.3%	10	2.3%	5
Days Off	7.4%	8	10.1%	22
Unsure	3.7%	4	4.6%	10
Fewer Evaluations	3.7%	4	3.2%	7
Promotion	1.9%	2	7.4%	16
None	0.0%	0	2.3%	5
Total	100.0%	108	100%	217

EVALUATION TOOLS

The most common evaluation tool used in the participants' schools was the Danielson Framework for Teaching, used in nearly 50% of schools, followed by 5 Dimensions of Teaching and Learning, used in 29% of schools (see Table 7).

Table 7

Evaluation Tools

Evaluation Tool	Number	Percent
Danielson	836	46.9%
5 Dimensions of Teaching & Learning	481	27.0%
Marzano	205	11.5%
The Thoughtful Classroom	104	5.8%
Other	96	5.4%
Danielson and Marzano	13	0.7%
5 Dimensions of Teaching & Learning and Danielson	12	0.7%
5 Dimensions of Teaching & Learning, and The Thoughtful Classroom	8	0.4%
Danielson and The Thoughtful Classroom	8	0.4%
5 Dimensions of Teaching & Learning and Marzano	7	0.4%
5 Dimensions of Teaching & Learning, Danielson, Marzano, and The Thoughtful Classroom	5	0.3%
Marzano and The Thoughtful Classroom	5	0.3%
5 Dimensions of Teaching & Learning, Danielson, and Marzano	1	0.1%
5 Dimensions of Teaching & Learning, Marzano, and The Thoughtful Classroom	1	0.1%
Danielson, Marzano and The Thoughtful Classroom	1	0.1%
Total	1783	100.0%

STUDENT GROWTH MEASURES

When asked as to how student growth was measured at their school, administrators provided a range of responses including school/district required assessments, state mandated assessments, and classroom based growth measures (see Table 8).

Table 8

Student Growth Measures: During Your Evaluations of Teachers,
What Data Are Used to Measure Student Growth?

Response	ADMINISTRATORS		TEACHERS	
	Percentage	Total	Percentage	Total
NWEA	177	19.6%	340	23.7%
Pre-post tests	118	13.1%	318	22.2%
M-Step	59	6.5%	71	5.0%
(P)SAT	74	8.2%	57	4.0%
MAP	49	5.4%	98	6.8%
Dibels	46	5.1%	47	3.3%
Dra	39	4.3%	76	5.3%
Local assessments	37	4.1%	13	0.9%
STAR	31	3.4%	92	6.4%
Slo	30	3.3%	16	1.1%
AimsWeb	29	3.2%	36	2.5%
Standardized tests	28	3.1%	54	3.8%
Fountas and Pinnell	24	2.7%	22	1.5%
Common assessments	22	2.4%	27	1.9%
Other	24	2.7%	32	2.2%
Formative assessments	21	2.3%	10	0.7%
State assessments	16	1.8%	6	0.4%
Summative assessments	12	1.3%	14	1.0%
Unit assessments	11	1.2%	6	0.4%
District assessments	10	1.1%	20	1.4%
Classroom assessments	8	0.9%	17	1.2%
IEP goals	8	0.9%	37	2.6%
Portfolios	8	0.9%	7	0.5%
iReady	7	0.8%	4	0.3%
MLPP	7	0.8%	3	0.2%
Delta	7	0.8%	9	0.6%
Total	902	100.0%	1432	100.0%

PURPOSE OF EVALUATION

The quantitative analysis of categorical data for each group of questionnaire items revealed that the sample distributions of responses by educators were not normally distributed. Descriptive statistics for each group of item responses by teachers were: teacher improvement (Median = .94, Range = 18.75), teacher well-being (Median = -.85, range = 20), student learning (Median = -1.00, Range = 20), and total (Median = -.28, range = 18.29). Descriptive statistics for each group of item responses by teachers were: teacher improvement (Median = 4.75, Range = 14.25), teacher well-being (Median = 2.60, range = 18), student learning (Median = 3.00, Range = 19.50), and total (Median = 3.33, range = 14.97). We used Kruskal-Wallace one-way analysis of variance (KWANOVA) comparisons across teacher (n = 1742) and administration responses (n = 475) for our identified groups of items (see Table 9).

Table 9.

Educator perception of teacher evaluation processes comparisons across grouped survey items

	Teacher Responses (n = 1272)				Administrator Responses (n = 474)			
	M	SD	Median	Range	M	SD	Median	Range
Teacher Improvement	1.20	3.48	.94*	18.75	4.55	2.89	4.75*	14.25
Teacher Well- Being	- .59	3.76	- .85*	20	2.53	3.26	2.60*	18
Student Learning	- .83	4.47	-1.00*	20	2.62	3.92	3.00*	19.50
Total	- .08	3.55	- .28*	18.29	3.24	3.02	3.33*	14.97

* The median score is the most representative of the sample due to non-normal distributions.

To gain a sense of how administrators and teachers perceived the evaluation process at their schools, we asked them: "What *is* the purpose of the teacher evaluation process in your school?" The most common administrator responses related to *improving instruction*, with 26% of principals listing that as the purpose of the evaluation process, followed by *student growth and learning* with 21% (see Table 9). It is important to note that while it might be implied that improving instruction would entail student learning, constructs of instructional improvement often center on lesson improvement through pathways such as increasing teachers' knowledge, commitment, and access to resources, rather than to student learning outcomes explicitly (Lewis, Perrry, & Murata, 2006). Nonetheless, combined with p*rofessional development and coaching*, 67% of the administrators cited purposes related to improving teaching and learning.

Nearly one in five administrators noted that the purpose of the teacher evaluation process is to meet compliance mandates, and 13% cited the negative aspects of the evaluation process, with 8% stating *rank and sort*, and 5% stating that the purpose is to punish incompetent teachers. Finally, despite the MDE's expressed goal of teacher evaluations supporting educator well-being, less than 1% of administrators listed teacher well-being as a purpose of the evaluation process.

Compared to the administrators, teachers most commonly cited compliance reasons (30%) as the current purpose of the teacher evaluation process, followed by student growth and learning (13%) and improved instruction (22%). Notably, more than one in ten teachers expressed that the purpose of the evaluation process is to punish teachers.

Table 10

Actual Purpose of Evaluation: What is the Purpose of the Teacher Evaluation Process in your School?

Response	ADMINISTRATORS		TEACHERS	
	Percentage	Total	Percentage	Total
Improve instruction	144	25.7%	273	22.0%
Student growth and learning	115	20.5%	160	12.9%
Compliance	108	19.3%	371	29.9%
Professional development	105	18.7%	80	6.5%
Rank and sort	45	8.0%	115	9.3%
Punitive	28	5.0%	132	10.7%
Coaching	12	2.1%	6	0.5%
Teacher well-being	4	0.7%	0	0.0%
Unsure	0	0.0%	102	8.2%
Total	561	100.0%	1239	100.0%

As a follow-up to the above question about the current purpose of the evaluation process, we also asked, "What should be the purpose of the teacher evaluation process in your school?" Similar to their responses to the previous question, administrators' responses predominantly listed improved teaching and learning with 96% of administrators citing student growth and learning, professional development, improved instruction, or coaching as purposes. Once again, teacher well-being received a modicum of responses (2.72%).

Teachers also expressed that the evaluation process should focus on teaching and learning at similar rates (97%) as the administrators. Surprisingly, 2.5% of teachers suggested that the purpose should be punitive, compared to less than 1% of administrators.

Table 11

Ideal Purpose of Evaluation: What Should the Purpose of Teacher Evaluation be in your School?

| Response	ADMINISTRATORS		TEACHERS	
	Percentage	Total	Percentage	Total
Improve instruction	165	26.4%	601	41.2%
Student growth and learning	220	35.1%	417	28.6%
Professional development	192	30.7%	355	24.3%
Coaching	27	4.3%	40	2.7%
Punitive	5	0.8%	36	2.5%
Teacher well-being	17	2.7%	9	0.6%
Total	626	100.0%	1458	100.0%

Student learning. There was a significant difference in responses to items grouped as *Student Learning* by teachers (mean ranking = 769.36) than administrator responses (mean ranking = 1156.27), c_2 (1, N = 1747) = 203.32, p < .01, h2 = 12. This is a medium effect size with 12% of the variability in responses accounted for by teacher versus administrators. Administrators perceived that the teacher evaluation process increased student learning more positively than teachers did as demonstrated by their responses to the items included in the teacher improvement group of questions.

Teachers' perspectives on teacher evaluation and student learning.

To determine teachers' perception of the role that the educator evaluation process plays in supporting student learning, we asked teachers to articulate *how* their school's evaluation system impacted their students' learning. Teachers most commonly indicated that the process had positively impacted their students' learning and resulted in improved instructional practices, including research-based strategies, sharing learning outcomes explicitly with students, and creating positive learning environments, as demonstrated by this quote: "Students benefit from a teacher who has set goals and reflects on them thoughtfully to show growth in a variety of ways. Students benefit from a teacher who has carefully evaluated the learning environment." Another teacher wrote, "The rubric focuses on student responses to teaching. This makes teachers focus on it, too. It is not just about the teacher but it is about the students. It is a subtle shift that is really, really important." Similarly, another teacher stated: "We are required to add learning targets. My students understood what was being taught. Knowing the targets they are more aware of their own learning. Teachers aim to achieve higher standards in order to achieve higher evaluation." Consistency in language and expectations was also commonly equated with supporting student learning as demonstrated by this quote, "One positive of everyone following the Danielson model is that we all use the same language and in many cases the same routines (i.e. clearly stated goals and objectives) and that helps the students learn better sim-

> " **Students benefit from a teacher who has set goals and reflects on them thoughtfully to show growth in a variety of ways.**"
>
> – Comment from Teacher Participant
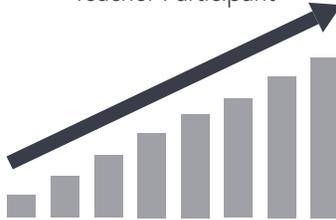
ply through consistency and repetition."

Alternatively, some teachers reported that the educator evaluation process had negatively impacted their students' learning citing a focus on data collection and compliance resulting in lack of autonomy for teachers. Comments such as the following capture this sentiment: "It has not impacted my students' learning in a positive way. If anything, it has hurt learning because more time is spent gathering data. Teaching time is lost in order to gather and score required data sets."

Many teachers wrote that the teacher evaluation process and student learning were unrelated or that they were unsure of any relationship between the two. Commonly, teachers stated that they see the teacher evaluation process and student learning as unrelated, instead sharing the sentiment that teachers impact student learning, not evaluations. The following quotes demonstrate this perception: "I doubt reading a document without any kind of accompanying training has probably dramatically altered what I was already doing on my own." and "Students are impacted by an evaluation system that supports growth, and acknowledges the accomplishments of the teacher. Contrarily, students will only benefit from a teacher being provided Professional Development in areas that require improvement. If neither happens, the evaluation is moot, and again punitive in nature. It's similar to assessing students without providing the necessary instruction to support their learning, or feedback needed to inform them of their successes."

Administrators' perspectives on evaluation and student learning. To determine administrators' perception of the role that the educator evaluation process plays in supporting student learning, we asked administrators to identify how the process has impacted student learning. Similar to the teachers, the administrators most commonly indicated that the process had positively impacted their students' learning and recognized this as a result of improved instructional practices. One principal shared, "Our teachers are more consistent in providing high quality instruction, due to the feedback provided through observations and walk-

throughs. Better quality instructional practices and attention to student performance have improved the results we get on assessments." Another principal wrote,

> I feel our core classroom practices have grown with a strong evaluation. Five years ago there wasn't necessarily a level of consistency in all classrooms. After a rigorous evaluation system has been in place, I observe in all classrooms some of the practices that we have worked on over the past four years directly related to our performance rubric.

Notably, principals often included the word "consistent" to describe the change that has occurred in instructional practices. In particular, instructional practices both within and across classrooms had become more "consistent" as a result of the evaluation system.

Principal responses also indicated that there was an increased focus on students and their involvement in the process and/or their own learning. One principal stated, "Students have a better understanding of what they are doing and why. It has increased the use of formative assessment, student self-assessment and goal setting. These things have impacted student learning." Woven throughout the principal responses was a recognition that the students' role in their own learning has shifted. This shift occurred through the improvement of instructional practices that positively impacted student engagement and students' involvement in achieving learning outcomes. This perception is demonstrated in the following quote, "Students have more opportunities to affect their own learning. Students who are somewhat successful are most strongly impacted by high quality teaching practices. Good teaching most profoundly impacts the large segment of 'typical' students who get the least attention."

The final notable response from principals is the acknowledgment of improved test scores. "I believe it has; our scores have gone up every year we have done it," and "Our test scores continue to increase as our fidelity within the systems continues to increase."

Alternatively, administrators also indicated that the evaluation processes either hasn't had an impact on student learning or they indicated that it was still too soon to tell based on the newness of the teacher evaluation process. Most often, unfamiliarity with the system was cited as the reason for no impact. This was evident in comments such as, "It's rather new to us and probably wasn't being used effectively in its early years, so we don't have the data yet," and "This

past year, I do not feel the system impacted student learning because the tool was new, and both I and my teachers were learning this new system."

Administrator and teacher perspectives on highly qualified teachers.

We also asked both the administrators and teachers if students whose teachers are rated as highly effective learn more than do students whose teacher receive lower ratings. Their responses were nearly even between "yes" and "no," with some indicating that it depended on circumstance or it was too soon to tell based on unfamiliarity with the new evaluation system. Those responding "no," often stated that the system was subjective and did not accurately label the best teachers as highly qualified. For example, one teacher remarked:

> No, they learn about the same. Since the evaluation system is subjective, in most cases, teachers are not properly evaluated by other administrators. Students who have teachers who are rated highly effective sometimes do learn more than their peers, and sometimes they learn less than their peers. Again, it depends on who the teacher's administrator is and how subjective their evaluation was.

Another teacher expressed a lack of confidence in the system, "No, I don't believe the ratings reflect the actual effectiveness. In fact, I believe there is an unspoken 'system' wherein teachers were moved up in their ratings after certain periods of time, regardless of actual performance." Likewise, one principal stated:

> No. I do not feel there is much of a difference between the students who learn with the teacher rated highly effective and one who is not. Most teachers have ceased to worry about the little hoops you need to be highly effective in my school.

Similar to the number of participants indicating "no," there was also a strong presence of respondents who shared a positive perception of the learning that occurs under highly effective teachers versus teachers receiving lower ratings. This perception is seen in comments such as, "I have had the privilege to work with several master teachers who are truly highly effective and their students do learn more than teachers who are rated effective." Many of the positive responses recognized instructional practices used by highly effective teachers to be a factor in student learning: "Yes, I believe they do. Teachers who strive to receive high ratings

are more likely to use good teaching practices that make sure their students are always learning"; "Yes, students are much more likely to be engaged from bell to bell. They are doing more one on one activities, more problem solving activities and demonstrate higher understanding on topics"; and finally, "Yes, They are able to achieve greater results with their students because they are differentiating and engaging students in deeper ways."

Teacher well-being. There was a significant difference in responses to items grouped as Teacher Well-being by teachers (mean ranking = 762.39) than administrator responses (mean ranking = 1172.88), c_2 (1, N = 1747) = 229.01, p < .01, h2 = .13. This is a medium effect size with 13% of the variability of responses accounted for by teachers versus administrators. Administrators perceived that the teacher evaluation process supported well-being more positively than teachers did as demonstrated by their responses to the items included in the teacher well-being group of questions.

When describing their increased stress, teachers frequently mentioned how the evaluation process has decreased collaboration.

described how the process has increased their stress and anxiety. Teachers made comments like: "It creates stress and negative feelings, which cause teachers to speak from emotion"; "It seems to bring stress to the building as a whole"; and "Higher stress and shorter fuses, especially at the end of the year."

Teachers suggested that the stress did not merely affect them personally, but also their teaching. For example, one teacher remarked, "I think that it cause teachers to stress out and worry too much about the system and not the students." Another teacher said, "It has created stress for everyone and draws our attention away from our students." Teachers described how the evaluation process makes them nervous and how that impacts their students. One teacher remarked, "The multiple meetings and paperwork and observations are scary. They make teachers nervous. When teachers are nervous, they are not at their best, and children suffer."

Teachers' perspectives on well-being. The educators' responses related to teacher well-being elicited the greatest difference in perspective between the teachers and the administrators. Two questions *(How has the teacher evaluation process at your school changed the relationships among teachers? How has the teacher evaluation process changed the culture of your school?)* yielded more than 2,000 teacher comments and nearly 1,000 administrator comments, which clearly revealed teachers do not perceive that the evaluation process supports their well-being.

A very small minority of teachers reported that the evaluation process improved teacher well-being. For example, one teacher declared, "Teachers have become more supportive in every way." Another teacher stated, "Some collaborate more with one another." Furthermore, a small minority reported that the teacher evaluation process has not affected teacher well-being or that they were unsure. The vast majority of teachers, however, described ways in which the evaluation process has affected them negatively.

Several themes emerged from the teachers' negative responses about the evaluation process. Frequently, teachers

When describing their increased stress, teachers frequently mentioned how the evaluation process has decreased collaboration. For example, one teacher remarked, "It has made people more stressed and some are more competitive or less likely to collaborate." Another teacher said, "Teachers feel like they must compete against each other now. This is not what is best for students. It invites corruption and will be less likely for teachers to collaborate and share."

Teachers described how the evaluation process increased competition, which they overwhelmingly viewed as negative rather than for its potential to improve their performance. One teacher commented, "It has pitted teachers against each other, created unfair "competition," and demoralized most teachers no matter what their rating is." Another teacher said, "Teachers are angrier, feel more backed against a wall, and it's much more competitive in a way that doesn't promote teacher growth or community."

Teachers explained how comparing themselves with others has undermined collaboration. For example, one teacher stated, "Teachers look at others and think that they may

be working just as hard but not getting recognized. It's a very non-collaborative culture." Another teacher remarked, "[The evaluation process] can cause people to pass blame and point fingers." Divisiveness was common theme, as demonstrated by comments like the following:

> There is a lot of pressure to perform well, and a lot of negative feelings when your class doesn't do well. There is a feeling of the "haves" and "have nots." Instead of coming together, teachers start doing whatever they can to raise their own scores and do not collaborate with other grade levels. Teachers become hesitant to listen to each other's ideas.

Teacher comparisons have lead many teachers to question the fairness of the process. One teacher remarked, "Ever since the rating system began some teachers have begun comparing what they do to others. This year with several ratings dropping it has led to some teachers questioning the system and how fair it is."

A commonly cited reason for unfairness was administrator bias. Teachers made comments like, "From my experience, my principal held everyone to different standards" and, "The principal plays favoritism." Teachers described how important it is to please their principals. Comments, like the following, revealed teacher perceptions of divisiveness and cliquishness:

> If you aren't in the principals "group" you might as well realize you won't be highly effective or effective if she has her way. Teachers are bragging that they are her tattletales as they walk around doing whatever they want because they don't have to worry about being targeted. They are all chummy. Our building is pitted teachers against one another. It used to be a family atmosphere, and now it's a divided mess.

Another teacher commented, "[The evaluation process] creates hostility and jealousy because the principal likes someone more than others. That's the reason why they received a good evaluation, not that they are a good teacher." Accordingly, the teachers reported that they worked to please their principals over what they believed to be good teaching. For example, one teacher stated, "We focus on what they are looking for in the lesson."

Overall, there were very few examples of teachers reporting that the teacher evaluation process increased their well-being. Comments like the following were not uncommon:

> I think overall, teachers are discouraged, stressed and afraid. It leads teachers to be secretive about how they really feel about things. Thus, our administrator hears one thing to her face, but teachers continually complain about anxiety, wanting to quit, cry behind closed doors......... Also it has led to teachers trying to step over one another in order to stand out and succeed. There is much less collaboration.

Teachers reported that they now like their jobs less. The following comment effectively captures the teachers' perceptions of the evaluation process on their well-being:

> There is so much more stress, anxiety, job dissatisfaction, secretiveness and much less collaboration. I do not recognize my school as it used to be. I am intrinsically motivated to do my best at all times, but it is not with the same energy that I used to have just a few years ago.

Administrators' perspectives on well-being. The administrators' responses about teacher-well-being revealed both similarities and differences compared to the teachers' responses. Like the teachers, approximately one-third of the administrators acknowledged that the teacher evaluation process has increased competitiveness and decreased collaboration. They described the stress teachers are under and the increased challenges they face because of the evaluation process. Another third of the administrators suggested that the teacher evaluation process has not had an effect on teacher well-being. Finally, unlike nearly all the teachers, one-third of the administrators reported that the teacher evaluation process has improved teacher well-being. We will explore some of these themes below, as contradictions.

Principals remarked that the evaluation process has "increased stress and anxiety." Some described how the process has created a "hostile environment" and has created an "us versus them" culture. One principal described how her school is a "scary and negative place during eval conversations." Another principal stated, "Teachers get really stressed out during evaluation time. This lends itself to an all-around negative atmosphere when we really should be celebrating student and teacher successes."

A number of administrators expressed a nuanced perspective on the role stress plays in the teacher evaluation process. Though these administrators recognized that the evaluation process increased teacher stress, they suggested that the stress can be positive. For example, one principal

remarked, "Our teachers have been stressed out about the change to a dramatically new system, even though most agree it is better." Another suggested that the evaluation process has "added stress, but fostered better conversations about teaching."

Some administrators proposed that the teacher evaluation process increased teacher collaboration. For example, one administrator said, "Teachers are more likely to feel like part of the conversation and contribute to the process." Another principal stated, "Teachers now share teaching strategies with each other and this is the key to success" and, "The teachers in my school are working together more collaboratively to share ideas and skills."

Modesty was absent from several administrators' responses, as they applauded their own roles in increasing teacher well-being. For example, one principal stated, "I think it's helped collaboration for the most part. But that's in part due to help from administration." Another suggested, "Teachers acknowledge and appreciate the support and feedback by the seasoned administrator." Some administrators acknowledged that teachers often "play the game" and seek to please their principal. One principal described how teachers quickly "figure out what your admin does and does not appreciate."

Similar to the teachers, the administrators described how the teacher evaluation process has increased competition among teachers. While some administrators dogmatically stated things like, "It has made teachers more competitive and less likely to help each other," other administrators were more nuanced in their replies, stating things like, "It has become more competitive rather than collegial in some instances" and, "There is an unhealthy competitive nature at times." According to administrators, comparison with others was the cause of diminished teacher well-being. For example, one principal remarked, "They don't want to share ideas with each other when they are worried that someone could be better at their job than they are." Another principal stated, "It can cause competition and bitterness if people compare their scores with each other. We strive for collaboration, but I think the evaluation process can hinder that when you start rank ordering teachers by a number score."

> "
> There is so much more stress, anxiety, job dissatisfaction, secretiveness and much less collaboration."
>
> – Comment from Teacher Participant

Administrators acknowledged the high-stakes culture that the teacher evaluation process has created. For example, one principal noted, "It has made some of them more competitive and proprietary about their teaching practices; it is a competition because if there are layoffs, the bottom person is the one let go." Another principal explained, "It created negative feelings and finger pointing at each other. I feel this could have been the case as well at my school if there were layoffs at the elementary level."

In a vast departure from the teacher responses, the administrators regularly cited how competition caused by the evaluation system is a force for good in the teachers' lives. For example, one administrator commented, "Teachers collaborate when preparing for end of year evaluations, they sometimes have healthy competition." Another principal suggested that competition was greater in the past but has since subsided:

> I think humans are competitive by nature. As we've moved to a system that requires us to label people "effective" vs "highly effective" it has created some friction. That said, we've been doing this long enough, much of that has ebbed away.

Administrators were often wishful in their descriptions of how teacher tension will dissipate over time. For example, one principal offered, "There always seems to be some competition. You want to see this drive teachers to be better at their craft and not create conflict among themselves."

Although the administrators expressed a more positive view of how competition influenced teacher well-being, they did not hold back their criticisms of teacher jealousy and lack of ability to accurately self-assess. For example, one principal stated,

> Teachers are a tough group. Many don't grasp the idea that teaching is a profession and often narrowly self-assess their teaching abilities. Being told that you are basic or proficient is not good enough when they know that distinguished is the top-level on the score sheet. Some teachers become defensive when evidence is provided illustrating they are not where they should be or want to be on the rubric.

The administrators addressed "bad apple" teachers and suggested that they tend to view the process differently from highly effective teachers. For example, one principal stated,

> Most of my teachers rate very high on the areas of assessment in the Charlotte Danielson review; however, when a teacher does not, the tool is viewed as arbitrary and subjective by those who score as ineffective. Also, the less than effective teachers tend to display jealousy toward those who are more successful and sometimes those who are more successful seem to be judgmental of the ones who need their support and guidance.

Likewise, one principal said, "Teachers question why they are not highly effective. That negatively impacts the culture." However, according to the principals, the well-being of highly effective teachers has not changed. As one principal said, "For teachers who are historically effective, the teacher evaluation process hasn't changed their relationships."

Overall, the administrators expressed mixed perspectives on how the teacher evaluation system has affected teacher well-being. Though a majority of the principals described the evaluation process as negative or neutral on teacher well-being, approximately one-third of the administrators described how competition and accountability associated with the process improved collaboration and student learning. Nonetheless, none of nearly 1,000 comments suggested that the process made the teachers more comfortable, healthy, or happy.

Critical feedback and directions for professional development. There was a significant difference in responses to items grouped as Teacher Improvement by teachers (mean ranking = 746.36) than administrator responses mean ranking = 1215.81), $c2$ (1, N = 1747) = 299.52, $p < .01$, $h2$ = .17. This is a medium effect size with 17% of the variability of responses accounted for by teachers versus administrators. Administrators perceived that the teacher evaluation process provided critical feedback more than teachers did

> "
> Teachers question why they are not highly effective. That negatively impacts the culture."
>
> – Comment from
> Administrator Participant

as demonstrated by their responses to the items included in the teacher feedback group of questions. To gain further insight in how educators perceive the teacher evaluation process provides teachers with critical feedback and professional development on how they can improve their own practice, we asked both teachers and administrators to respond to questions related to these two concepts.

Teachers' perspectives on critical feedback. Effective and critical feedback needs to be individualized to reflect grade level, content area, and specific teacher need (Hill & Grossman, 2013). We asked teachers if their school's educator evaluation system was personalized for their position or teaching assignment. The majority of teachers felt that the system, including the process used, the rubric, and the required data components were not flexible enough to be personalized. Teachers stated that the educator evaluation tools seemed to reflect a focus on the general education setting making it challenging for specials teachers (art, music, physical education, etc.), early childhood/kindergarten teachers, school counselors, vocational/career and technical education teachers, interventionists and special education teachers (categorical, resource room, teacher consultants, RESA program teachers) to receive effective and critical feedback from the evaluation process. Additionally, teachers noted the need for educator evaluation systems to address the diversity present in classroom. For example, one teacher wrote, "Some teachers teach AP classes, some teach remedial classes, or bilingual or special ed students. But we are all evaluated in the same way. I really feel like a more personalized approach would be better for everyone."

Administrators' perspectives on critical feedback. Administrators overwhelmingly indicated that the educator evaluation process resulted in positive professional growth and that the process drove professional development decisions. For example, one administrator stated:

> Teachers have been more likely to converse with other teachers about professional practices. We also try to provide more targeted professional learning.

Teachers are also more willing to be reflective about their teaching practices. Those that are unwilling to grow have been impacted by a lower effectiveness rating.

Another principal positively described the feedback process as follows:

The teachers know what they need to focus on and can advocate for PD that is meaningful for them to achieve their goals. They discuss these items to focus on at PLCs throughout the year. The evaluation and student assessment data is used to plan for upcoming PD.

We also asked administrators if the school's educator evaluation system was personalized for their teachers. Overall, administrators wrote that the system -- Danielson, Marzano, etc. -- was not personalized to meet individual teachers' needs. One administrator wrote,

I don't think it is personalized for teachers. I think it's very generic and does not lend itself to be used as a growing tool but as a compliant tool. Teachers should be able to use it to help themselves instead of feeling like they have to comply to it.

However, beyond the rubric or the tool used within the educator evaluation process, opportunities for collaborative goal setting between the teacher and the administrator and opportunities for teachers to make individualized choices about components of the broader evaluation process were noted. Administrators wrote that teachers could set their own goals, identify the evidence to be used to document achievement of goals, determine the data measures for student growth, influence observation foci through the pre-conference process, and use self-assessment tools to personalize the evaluation process.

BARRIERS TO EFFECTIVE TEACHER EVALUATIONS

On the questionnaire and the electronic interviews, we asked the administrators and teachers to describe the barriers to an effective evaluation process. Overwhelmingly, the participants identified lack of time as the primary barrier. The following comment from a teacher encapsulates what many respondents expressed:

Our administrators simply do not have the time to conduct proper classroom observations. There

are too many standardized tests to be organized, discipline issues to be dealt with, building maintenance problems to be handled, etc. There is just too much expected of them, and not enough time to do it all.

Despite teachers' disdain for several aspects of the teacher evaluation process, none of them referred to their administrators being too busy as a positive. Rather, the teachers expressed a desire to have their principals spend more time in their classrooms. For example, one teacher said, "Our administrators don't have time to meet with us or see us in action accept when we're being observed. I'd like to see him more."

Administrators described several ways in which lack of time limited their ability to evaluate teachers. They noted how the evaluation process should include a pre-conference, an observation, and a post-conference; however, that process is often accelerated or ignored because of time constraints. For example, one administrator remarked:

It's impossible. You are supposed to have a conference before and after the evaluation. In a perfect world, that would be great. But, when you have one principal that has to deal with discipline, professional development, and other areas of a school...there is only so much time you can get in teacher evaluation.

Administrators also described how the teacher evaluation mandates require them to spend time evaluating effective teachers. Once principal said, "I don't get to spend the extra time with less effective teachers in a coaching capacity because I am using that time to evaluate teachers that are doing a great job." Likewise, one principal quipped, "We spend too much time telling teachers what they need to do rather than helping them do it."

A number of principals mentioned the evaluation tool specifically, with comments like, "The evaluation tool is very long and extensive and requires a lot of time to complete." Another principal remarked, "It takes too long to learn the new evaluation tool. We need that time to sit down with teachers and have conversations over instruction and student growth data." The administrators reported that they liked observing teacher and helping them to improve, but they despised the paperwork. One principal remarked, "All the forms and bureaucracy get in the way of simply helping teachers to reflect on their practice."

After citing lack of time as the primary barrier to effective teacher evaluations, respondents expressed how the process is often inconsistent or inauthentic. Teachers remarked on how the process is too subjective. For example, one teacher said that a barrier is "subjective ratings which are influenced by favoritism, and negative feelings toward individual teachers." Another teacher expressed, "Some administrators are more thorough than others. Sometimes you get observed on one day a year and your whole evaluation depends on that." Teachers commonly expressed a desire to be evaluated by an outside evaluator or to participate in peer evaluations, rather than to be evaluated by their principals. For example, one teacher said, "I wish we used third-party evaluators who didn't know any backgrounds or have any biases." Another teacher stated, "We should be doing peer evals. They would be much more beneficial for both teachers, and to be honest, my peers know much more about good teaching than my principal." A common sentiment was that principals are often required to evaluate teachers in subjects or grade levels in which they are lacking experience or expertise.

Teachers also expressed frustration with the scoring system and the limits placed on those who are able to receive ratings of highly effective. For example, one teacher said:

> I don't think it's fair to say new teachers should never be marked highly effective because they are new and still learning. I don't think it's fair that there is only a certain number of teachers who can be marked highly effective.

Ostensibly, in some districts, "The administration is marked down by the superintendent if they give too many highly effectives." In other districts, teachers blamed their union leadership for placing limits on their ratings, which in turn, demotivated those teachers:

> Since it is set up for almost everyone to be rated effective, it may as well not exist. The union negotiated it so that it is virtually impossible to be rated highly effective. Since there is no chance, why knock yourself out?

In addition to expressing that the teacher evaluation process is often inconsistent and inauthentic, administrators and teachers complained that unclear and inadequate evaluation criteria were a major barrier to effective teacher evaluations. For example, teachers made comments like, "The descriptions are too ambiguous," and, "The rubric should be made more clear. There are a lot of grey areas that are left for interpretation." Contradicting the administrators and teachers who complained about the tool being too long and cumbersome, these educators sought greater clarity and detail.

Much of their concern, however, related to the diverse needs of teachers and the perceived futility of using one tool or rubric for all teachers. For example, one teacher remarked, "The one-size-fits-all evaluation rubric is a barrier. For example, special education teachers are evaluated using the same rubric as general education teachers although their jobs are very different." This administrator expressed how universalizing the evaluation process is flawed:

> The cut and dry whole number ratings can make a "messy" and complicated profession seem much more simplistic than it really is. Teaching is a difficult profession with a lot of variables and moving parts. Our evaluation systems simplifies this process and makes it difficult for administrators to account for all of the difficult variables in the classroom that the teachers need to deal with on a daily and yearly basis.

Administrators and teachers alike frequently also criticized the process for its overreliance on standardized test scores. Teachers made comments like, "Teachers are basically required to set goals based on NWEA scores even when other goals would be better." An early-elementary teacher complained:

> The process should focus less on data collection and more emphasis on teaching the whole child. Each student has the potential to achieve their own personal goals. However, testing a first grader to see if they are college ready is simply ludicrous. They are six!!! We need to support their social, emotional and academic growth!

Moreover, teachers shared that they were often evaluated on students' test scores unrelated to the subjects they teach. For example, one teacher remarked, "I am evaluated based on standardized tests on a subject I don't even teach, and since the subject I teach does not have a lot of data sources (SAT or NWEA), there is no way to do it more effectively."

Another theme related to educators' perceptions of the barriers to effective teacher evaluation was teacher resistance and mistrust. Teachers and principals alike mentioned lack of trust in each other and in the process frequently. Teachers made comments like, "The current administration does
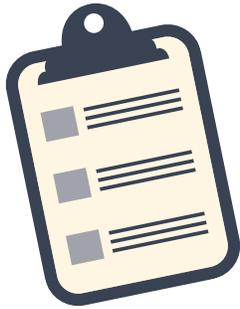
In addition to expressing that the teacher evaluation process is often inconsistent and inauthentic, administrators and teachers complained that unclear and inadequate evaluation criteria were a major barrier to effective teacher evaluations.

not trust the teachers," and, "Our administration is terrible with criticism, communication, and providing feedback in a timely manner. This makes the staff frustrated and causes a sense of animosity regarding the evaluation system."

The administrators, however, made the most biting comments. Administrators made comments like, "I don't think I have the teacher buy-in that I'd like. I wish they would see the value in the process instead of seeing it as an attack," and, "Some teachers don't feel they need to grow and just simply don't buy into the research that goes into the rubrics we are using to increase teacher effectiveness which makes it difficult for those teachers to grow." A number of administrator comments like the following were directed at late-career teachers:

> Teacher attitude. So many teachers are so against getting feedback because all they think about his how the state says this and the state says that. They aren't taking into consideration that the evaluation could actually be a good tool for them. I feel like older teachers are the ones with this attitude and they don't grow. They seem to think since they've taught so long they have no room left to improve. What they don't realize is that today's students are different than students 10 years ago.

Some administrators acknowledged how the process contributes to teacher stress. For example, one principal remarked, "Too often, even my most trusting staff still demonstrate a form of PTSD from what they hear their policy makers say about their profession. As a result, this leads to a less than fully authentic experience for most staff." Another principal eloquently summarized this theme, "The 'done-on-to' experience eroded an already fragile trust between teachers and administration."

SOLUTIONS FOR IMPROVING THE TEACHER EVALUATION PROCESS

We also asked the teachers and administrators for their remedies to the barriers they described. Expectedly, their most commonly described remedies corresponded with their most commonly described barriers. For example, they explained how the teacher evaluation process could be improved by finding ways to give teachers and administrators more time. Solutions for freeing up more time so principals could spend more time on the teacher evaluation process varied. Some suggested a shift from the traditional observation format to the walk-through model. One teacher remarked, "I just wish our principals had more time to come in casually so they could see us in action at different points of the year and doing different subjects." An administrator suggested, "I would shift to more frequent and less-formal mini-observations." A number of administrators expressed a desire to forgo observing teachers who have already proved to be highly qualified. One principal suggested, "If I didn't have to observe my exemplary teachers every year, I could spend more time on the teachers who need it."

A number of teachers and administrators mentioned streamlining the evaluation process to save time. For example, one teacher remarked, "The evaluation system needs to be streamlined and made less cumbersome for both teachers and administrators." This administrator noted the paperwork burden:

> The biggest barrier that needs to be addressed in order to make the evaluation system more effective is streamlining the process. There is a lot of repetitive paperwork, if we could streamline the process, it would be much more effective.

Others suggested shortening the pre and post conferences to save time.

Most commonly, administrators and teachers suggested either lessening the principals' duties or hiring more administrators. Ideas for reducing principal duties included decreasing the number of meetings and eliminating lunch supervision. For example, one principal expressed the following:

Administrators also need to be afforded the time support to be in classrooms making observations throughout the instructional day to provide specific feedback. Current district expectations are that administrators provide direct student supervision throughout the lunch period. In a building of 500 4th-6th students, this consumes 2.5 hours during the time that instruction is occurring.

Many administrators described how they wish they could spend more time on evaluations but were short-staffed. For example, one principal stated:

> I am the lone administrator in my building and wear many hats, making it difficult to do the evaluations the way I feel they should be done. I would love to get in every teacher's classroom at least ten times a year, but it can be difficult with everything else I have to do.

Despite the widespread sentiment that schools needed more administrators in order to conduct teacher evaluations properly, few were optimistic about that solution. Most made comments like the following, "If another administrator were hired, that might make a difference. But with the budget constraints placed on public schools, that is highly unlikely." Another principal remarked, "The time would only come from having more manpower and manpower takes money." The educators recognized that their problems were not unique. For example, one teacher stated, "Money is an ongoing issue for all districts. I wish I did have the answers."

Fortunately, the educators provided a number of low-cost or cost-free solutions for improving the teacher evaluation process. The administrators and teachers commonly described how the process could be made more collaborative. Teachers, in particular, wanted to be included more in the process. For example, one teacher suggested, "Involve teachers more in the process. It should be a team effort, not just a top-down judgment." Increasing trust through collaboration was a common theme. For example, one teacher stated, "Trust that teachers are invested in growing! Having a system that is more organic would make teachers more invested and create more growth. It would allow teachers to really take positive risks."

Teachers who were critical of the process typically were not opposed to having their teaching evaluated. Rather they wanted the process to be more authentic and meaningful. One teacher suggested, "Make the evaluative process more engaging and interactive if you really want it to make a change. I wouldn't mind if my administrator recorded me and we watched it together to pick apart and evaluate the lesson." A number of teachers suggested peer observations. For example, one teacher recommended, "Look at evaluations from a team perspective and provide more peer coaching opportunities."

The teachers wanted to help make the evaluation process transparent and reliable. One teacher stated that her school should, "come up with what everything means together. Everyone understands them differently. A committee should be formed to go over and understand what the rubric means." Teachers expressed a desire to have a shared understanding of expectations. For example, one teacher recommended, "Review the rubric more as a staff. Give more examples of highly effective in the 5d. If teachers had an opportunity to visit classrooms that are highly effective and observe that would help." Teachers wanted clarity and consistency, as illustrated by this comment:

> Teachers who were critical of the process typically were not opposed to having their teaching evaluated. Rather they wanted the process to be more authentic and meaningful.

Make each observation meaningful and clear as to what exactly they are looking for. It is often a mystery until the final discussion as to what type of evaluation a teacher will receive. This year alone the scale was changed and teachers were not notified until the end if it affected their rating.

Similar to suggestions related to increased collabora-

tion, the educators expressed the need for increased teacher training and resources. For example, one administrator stated, "I wish there was time and resources for the teachers to have the same training in the Danielson model." A teacher remarked, "We should get training on what they are expecting on the evals. If that's how they want us to teach, they should show us what they mean." Teachers frequently suggested using their mandated professional development days for the training. For example, one teacher wrote, "I think the principals and administrators should use the evaluation to provide appropriate professional development time to address these goals. Appropriate resources and professional development directly related to the goals set out in the evaluation tools." Another teacher suggested, "Choose a focus each year, provide professional development opportunities, and group teachers according to needs and interests for personalized learning." Teachers bemoaned professional development sessions that did not address their needs. Instead, they recommended targeted professional development, as expressed by this teacher, "Administrators need to continue to identify trends in the evaluation data to be more specific with our PD and show staff that we want to help them improve."

A final theme from the teachers' suggestions for removing barriers was their desire to improve the evaluation tool. A number of educators made comments like the following: "The rubric should be tailored for early elementary, upper, elementary, and specific specials classes." Another teacher suggested, "New rubrics should be made for each grade level based on research that shows what is most effective for those aged students." Although administrators mentioned their lack of subject- and grade-specific experience and knowledge when discussing barriers, few suggested the creation of more evaluations tools or rubrics to overcome those barriers.

8 DISCUSSION

The analyses of this study revealed valuable yet not unprecedented findings whereby Michigan administrators were shown to be on trend with those nationwide (Kraft & Gilmour, 2017; Louwes, 2017). Consistent with the findings of Kraft and Gilmour (2017), who also studied Michigan teachers and found that 97% received effective or higher ratings, 99% of the teachers in this study were rated as highly effective or effective in 2017. Corroborating the findings of Ford et al. (2017), this study found that highly effective/effective ratings do not determine teachers' job satisfaction and well-being. While participating administrators felt the evaluation process resulted in positive professional growth, participating teachers experienced the contrary. Participating teachers perceived the added stress, competitive culture, and distrust of the system as interferences in job satisfaction.

Furthermore, such high ratings seemingly undermine participating administrators' claims that the teacher evaluation process creates healthy stress and competition among teachers, putting pressure on apathetic or oppositional teachers. However, as Berliner and Glass (2014) have explained, competition is anything but healthy when teachers feel they have been given a Hobson's choice on actions that could save their evaluations and therefore their jobs (Amrein-Beardsley & Collins, 2012; Koretz, 2008).

Despite the differences in participating administrators' perceptions and teachers' perceptions on the purpose of teacher evaluations, both groups identified time as a barrier to successful implementation of the process. As with the findings in other studies (Lavigne & Chamberlain, 2017; Ramirez et al., 2014), participating administrators lacked sufficient time to conduct multiple observations and hold meetings with teachers while participating teachers cited the evidence component as time-consuming and overwhelming.

Consistent with prior research (Lavigne & Chamberlain, 2017; Norris et al., 2017; Reid, 2017; Rowan et al., 2013), administrators and teachers in this study expressed concerns over administrators' adeptness with conducting evaluations. Administrators felt they were still new to the evaluation tools and therefore, still learning the evaluation criteria. Teachers lacked confidence in administrators' abilities to evaluate unbiasedly and in administrators' knowledge of effective pedagogy for given content areas. Moreover, teachers, sharing Darling-Hammond et al.'s (2012) concerns, questioned the overreliance on standardized test scores and echoing Hill and Grossman (2015), criticized the one-size-fits-all evaluation rubrics.

This study, congruent with other research, highlights the many facets of teacher evaluation that have limited the extent to which the process can actually contribute to teacher well-being and influence the quality of instruction in Michigan schools.

9 CONCLUSIONS AND RECOMMENDATIONS

In its support of new mandates (PA 173, 2015) for evaluating teachers, the Michigan Department of Education claimed that the new teacher evaluation process would "support both student learning as well as educator well-being" and would "provide teachers with critical feedback on how they can improve their own practice to impact the lives of students" (Michigan Department of Education, 2016, p. 4). This comprehensive study of the perceptions of a stratified convenience sample of 1,746 public school educators (teachers; n = 1274) and administrators; n = 474) from Michigan's 56 Intermediate School Districts revealed that MDE's claims are not being met; and, arguably, evidence suggests that the new teacher evaluation mandates have had the opposite effect they intended.

On a scale from -10 (strongly disagree) to +10 (strongly agree), the median administrator scores were 3.00 for student learning, 4.75 for teacher improvement, and 2.6 for teacher well-being. The administrators' perceptions of the effectiveness of the evaluation system at meeting the espoused goals were lukewarm at best. The teachers' median scores were -1.00 for student learning, .94 for teacher improvement, and -.85 for teacher well-being. In two categories, teachers reported the teacher evaluation process has had a negative effect.

Fewer than half of teachers and two thirds of administrators stated that Michigan's teacher evaluation process in its current form is improving teaching and learning.

Moreover, the teachers' scores in all three areas were significantly lower than administrators' scores. In other words, there is a chasm between what the administrators think about the teacher evaluation process and what the teachers think. Qualitative data from thousands of questionnaire and interview responses from this study confirms this disconnect. While the administrators expressed greater variation in their responses, the teachers were overwhelmingly negative about the impact of the evaluation system.

The system is not working as it was intended, and it is not working the way administrators and teachers proclaimed it should be. Fewer than half of teachers and two thirds of administrators stated that Michigan's teacher evaluation process in its current form is improving teaching and learning, compared with their nearly unanimous (97% and 96%

respectively) assertion that improving teaching and learning should be the goal of the evaluation process.

Educator discontent with the teacher evaluation process is problematic beyond its failure to meet the espoused goals. Teacher shortage is a growing concern in Michigan, and a reduction in teacher well-being is likely to make the profession less attractive to teacher candidates and increase teacher turnover. The teachers and administrators in this study commonly described the negative impact the teacher evaluation system has had on their job satisfaction. Low job satisfaction is a primary reason teachers leave the profession (Perrachione, Petersen, & Rosser, 2008), and teacher perceptions are most accurate predictor of job satisfaction (Cunningham, 2016). Student achievement in schools with higher teacher turnover is significantly lower than in schools with less teacher turnover (Ronfeldt, Loeb, & Wyckoff, 2013). Accountability measures, particularly their impact on teacher well-being, have increased the number of teachers leaving the profession (Ingersoll, Merrill, & May, 2016). What's more, minority teachers have been leaving the profession at a higher rate than non-minority teachers (Ingersoll & May, 2016).

With nearly 99% of teachers receiving an effective or highly effective rating in 2017, yet commonly reporting that they collaborate less, have lower job satisfaction, and do not perceive that teacher evaluation process increases student achievement, the system is clearly broken. Teachers and administrators alike long for more time and resources to collaborate on improving teaching and student learning. However, as frequently noted by the educators in the study, hiring more administrators or increasing related resources is impeded by school budget limitations. Likewise, additional time for collaboration is unlikely to come from an increase in instructional hours, whether mandated by the legislature or negotiated locally. Therefore, educational leaders must find ways to free-up administrator and teacher time by reducing less-important tasks.

For teachers, restructuring their professional development time would be a good start. Professional development in its
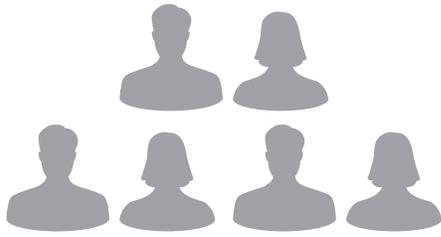
> Teacher shortage is a growing concern in Michigan, and a reduction in teacher well-being is likely to make the profession less attractive to teacher candidates and increase teacher turnover.

most common forms – workshops and trainings – is widely viewed as ineffective (Hill, 2009). Devoting the majority of their mandated professional development time to collaborating toward improving teachers' instruction and well-being, as well as increasing student learning would likely mediate some of the barriers the teachers corroborated.

For administrators, reimagining their workday and task responsibilities could allocate more of their time for instructional leadership. Administrators' workdays often consist more of managerial duties than leadership. "Administrators need to find better ways to involve teachers and other staff members and to help them adjust conflicting interests" (Murphy, 2013, p. 36). If principals could transfer some of their tasks to other school employees, they could recoup much-needed time for making the teacher evaluation process more effective. Take, for example, one task most principals undertake: lunchroom supervision. If a principal who spends two hours per day supervising lunch were to delegate that duty, she would have 360 more hours per year to provide teachers with critical feedback to support student learning and educator well-being, as PA 173 intended.

REFERENCES

Allegretto, S. A., & Mishel, L. (2016). The Teacher Pay Gap Is Wider than Ever: Teachers' Pay Continues to Fall Further behind Pay of Comparable Workers. *Economic Policy Institute.*

Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS® EVAAS®) in the Houston Independent School District [HISD]: Intended and unintended consequences. *Education Policy Analysis Archives, 20*(12). Retrieved from http://epaa.asu.edu/ojs/article/view/1096/982

Berliner, D. C., & Glass, G. V. (2014). *50 myths & lies that threaten America's public schools: The real crisis in education.* New York, NY: Teachers College Press.

Bill and Melinda Gates Foundation (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study.* MET Project Policy & Practice Brief. Bill & Melinda Gates Foundation.

Borman, G. D., & Kimball, S. M. (2005). Teacher quality and educational equity: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *The Elementary School Journal, 106*(1), 3-20.

Cantrell, S., & Kane T. J. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study.* Bill & Melinda Gates Foundation.

Center for Educational Leadership. (2015). Retrieved from https://www.k-12leadership.org/.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity.* Washington, D. C.: U. S. Government Printing Office.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Cunningham, S. L. (2016). *A quantitative analysis of the factors associated with teacher attitudes and perceptions towards job satisfaction* (Doctoral dissertation, Seton Hall University).

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2007). Enhancing professional practice: A framework for teaching (2nd ed.).Alexandria, VA: Association for Supervision and Curriculum Development. Retrieved from http://www.ascd.org/publications/books/106034.aspx.

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

Eun, B. (2008). Making connections: Grounding professional development in the developmental theories of Vygotsky. *The Teacher Educator, 43*, 134-155. doi:10.1080/08878730701838934

Every Child Succeeds Act (EESA) of 2015, Public Law No. 114-95, S.1177, 114th Cong. (2015). Retrieved from https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher, 43*(2), 100-107. doi:10.3102/0013189X14521864

Ford, T. G., Van Sickle, E. V., Clark, L. V., Fazio-Brunson, M., & Schween, D. C. (2017). Teacher self-efficacy, professional commitment, and high-stakes teacher evaluation policy in Louisiana. *Educational Policy, 31*(2), 202-248. doi:10.1177/0895904815586855

Friedman, I. A. (1999). Teacher-perceived work autonomy: The concept and its measurement. *Educational and Psychological Measurement, 59*(1), 58-76.

Gangé, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior, 26*, 331-362. doi:10.1002/job.322

Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*(2), 371-384.

Hill, H. C. (2009). Fixing teacher professional development. *Phi Delta Kappan, 90*(7), 470-476.

Hill, H. C., Kapitula, L., & Umland, K. (2010). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal, 48*(3), 794-831. doi:10.3102/0002831210387916

Hunter, M. (1980). Six types of supervisory conferences. *Educational Leadership, 37*(5), 408-412.

Hunter, M. (1984). Knowing, teaching, and supervising. In P. Hosford (Ed.), *Using what we know about teaching* (pp. 169-192). Alexandria, VA: Association for Supervision and Curriculum Development.

Ingersoll, R., & May, H. (2016). Minority Teacher Recruitment, Employment and Retention: 1987 to 2013. *Learning Policy Institute, Stanford, CA.*

Ingersoll, R., Merrill, L., & May, H. (2016). Do Accountability Policies Push Teachers Out? *Educational Leadership, 73*(8), 44-49.

Ingersoll, R. M. (1999). The problem of underqualified teachers in American secondary schools. *Educational Researcher, 28*(2), 26-37.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal, 38*(3), 499-534.

Jewell, J. W. (2017). From inspection, supervision, and observation to value-added evaluation: A brief history of U. S. teacher performance evaluations. *Drake Law Review, 65*(2), 363-419.

Karier, C. J. (1986). *Scientists of the mind: Intellectual founders of modern psychology.* Chicago, IL: University of Chicago Press.

Koretz, D. (2008). *Measuring up: What educational testing really tells us.* Cambridge, MA: Harvard University Press.

Kraft, M.A. & Gilmour, A.F. (2017). Revisiting the Widget Effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher, 46*(5), 234-249.

Lavigne, A. L., & Chamberlain, R. W. (2017). Teacher evaluation in Illinois: School leaders' perceptions and practices. *Education Assessment, Evaluation, & Accountability, 29*(2), 179-209. doi:10.1007/s11092-016-9250-0

Learningforward (2017, October 15). Definition of professional development. Retrieved from https://learningforward.org/who-we-are/professional-learning-definition?printToPDF=true

Learning Sciences Marzano Center (2016). *The research base for the Marzano teacher evaluation model and correlations to state VAM.* West Palm Beach, FL: Learning Sciences International.

Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research. *Quality & Quantity, 43*, 265-275.

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher, 35*(3), 3-14.

Loewus, L. (2017). Principals are loathe to give teachers bad ratings: Most teachers are still rated as effective. *Education Week, 36*(27), 7.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Michigan Council for Educator Effectiveness (2013). *Building an improvement-focused system of educator evaluation in Michigan: Final recommendations.* Retrieved from http://www.mcede.org/reports

Michigan Department of Education (2015). *Michigan educator evaluations at-a-glance.* Retrieved from https://www.michigan.gov/documents/mde/Educator_Evaluations_At-A-Glance_522133_7.pdf

Michigan Department of Education (2016). *Michigan department of education strategic plan.* Retrieved from http://www.michigan.gov/documents/mde/10_in_10_Action_Plan_543856_7.pdf

Murphy, J. (2013). The unheroic side of leadership: Notes from the swamp. In M. Grogan (Ed.), *The Jossey-Bass reader on educational leadership* (3rd ed.) (pp. 28-39). San Francisco, CA: Jossey-Bass.

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

Norris, J., van de Mars, H., Kulina, P., Amrein-Beardsley, B., Kwon, J., & Hodges, M. (2017). Physical education teacher perceptions of teacher evaluation. *The Physical Educator, 74*(1), 41-62. doi:10.1866/TPE-2017-V74-I1-6882

Onwuegbuzie, A. J., & Leech, N. L. (2004). Enhancing the interpretation of significant findings: The role of mixed methods research. *The Qualitative Report, 9*(4), 770-792.

Pearson, L. C., & Moomaw, W. (2005). The relationship between teacher autonomy and stress, work satisfaction, empowerment, and professionalism. *Educational Research Quarterly, 29*(1), 38-54.

Perrachione, B. A., Petersen, G. J., & Rosser, V. J. (2008). Why do they stay? Elementary teachers' perceptions of job satisfaction and retention. *The Professional Educator, 32*(2), 1-18.

Peterson, P. E. (2010). *Saving schools: From Horace Mann to virtual learning.* Cambridge, MA: The Belknap Press of Harvard University Press.

Ramirez, A., Clouse, W., & Davis, K. W. (2014). Teacher evaluation in Colorado: How policy frustrates practice. *Management in Education, 28*(2), 44-51. doi:10.1177/0892020613511264

Reid, D. B. (2017). U. S. principals' interpretation and implementation of teacher evaluation policies. *The Qualitative Report, 22*(5), 1457-1470.

Ripmaster, C. [Colin Ripmaster]. (2014, December 12). *Overview of CEL's 5 Dimensions of Teaching and Learning* [Video file]. Retrieved from https://www.youtube.com/watch?v=tRAjgUBkTbs

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal, 50*(1), 4-36.

Rossii, P. H., & Freeman, H. E. (1993). *Evaluation: A systemic approach.* Newbury Park, CA: Sage.

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

Ryan, R. M. & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology,* (25), 54-67. doi:10.1006/ceps.1999.1020

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.*

Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sanders, W., & Horn, S. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*(3), 299-311.

Schleicher, A. (Producer). (2017). *What does PISA reveal about teacher policy and practice* [video webinar]. Retrieved from http://www.oecd.org/pisa/webinars/

Shulman, L. S. (1992). Research on teaching: A historical and personal perspective. In F. K. Oser, A. Dick, & J-L. Patry (Eds.), *Effective and responsible teaching: The new synthesis* (pp. 14-34). San Francisco, CA: Jossey-Bass Publishers.

Silver Strong & Associates (n.d.). *The simple and deep approach to teacher evaluation.* Franklin Lanes, NJ: Thoughtful Education Press. Retrieved from http://www.thoughtfulclassroom.com/TCTEF_flipbook/#p=1

Strong, L. E. G., & Yoshida, R. K. (2014). Teachers' autonomy in today's educational climate: Current perceptions from an acceptable instrument. *Educational Studies, 50*(2), 123-145. doi:10.1080/00131946.2014.880922

Tucker, M. S. (2011). Standing on the shoulders of giants: An American agenda for education reform. *National Center on Education and Economy (NJ3).*

Tuytens, M., & Devos, G. (2017). The role of feedback from the school leader during teacher evaluation for teacher and school improvement. *Teachers and Teaching, 23*(1), 6-24. doi:10.1080/13540602.2016.1203770

U. S. Department of Education (2016, June 6). Race to the Top funds. Retrieved from https://www2.ed.gov/programs/racetothetop/index.html

United States. National Commission on Excellence in Education. Department of Education. (1983). *A nation at risk: A report to the nation and the Secretary of Education, United States Department of Education.* Washington, D. C.: The Commission: [Supt of Docs., U. S. G. P. P distributor].

United States. National Commission on Excellence in Education. Department of Education. (1983). *A nation at risk: the imperative for educational reform: a report to the Nation and the Secretary of Education, United States Department of Education.* Washington, D.C.: The Commission: [Supt. of Doc., U. S. G. P. O. distributor]

Value-added measures (2013). The glossary of education reform. Retrieved from http://edglossary.org/value-added-measures/

Visscher, A. J., & Coe, R. (2003). School performance feedback systems: Conceptualisation, analysis, and reflection. *School Effectiveness and School Improvement, 25*, 509-530.

Wei, R. C., Darling-Hammond, L., & Adamson, F. (2010).

Professional development in the United States: Trends and challenges. Dallas, TX: National Staff Development Council.

Wells, K. (2016, January 27). *Teachers suspended from Wayne County school, accused of helping students cheat on state test.* Michigan Radio. Retrieved from http://michiganradio.org/post/teachers-suspended-wayne-county-school-accused-helping-kids-cheat-state-test

Wise, K. C., Darling-Hammond, L., McLaughlin, M., & Bernstein, H. (1984). *Teacher evaluation: A study of effective practices.* Santa Monica, CA: RAND.

APPENDICES

Appendix A
Reliability Coefficients for Teacher Improvement Survey Items

Sixteen survey Items for Teacher Improvement (α = .89)	M	SD	r
The teacher evaluation system implemented by my school provides a platform to recognize excellent teachers for their best practices.	1.34	6.22	.71
The teacher evaluation system implemented by my school leads to more innovative teaching.	- .85	6.15	.72
The teacher evaluation system implemented by my school is an effective tool for helping teachers to improve their performance.	1.40	6.30	.75
The teacher evaluation tool implemented by my school results in objective and actionable feedback.	.31	6.47	.74
The teacher evaluation tool implemented by my school documents ways to improve educational practice.	5.34	6.41	.76
The teacher evaluation system implemented by my school is utilized to inform professional development.	-1.50	6.41	.70
The teacher evaluation system implemented by my school is used to inform staffing decisions.	- .51	6.59	.50
I think principals should include a teacher's attendance as part of the evaluation.	.71	7.4.39	
I think the primary role of instructional supervision should be to improve instruction.	5.52	4.51	.35
I think the role of the principal in the teacher evaluation system should be more closely aligned with that of a coach.	5.40	4.89	.24
I think a teacher's special training beyond what is required by the district and continuing education should be considered in the evaluation process.	4.44	5.62	.25
My school's teacher evaluation system is utilized to impact school improvement.	.15	6.44	.69
I have a clear understanding of the definition of effective teaching.	3.68	6.10	.55
I have a clear understanding of the rubric language.	2.95	6.24	.56
I think relevant accomplishments and contributions to the school should be considered in teacher evaluation.	5.60	5.09	.27
I think information gained through regular contact with teachers should be part of the teacher evaluation system.	5.93	4.75	.28

Appendix B

Reliability Coefficients for Teacher Wellbeing Survey Items

Eleven Survey Items for Teacher Wellbeing (α = .92)	M	SD	r
The teacher evaluation tool implemented by my school is utilized to foster a positive environment for teachers.	-1.34	6.48	.86
The teacher evaluation tool implemented by my school is utilized to foster a productive environment for teachers.	- .61	6.46	.86
The teacher evaluation tool implemented by my school is utilized to foster a fair environment for teachers.	-.47	6.61	.85
The teacher evaluation tool implemented by my school plays an essential role in providing targeted professional development responsive to the needs of teachers.	-1.20	6.53	.76
The teacher evaluation tool implemented by my school improves retention rates for effective teachers.	-2.21	6.32	.74
The teacher evaluation system implemented by my school is implemented with fidelity.	.17	6.64	.67
The teacher evaluation tool implemented by my school is fair and accurate.	- .65	6.53	.83
The teacher evaluation tool implemented by my school is stressful for teachers*	-5.06	5.95	.38
The teacher evaluation tool implemented by my school creates a negative environment for teachers*	-1.02	6.72	.58
I have a clear picture of the criteria that are used to evaluate teachers.	2.84	6.40	.50

*Items recoded to accommodate counter balancing

Appendix C

Reliability Coefficients for Student Learning Survey Items

Four Survey Items for Student Learning (α = .82)	M	SD	r
My school's teacher evaluation system supports student learning.	.85	6.53	.80
I think teacher evaluations should consider individual student growth on standardized assessments.	.07	6.89	.27
My school's teacher evaluation system is utilized to impact student achievement.	.24	6.43	.81
My school's teacher evaluation system leads to students learning more.	-.82	6.47	.79

Non-parametric KWANOVA Comparisons for Groups of Survey Items for Teachers and Administrators

	Teachers (n = 1272) Mean Ranking	Administrators (n = 475) Mean Ranking	χ^2	p	η^2
Teacher Improvement	746.36	1215.81	299.52	< .01	.17
Teacher Well- Being	762.39	1172.88	229.01	< .01	.13
Student Learning	769.36	1156.27	203.32	< .01	.12
Total	750.95	1203.52	278.37	< .01	.16

NORTHERN MICHIGAN
UNIVERSITY