

THE RESEARCH BASE FOR THE MARZANO TEACHER EVALUATION MODEL AND CORRELATIONS TO STATE VAM

Learning Sciences Marzano Center | March 2016

Background on the Research

The Marzano Teacher Evaluation Model was initially based on more than 5,000 studies that span five decades. These studies have been chronicled and cataloged in books widely disseminated to teachers and principals in the United States. More than two million copies have been purchased by K-12 educators. They include *What Works in Schools* (Marzano, 2003), *Classroom Instruction that Works* (Marzano, Pickering, & Pollock, 2001), *Classroom Management That Works* (Marzano, Pickering, & Marzano, 2003), *Classroom Assessment and Grading that Work* (Marzano, 2006), *The Art and Science of Teaching* (Marzano, 2007), and *Effective Supervision: Supporting the Art and Science of Teaching* (Marzano, Frontier, & Livingston, 2011). Each of these works was generated from a synthesis of research.

Thus, the Marzano Teacher Evaluation Model is an aggregation of the research on elements traditionally shown to correlate with student academic achievement.

To further test the effectiveness of the model, Dr. Marzano has partnered with state departments of education, districts, and schools across the nation to investigate the effectiveness of the Marzano Teacher Evaluation Model—specifically, to test the whether teachers’ use of the model can increase student achievement.

Over 500 teachers in 87 schools across the country participated in these experimental/control studies. The results showed a correlation between the model and student achievement. Furthermore, achievement was correlated not only for the model as a whole, but between the 41 specific strategies in Domain 1 and student test scores. Those correlations were positive in all cases.

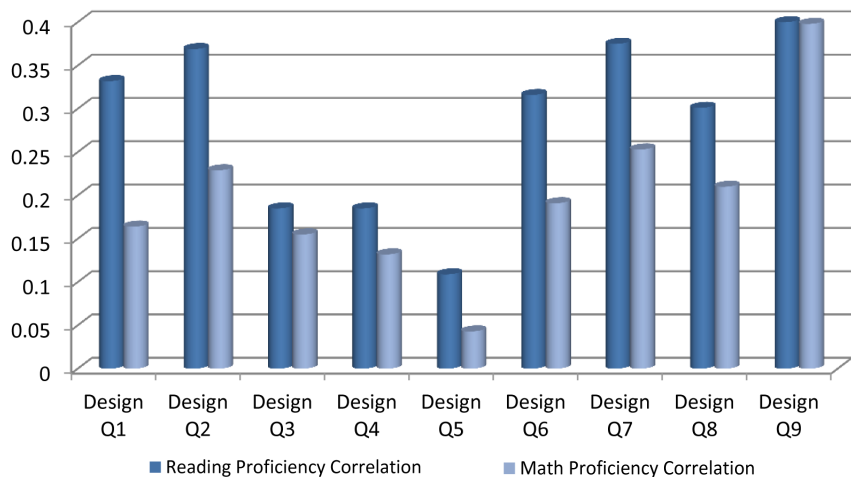
“The power of the Marzano Teacher Evaluation Model grows geometrically when applied regionally, as it provides for tightly coupled systems to maximize our instructional talent across all of our districts in the drive for increased student performance! The data that we receive from iObservation feeds our short cycle improvement process and supports our leadership and instructional decision making with enhanced accuracy and precision.”

Jason Jeffrey
Assistant Superintendent
Traverse Bay Area Intermediate School District

Math & Reading Proficiency Scores Correlated with 9 Design Questions

The Oklahoma State Department of Education commissioned Dr. Marzano in 2009-2010 to conduct a three-part study of Oklahoma schools. The study found a strong correlation between Dr. Marzano’s nine Design Questions and increased student achievement on state math and reading scores.

The studies aggregated student data from reading and math scores across the nine design questions in Domain 1. The highest correlations for D9, are associated with a 31 percentile point increase in student learning gains.



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, 2016¹), 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 the validation of 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 coefficients 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 a framework increases student achievement should also include either experimental or quasi-experimental designs, to demonstrate that the framework increases student achievement.

Marzano Observation Correlations With Florida VAM

Basileo and Toth (2016) investigated the magnitude of correlations using three 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 coefficients for the framework. The study included three years of data from the 2012-13, 2013-14 and 2014-15 school years. Additionally, each teacher's average score for each element within the model 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 extensive because within state files, observation scores could be matched only by teacher name, district and school. 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 .15.

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

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

¹This 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.

MARZANO TEACHER EVALUATION MODEL

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 58,520 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 statistically significant with the coefficients ranging from .14 to .21.

	Avg. Obs. Score	Read VAM	Math VAM	Algebra VAM
Avg. Obs. Score	1.00	.140**	.177**	.205**
N	58,520	12,099	8,262	1,217

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

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.

Lastly, for the 2014-15 results, the findings were similar if not stronger. During this year, the Florida Standards Assessment (FSA) included more rigorous items to assess state standards. There were a total of

59,412 teachers who had an observation score. Researchers were able to match 11,452 (20%) of those teachers to a reading VAM, math and/or algebra VAM. Table 3 shows the correlations between the average teacher observation score and the reading VAM, math or algebra VAM. **As noted below, correlations were small and statistically significant (p< .01) with the coefficients ranging in size from .21 to .26.**

	Avg. Obs. Score	Read VAM	Math VAM	Algebra VAM
Avg. Obs. Score	1.00	.210**	.263**	.209**
N	59,412	9,669	6,479	887

Table 3. 2014-15 Marzano Observation Correlations and Florida VAM scores

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= 4,930). Significant coefficients were small and ranged from .04 to .19. **Forty-one, or 100%, of the elements were significantly correlated with the math VAM** (n= 3,270). Significant coefficients were small and ranged from .10 to .26. Lastly, 29, or 71%, of the elements in the model were significantly correlated with the algebra VAM (n= 426). Significant coefficients were small and ranged from -.01 to .421.

This in-progress study is one of the largest validation studies on an observation framework for an entire state. **The study has found that across three 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 coefficients by element, each element almost always had a small and significant correlation with teacher value-added scores. **Taken as a whole, these findings support the model as a valid system to measure teacher proficiency.**

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 effectiveness 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 offering 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 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 different VAMs in Pinellas county. Significant coefficients 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

	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 2 State VAM Combined	Year 3 State VAM Read	Year 3 State VAM Math	Year 3 State VAM Combined
Obs. Score S2	.104	.135*	.168	.444**	.239*	.221	.460**	.287*	.251*	.532**	.347**
N	127	249	61	40	75	64	41	75	64	45	75

Table 3. 2013-14 Validity Coefficients in Pinellas County

MARZANO TEACHER EVALUATION MODEL

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 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 correlations with teacher VAMs demonstrating that educators can rely on the model. The second study found evidence that student achievement significantly increased where the model was coupled with leadership coaching and implemented with fidelity.

Endnotes

¹Overview of Current Studies

The following comprises a brief outline of current research on the magnitude and range of correlation coefficients between observation data and VAM estimates. Standards set by Cohen (1988) are as follows: .1 coefficients are classified as small correlations, .30 are medium, and .50 or above are large. Research has shown that correlations between observation data and VAM estimates are small to moderate. For example, Chaplin and colleagues (2014) found a small and significant correlation coefficient of .20 between the RISE observational instrument and VAM estimates using a sample of 358 teachers. Kane and Staiger (2012) had similar findings of their comparison between observation data and VAM scores across two years of

data. For math courses, they found small correlations between the two metrics that ranged from .09 to .18 on four observation instruments. For their implied measure of VAM (pp. 43–44), correlation coefficients were small to moderate, ranging from .12 to .34. Correlation coefficients were also small for ELA courses. They found that correlations between the two metrics ranged from .06 to .08 on three observation instruments. For their implied measure of VAM, correlation coefficients were also small and ranged from .09 to .12. Overall, the correlation coefficients found between observation data and VAM estimates are relatively small, with few ranging at the moderate level.

Research References

- Deliberate Practice for Deliberate Growth (*Marzano, Toth, 2013*) MarzanoDeliberatePractice.com
- Common Language, Common Goals, Hierarchical Evaluation & Growth System (*Carbaugh, Marzano, Toth, 2013*) MarzanoCenter.com/District-Leader-Evaluation/MC-Hierarchical-whitepaper/
- Examining the Role of Teacher Evaluation in Student Achievement (*Marzano, Schooling, Toth, 2012*), MarzanoCenter.com/Teacher-Evaluation/MC-whitepaper
- Dr. Marzano's Meta-Analytic Synthesis of Studies on Instructional Strategies (*Haystead & Marzano, 2009*), [MarzanoEvaluation.com/files/Instructional Strategies Report 9 2 09.pdf](http://MarzanoEvaluation.com/files/Instructional%20Strategies%20Report%2009.pdf)



Learning Sciences
MARZANO
C E N T E R

1.877.411.7114
MarzanoCenter.com
West Palm Beach, FL

© 2016 Learning Sciences International

