The Marzano Teacher Evaluation Model

Learning Sciences International
1400 Centrepark Boulevard, Suite 1000
West Palm Beach, FL 33401
717.845.6300

Table of Contents

1. The Research Base for the Marzano Teacher Evaluation Model

(For an in-depth examination of the research base, please see: Marzano, Toth, Schooling, "Examining the Role of Teacher Evaluation in Student Achievement: Contemporary Research Base for the Marzano Causal Teacher Evaluation Model," 2012. http://www.marzanocenter.com/teacher-evaluation/mc-whitepaper/ and Basilio and Toth, "The Research Base for the Marzano Teacher Evaluation Model and Correlations to State VAM," 2016. http://www.marzanocenter.com/MCTeacherEval_VAM%2020160328.pdf)

The Marzano Evaluation Model is based on a number of previous, related works that include: *What Works in Schools* (Marzano, 2003), *Classroom Instruction that Works* (Marzano, Pickering, & Pollock, 2001), Cla*ssroom Management that Works* (Marzano, Pickering, & Marzano, 2003), *Classroom Assessment and Grading that Work* (Marzano, 2006), *The Art and Science of Teaching* (Marzano, 2007), *Effective Supervision: Supporting the Art and Science of Teaching* (Marzano, Frontier, & Livingston, 2011). Each of these works was generated from a synthesis of the research and theory. Thus the model can be considered an aggregation of the research on those elements that have traditionally been shown to correlate with student academic achievement. The model includes four domains:

Domain 1: Classroom Strategies and Behaviors
Domain 2: Preparing and Planning
Domain 3: Reflecting on Teaching
Domain 4: Collegiality and Professionalism

The four domains include 60 elements: 41 in Domain 1, 8 elements in Domain 2, 5 elements in Domain 3 and 6 elements in Domain 4. For a detailed discussion of these elements see *Effective Supervision: Supporting the Art and Science of Teaching* (Marzano, Frontier, & Livingston, 2011).

Each of the works (cited above) from which the model was developed report substantial research on the elements they address. For example, *The Art and Science of Teaching* includes over 25 tables reporting the research on the various elements

of Domain 1. These tables report the findings from meta-analytic studies and the average effect sizes computed in these studies. In all, over 5,000 studies (i.e., effect sizes) are covered in the tables representing research over the last five decades. The same can be said for the other titles listed above**. Thus, one can say that the model was initially based on thousands of studies that span multiple decades and these studies were chronicled and catalogued in books that have been widely disseminated in the United States.** Specifically, over 2,000,000 copies of the books cited above have been purchased and disseminated to K-12 educators across the United States.

Experimental/Control Studies

Perhaps one of the more unique aspects of the research on this model is that it has a growing number of experimental/control studies that have been conducted by practicing teachers on the effectives of specific strategies in their classrooms. This is unusual in the sense that these studies are designed to establish a direct causal link between elements of the model and student achievement. Studies that use correlation analysis techniques (see next section) can establish a link between elements of a model and student achievement; however, causality cannot be easily inferred. Other evaluation models currently used throughout the country only have correlational data regarding the relationship between their elements and student achievement.

To date over 300 experimental/control studies have been conducted. Those studies involved over 14,000 students, 300 teachers, across 38 schools in 14 districts. The average effect size for strategies addressed in the studies was .42 with some studies reporting effect sizes of 2.00 and higher. An average effect size of .42 is associated with a 16 percentile point gain in student achievement. Stated differently: on the average, when teachers use the classroom strategies and behaviors in the Marzano Evaluation Model, their typical student achievement increased by 16 percentile points. However, great gains (i.e., those associated with an effect size of 2.00) can be realized if specific strategies are use in specific ways.

Correlational Studies

As mentioned above, correlational studies are the most common approach to examining the validity of an evaluation model. Such studies have been, and continue to be conducted, on various elements of the Marzano Evaluation Model. For example, one such study was recently conducted in the state of Oklahoma as a part of their examination of elements that are related to student achievement in K-12 schools (see What Works in Oklahoma Schools: Phase I Report and What Works in Oklahoma School: Phase II Report, by Marzano Research Laboratory, 2010 and 2011 respectively). Those studies involved 59 schools, 117 teachers and over 13,000 K-12 students. Collectively, those reports indicate positive relationships with various elements of the Marzano Evaluation Model across the domains. Specific emphasis was placed on Domain 1 particularly in the Phase II report. Using state mathematics and reading test data, 96% of the 82 correlations (i.e., 41 correlations for mathematics and 41 for reading) were found to be positive with some as high as .40 and greater. A .40 correlation translates to an effect size (i.e., standardized mean difference) of .87 which is associated with a 31 percentile point gain in student achievement. These studies also aggregated data across the nine design questions in Domain 1. All correlations were positive for this aggregated data. Seven of those correlations ranged from .33 to .40. These correlations translate into effect sizes of .70 and higher. High correlations such as these were also reported for the total number of Domain 1 strategies teachers used in a school. Specifically the number of Domain 1 strategies teachers used in school had a .35 correlation with reaching proficiency and a .26 correlation with mathematics proficiency.

Technology Studies

Another unique aspect of the research conducted on the model is that its effects have been examined in the context of technology. For example, a two year study was conducted to determine (in part) the relationship between selected elements from

Domain 1 and the effectiveness of interactive whiteboards in enhancing student achievement (see Final Report: A Second Year Evaluation Study of Promethean ActivClassroom by Haystead and Marzano, 2010). In all, 131 experimental/control studies were conducted across the spectrum of grade levels. Selected elements of Domain 1 were correlated with the effect sizes for use of the interactive white boards. All correlations for Domain 1 elements were positive with some as high as .70. This implies that the effectiveness of the interactive whiteboards as used in these 131 studies was greatly enhanced by the use of Domain 1 strategies.

Summary of Research Base

In summary, the Marzano Evaluation Model was designed using literally thousands of studies conducted over the past five or more decades and published in books that have been widely used by K-12 educators. In addition, experimental/control studies have been conducted that establish a more direct causal linkages with enhanced student achievement that can be made with other types of data analysis. Correlation studies (the more typical approach to examining the viability of a model) have also been conducted indicating positive correlations between the elements of the model and student mathematics and reading achievement. Finally, the model has been studied as to its effects on the use of technology (i.e., interactive whiteboards) and found it to be highly correlated with the effectiveness of that technology.

References

Haystead, M. W. & Marzano, R.J. (2010) Final Report: A Second Year Evaluation Study of Promethean ActivClassroom. Englewood, CO: Marzano Research Laboratory (marzanoresearch.com)

Haystead, M. W. & Marzano, R.J. (2010). Meta-Analytic Synthesis of Studies Conducted at Marzano Research Laboratory on instructional Strategies. Englewood, CO: Marzano Research Laboratory (marzanoresearch.com)

Marzano, R.J. (2003). What works in schools. Alexandria, VA: ASCD

Marzano, R. J. (2006).Classroom assessment and grading that work. Alexandria, VA: ASCD.

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

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

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). Classroom instruction that works. Alexandria, VA: ASCD.

Marzano, R.J., Marzano, J. S., & Pickering, D. J. (2003). Classroom management that works. Alexandria, VA: ASCD

Marzano Research Laboratory. (2010) What Works in Oklahoma Schools: Phase I Report. Englewood, CO: Marzano Research Laboratory (marzanoresearch.com)

Marzano Research Laboratory. (2011) What Works in Oklahoma Schools: Phase II Report. Englewood, CO: Marzano Research Laboratory (marzanoresearch.com)

2. About Robert Marzano and Learning Sciences International

Robert J. Marzano, PhD, is a nationally recognized researcher in education, speaker, trainer, and author of more than 30 books and 150 articles on topics such as instruction, assessment, writing and implementing standards, cognition, effective leadership, and school intervention. His books include *District Leadership That Works*, *School Leadership that Works*, *Making Standards Useful in the Classroom*, *The Art and Science of Teaching*, and *Effective Supervision*.

His practical translations of the most current research and theory into classroom strategies are internationally known and widely practiced by both teachers and administrators. He received a bachelor's degree from Iona College in New York, a

master's degree from Seattle University, and a doctorate from the University of Washington. He is also Executive Director of the Learning Sciences Marzano Center located in West Palm Beach, Florida, and of Marzano Research in Colorado.

Dr. Marzano believes that great teachers make great students: His Marzano Teacher Evaluation Model has been adopted by school districts in all 50 states because it doesn't just measure teacher ability, it helps teachers get better, improving their instruction over time. Dr. Marzano has partnered with Learning Sciences International to develop and implement the Marzano Teacher Evaluation Model, the School Leader and District Leader Evaluation Models, and the Non-Classroom Instructional Personnel Evaluation model, four complimentary evaluation systems that may be used with the iObservation technology platform.

Founded in 2002, **Learning Sciences International** partners with schools and districts to develop custom solutions for school improvement and professional development. With Robert Marzano, Learning Sciences co-developed the Marzano Evaluation Models and was selected as the statewide technical assistance provider for teacher evaluation implementation throughout the state of Florida. Learning Sciences was selected by the Michigan Department of Education's School Reform Office to provide monitoring and technical assistance to Priority Schools. Learning Sciences offers innovative technology, data analysis, research, consultation, and the tools and training to help schools meet their challenges and reach their greatest potential in today's high-stakes educational environment. For further information, visit www.LearningSciences.com.

3. Evidence of reliability, validity, and efficacy of the Marzano Teacher Evaluation Model

Recent Research Validating the Marzano Teacher Evaluation Model

Two recent studies address whether the Marzano Teacher Evaluation Model is a validated framework. The first, (Basileo and Toth, In Progress[1]), investigates whether the observation data from the Marzano Teacher Evaluation Model correlates with teacher value-added measures (VAMs) across the state of Florida. The second study, which was featured in a US Department of Education report in 2015, directly tested whether a professional development program based on the Marzano Teacher Evaluation Model increased student achievement in a pilot in Pinellas County Public Schools, Florida (see Basileo, Toth, & Kennedy, 2015). Both studies support and validate the Marzano Teacher Evaluation Model in Florida.

When evaluating the validity of observation protocols, studies typically assess the correlations between teacher observation scores and their value-added scores. Small to moderate correlations permit researchers to claim that the framework is validated (Kane, Taylor, Tyler, & Wooten, 2010). (See Endnote i for an overview of current research on the magnitude and range of correlation coe cients between observation data and VAM estimates).

A correlation between two variables does not necessarily mean that X causes Y; it merely provides evidence that there is a relationship between the two. Thus, validity studies that investigate whether the framework increases student achievement should include either experimental or quasi-experimental designs, to demonstrate that the framework increases student achievement.

Marzano Observation Correlations With Florida VAM

Basileo and Toth[1] investigated the magnitude of correlations using two years of data including all teachers in the state of Florida where districts were implementing the Marzano Teacher Evaluation Model and using the iObservation technology platform to collect observation data. Teachers' average observations scores were matched to state VAMs to assess validity coeffcients for the framework. The study included two years of data from the 2012-13 and 2013-14 school years. Additionally, each teacher's average score for each element was correlated to the state reading VAM, math VAM, and algebra VAM to investigate whether certain elements in the Marzano Evaluation Model had larger correlations to student achievement than others.

For the 2012-13 results, there were a total of 62,742 teachers who had an observation score. Researchers were able to match 13,236 (21%) of those teachers to a reading VAM and/or math VAM. The matching process was quite intensive because within state les, observation scores could be matched only by teacher name. Table 1 shows the correlations between the average teacher observation score and the reading VAM or math VAM. As noted below, both correlations were small and statistically significant ($p < .01$) with the coefficients ranging in size from .13 to .14.

Table 1. 2012-13 Marzano Observation Correlations and Florida VAM Scores

	Avg. Obs. Score	Read VAM	Math VAM
Avg. Obs. Score	1.00	**.132****	**.145****
N	62,742	8,511	6,001

[1] This study is in progress and will be published after the 2014-15 state VAM scores are released and analyzed. Check http://www.learningsciences.com/resources/ for more information.

Additionally, the average score for each element in the model was correlated to the reading and math state VAM. Thirty-eight, or 92%, of the elements were significantly correlated with the reading VAM (n = 5,021). Significant coefficients were small and ranged from .05 to .13. Thirty-six, or 87%, of the elements were significantly correlated with the math VAM (n = 3,515). Significant coefficients were small and ranged from .06 to .13.

For the 2013-14 results, there were a total of 58,527 teachers who had an observation score. Researchers were able to match 15,452 teachers (26%) to VAM data. In the 2013-14 school year, students were also tested in algebra. Table 2 shows the correlations between the average teacher observation score and the reading, math, or algebra VAM. Correlations were small and sta- tistically signi cant with the coefficients ranging from .14 to .21.

Additionally, the average score for each element in the model was correlated to the reading, math, and algebra VAM. **Forty, or 98%, of the elements in the model were significantly correlated with the reading VAM** (n= 6,720). Significant coefficients were small and ranged from .05 to .13. **Thirty-eight, or 93%, of the elements were significantly correlated with the math VAM** (n= 4,464). Significant coefficients were small and ranged from .06 to .17. Lastly, 29, or 71%, of the elements in the model were significantly correlated with the algebra VAM (n= 642). Significant coefficients were small and ranged from -.02 to .27.

This in-progress study is one of the largest validation studies on an observation framework. **The study has found that across two years of data, the Marzano Teacher Evaluation Model had significant and small correlations with teacher state VAMs.** Moreover, while there were small variations in the correlations coe cients by element, each element almost always had a significant correlation with teacher value-added scores. **Taken as a whole, these findings support the model as a valid, reliable, and accurate system to measure teacher proficiency.**

Educators can rely on the model to accurately determine teacher effectiveness.

2013-14 Pinellas Pilot Findings

In the spring 2012-2013 school year, Pinellas County Schools (PCS) received Florida Department of Education approval for a research project to develop a teacher e ectiveness system that would help teachers grow professionally. The new system would revitalize the evaluation system, diagnosing teacher pedagogical strengths and areas for growth, providing targeted support for individual professional skill development, and o ering a foundation in research-based classroom strategies to improve teacher practice. The projected outcome of the pilot was to increase student achievement as teachers improved their pedagogy through immersion in, and practice with, the Marzano Teacher Evaluation Model.

One innovation of the pilot was to employ short- duration student growth metrics for teacher evaluation. In contrast to evaluation measures that scored teacher practice long after students had left the classroom (in effect, generating scores when it was too late for teachers to make adjustments), the idea was to improve teacher practice within a single year while students were still in the classroom. The pilot included the use of multiple metrics: teacher self-assessment, principal observation scores, student perception surveys, and a short-duration value-added

Table 2. 2013-14 Marzano Observation Correlations and Florida VAM scores

	Avg. Obs. Score	Read VAM	Math VAM	Algebra VAM
Avg. Obs. Score	1.00	**.132****	**.145****	**.205****
N	62,742	8,511	6,001	1,217

measure (VAM) based at the unit level. The pilot had two additional, overarching aims: first, to create the diagnostic measures of teacher effectiveness, and second, to document and empirically test whether the professional development and coaching received by teachers and leaders throughout the year on the MTEM increased student achievement by the end of the year.

To assess program effects, a process and outcome evaluation was conducted to investigate whether the program had the intended effects of increasing student achievement. In total, five treatment schools and five statistically matched control schools were included in the study. Only the treatment schools received the training, coaching, and diagnostic measures of effectiveness.

Two sets of findings from this study are relevant to the validity of the Marzano Teacher Evaluation Model. The first finding pertains to the magnitudes of the correlation coefficients with VAMs. While the sample size is much smaller than the state level study, the magnitudes of the correlations are much higher when the model is implemented with fidelity. Table 3 shows correlation coefficients between observation scores and several di erent VAMs in Pinellas county. Significant coeffcients ranged from small to large (.14 to .53) with the largest correlation for the three-year aggregated math VAM at .53.

The outcome evaluation used several different methods to assess program effects, including independent sample t-tests, ordinary least squares regression, and hierarchical linear modeling. Out of the 26 assessments that had a control group match, 21 showed positive and significant growth for students at treatment schools ($p < .10$). Consequently, favorable and significant results were shown for treatment students in 81% of administered assessments. Moreover, fixed effects models showed similar results: Students who attended treatment schools had significantly increased growth scores (.37 to .39 standard deviations above prediction) compared to students at control schools, which accounted for both individual and

school characteristics (Basileo, Toth, & Kennedy, 2015).

"Students who attended treatment schools had significantly increased growth scores (.37 to .39 standard deviations above prediction) compared to students at control schools, which accounted for both individual and school characteristics."

The Pinellas pilot gained national attention from the Research Support Network and US Department of Education for these innovative efforts to reform teacher evaluation.

Overall, both studies outlined here provide ample support that the Marzano Teacher Evaluation Model has been validated in the state of Florida. Specifically, the first study, one of the largest validation studies conducted on an observation framework, found small to moderate correlations with teacher VAMs demonstrating that educators can rely on the model to accurately determine teacher effectiveness. The second study found evidence that student achievement significantly increased where the model was coupled with leadership coaching and implemented with fidelity.

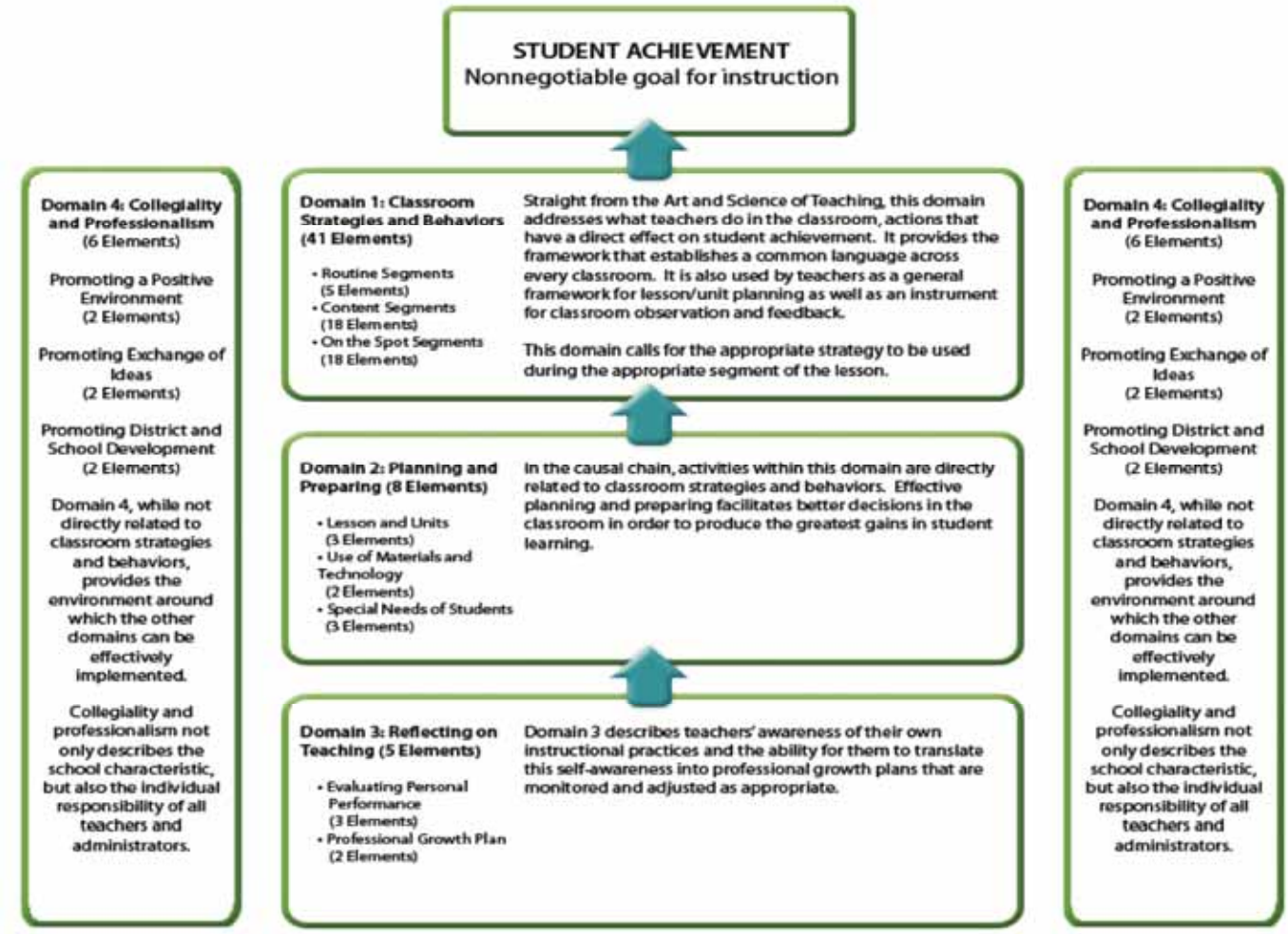
Table 3. 2013-14 Validity Coefficients in Pinellas County

	Unit VAM S1	Unit VAM S2	Year 1 State VAM Read	Year 1 State VAM Math	Year 1 State VAM Combined	Year 2 State VAM Read	Year 2 State VAM Math	Year State VAM Com d
Obs. Score S2	.104	**.135***	**.168**	**.444****	**.239***	**.221**	**.460****	**.287**

N	127	249	61	40	75	64	41	75

4. Overview of the Marzano Teacher Evaluation Model and Rubrics

The Research-Based Model: Four Domains Directly Tied to Student Achievement



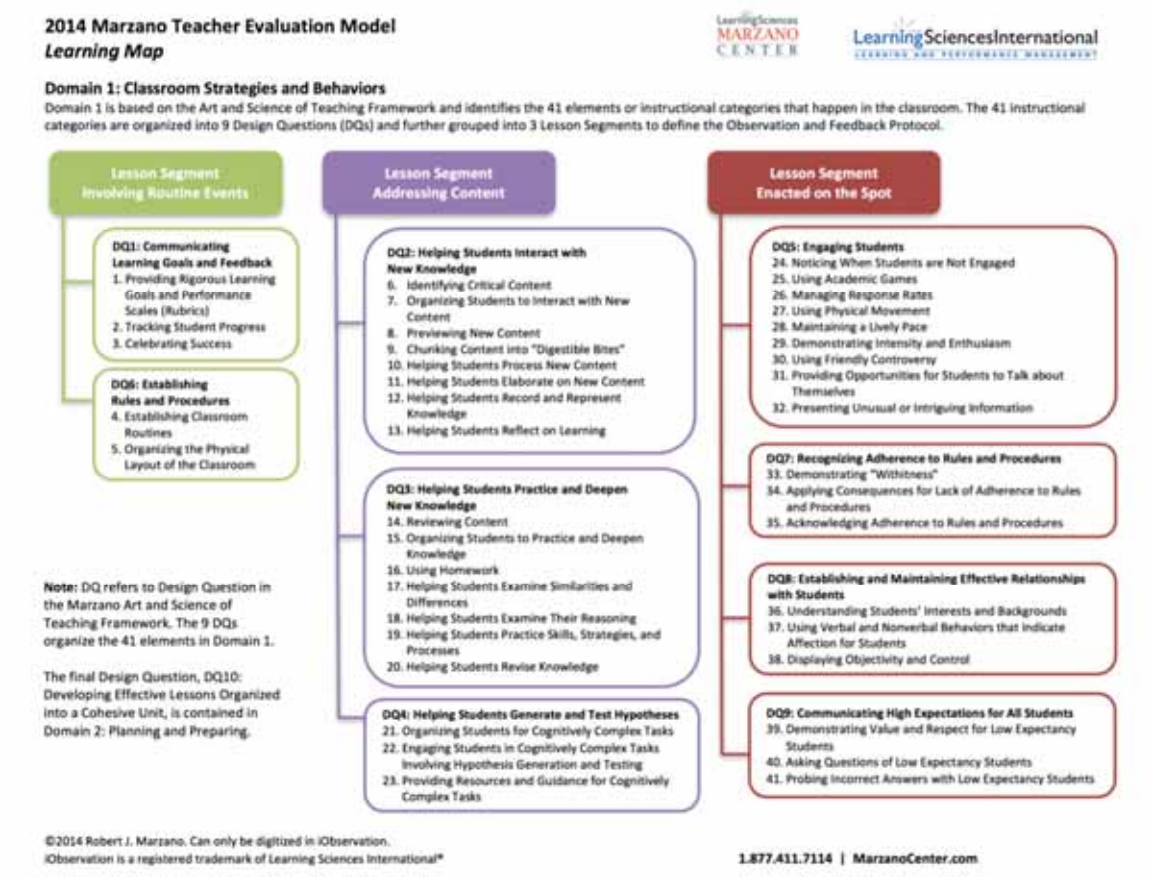
Domain 1, which contains 41 of the 60 elements, focuses on ***pedagogical strategies*** that have a direct link with student achievement. Domain 1 addresses what teachers do in the classroom: It reflects the intricacy of what happens during any given lesson and the natural flow of activities. The model is based on the premise that lessons are constructed with multiple parts and that each part of a lesson has distinct characteristics, routines, and processes. A model built to support teachers as they develop their skills must necessarily reflect the complexity of their work. But not all of the 41 elements need to be, or should be, observed in a single lesson. Domain 1 breaks down teaching into "thin slices" for richer diagnostic and feedback purposes.

Domain 2 focuses on ***planning and preparing*** for units of instruction and lessons within units. Because these elements are directly related to Domain 1, the better a teacher prepares, the more effective are his or her instructional choices.

Domain 3 addresses *deliberate practice*. It encourages teacher self-reflection in the areas of evaluating personal performance and developing and implementing a professional growth plan. When teachers receive specific and focused feedback using a common language of instruction, they increase their expertise and subsequently, student performance.

Domain 4 is the backdrop for the other domains and encourages a supportive culture. It addresses ***collegiality and professionalism,*** emphasizing opportunities to observe and discuss strategies. This domain supports teacher participation in lesson study, instructional rounds, teacher-led professional development, and professional learning communities in which teachers collaboratively examine evidence of student learning and the impact that specific instructional strategies have on learning.

The Marzano Teacher Evaluation Model Learning Map



5. Process for Classroom Observations

(Note: The charts below summarize LSI recommendations for implementation. Please see additional district attachments)

Protocols for collecting evidence, conducting evaluation conferences, developing learning, and developing performance improvement plans.

Types of Observations

	Announced	Unannounced
Formal	• Class Period • Pre-Conference • Post-Conference • Results used for annual evaluation • Written feedback provided to the teacher	
Informal	• At least 10 minutes long • Teacher is informed • Results used for the annual evaluation • May include written feedback	• At least 10 minutes long • Teacher is not informed • Results are used for the annual evaluation • May include written feedback
Targeted	• Usually 5-10 minutes • Planned so feedback for a single element can be given • Used for Deliberate Practice	
Walkthroughs		• Usually 3-10 minutes • Teacher is not informed • Results may be used for the annual evaluation

PLANNING CONFERENCE

Purpose: To discuss lesson that will be observed

Tips
• Planning conference should be scheduled • Set expectations including what forms and documents should be brought to conference *(Forms should be provided to the teacher before the conference)* • Both observer and teacher should have a clear understanding of the planned unit and lesson to be observed • Prepare responses and questions ahead of time *(Teacher can prepare responses ahead of time; Observer can prepare questions ahead of time)* Note: If possible, conduct the conference in the teacher's classroom.
Observer Role
• Clarify expectations with regard to the process • Promote dialogue about teaching and learning • Question, probe, and clarify • Gain as much information prior to the observation as possible • Identify elements both the observer and teacher have determined to be the focus of the observation • Discuss Domain 2 elements not readily observable
Teacher Role
• Participate in a dialogue about teaching and learning • Brief the administrator about the makeup of the classroom, (e.g. individual needs, levels, abilities, and special needs) • Identify the goals, instructional strategies, and assessment processes that will be used • Explain how their collegial relationships impact planning and teaching of lessons and units • Revise the upcoming lesson based on the conversation

REFLECTION CONFERENCE

Purpose: To discuss observed lesson, related documents, and student work/data from the lesson and to plan for future practice

Tips
• Before the conference, observer should provide and review documents that will be used and expectations for teachers • Observer should explain that the conference and the documents used will be a method to document elements in Domains 3 and 4 • Reflection conference should be completed soon after the observation • Observer should make sure to help the teacher see next steps as a result of the formal observation cycle • Prepare responses and questions ahead of time
Observer Role
• Clarify expectations regarding the process • Probe, clarify, question, affirm • Model a reflection process to include insights made by the observer • Help the teacher summarize their lesson • Help the teacher consider the impact of the lesson on student learning • Help the teacher consider future adjustments • Help identify supports (mentors) for areas of improvement as well as areas of expertise
Teacher Role
• Summarize the lesson (e.g. what worked well, what could be improved) • Identify the impact of the lesson on student learning • Share evidence of student learning • Identify new insights, potential adjustments

Scales and Evidences for the Marzano Teacher Evaluation Model.

The full scales and evidences for the model may be downloaded here:
http://www.marzanocenter.com/files/Marzano_AST_Domain1234_20130107.pdf

6. Training Plan for Evaluators and Observers

(Please see district attachment)

7. The Marzano Center Non Classroom Instructional Personnel Evaluation Model

A full report on the model may be downloaded here.
http://www.marzanocenter.com/non-classroom-instructional-support-personnel/

8. Appendix

Paper: *Marzano Teacher Evaluation Model: Prepared for the Michigan Department of Education.* *http://www.marzanocenter.com/files/MI_White_Papers_20131120.pdf*

For further information, contact us at LearningSciences.com.