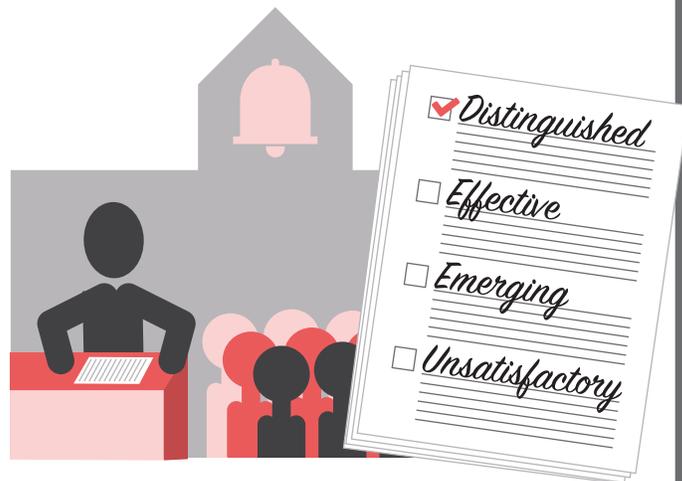# How are Teacher Evaluation Scores Related to School Achievement?

ROSLYN MILLER & ANNE HIERHOLZER

NOVEMBER 2015

What was the purpose of this study?

This study examined the results of the 2013-2014 Mississippi Statewide Teacher Appraisal Rubric (M-STAR) teacher-effectiveness ratings. We were interested in discovering:

- Are M-STAR scores related to school accountability grades?
- In which areas of practice do teachers need additional training?

How are teachers evaluated in Mississippi?

Since 2012, the Mississippi Department of Education (MDE) has worked to develop a new teacher-evaluation system that takes into account a variety of measures to determine teacher effectiveness. The 2013-2014 school year marked the first year of a gradual implementation of the Mississippi Teacher Evaluation System (MTES). Two of the four evaluation components—M-STAR and a measure of schoolwide student growth—were field-tested during the 2013-2014 school year. The study described here focused on M-STAR, not on the other components of the MTES.

During the M-STAR process, teachers are scored on 20 standards across five domains of teacher practice:

- Domain I: Planning
- Domain II: Assessment
- Domain III: Instruction
- Domain IV: Learning Environment
- Domain V: Professional Responsibilities

M-STAR PROCESS

- ✔ Formal classroom observations
- ✔ Preobservation and postobservaton conferences
- ✔ Informal walkthrough observations
- ✔ Review of artifacts
- ✔ Teacher self-assessments
- ✔ Optional student surveys
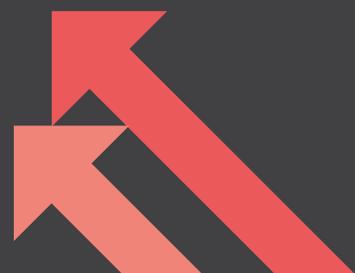
Scores for each standard are awarded by a school administrator who observes each teacher in action and collects evidence of teacher practice using multiple methods, including formal classroom observations, informal walkthrough observations, and artifact review. Administrators are trained on how to collect, observe, and score evidence, and they award scores ranging from 1 (Ineffective) to 4 (Distinguished) on each standard.

Teacher domain averages are calculated by averaging the scores on each standard within a domain, and each teacher's overall summative M-STAR score is equal to the mean of the five domain averages.

How was this study conducted?

This study analyzed M-STAR data from 18,429 teachers across Mississippi during the 2013-2014 school year. The following predictor variables were included in the analyses:

- School accountability grade
- Percentage of students receiving free- or reduced-price lunches (FRLs)
- School location (i.e., urban or rural)
- Teaching level (i.e., elementary school, middle school, high school)
- Years of teaching experience

Descriptive Statistics of M-STAR Domain Averages and Summative Score with Predictors

Variable	Mean	SD	N
School Grade	1.97	1.04	17,581
School Total Points	462.71	134.13	17,623
School FRL Percent	65.74	20.08	18,047
Rural Urban	.43	.49	18,047
Level Taught	.83	.88	18,047
Years Experience	7.57	9.17	18,047
Domain I Average	3.12	.56	18,047
Domain II Average	3.04	.63	18,047
Domain III Average	3.13	.54	18,047
Domain IV Average	3.20	.56	18,047
Domain V Average	3.20	.53	18,047
Summative Score	3.14	.49	18,047

A number of statistical analyses, including multiple linear regression and hierarchical linear modeling (HLM), were employed to determine links between teachers' M-STAR scores and the predictor variables. HLM was used to account for variation in M-STAR scores at both the teacher and school levels because teacher performance on M-STAR can be influenced by factors at a school.
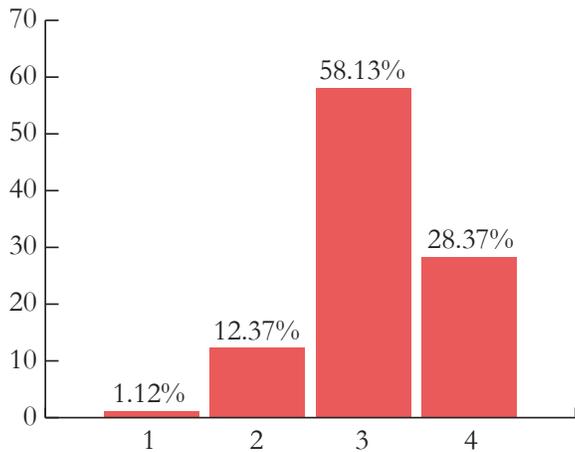
Across all domains, M-STAR scores were positively correlated with schoolwide measures of student achievement.
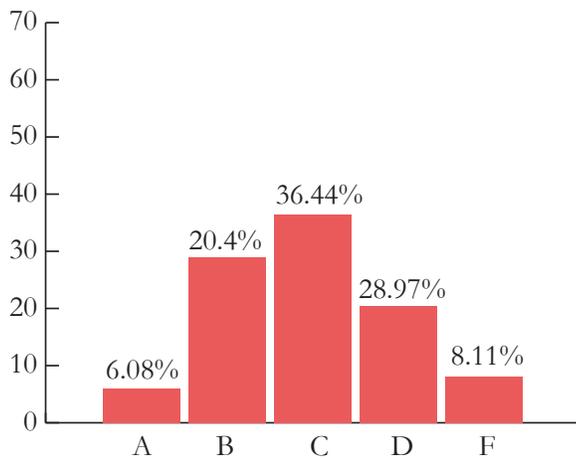
What were the results of the study?

Summative Score Results

On average across the state, teachers were rated Effective, with an average summative score of 3.13 on the four-point scale.

M-STAR Summative Ratings

School Accountability Grades of Teachers with M-STAR Ratings

Domain Results

Seventeen (17) of the 20 tested standards had mean scores in the Effective range. On each of the standards, at least 76% of teachers received either a 3 or 4.

Domain II (Assessment) had the lowest average scores, while Domains IV (Learning Environment) and V (Professional Responsibilities) had the highest mean scores. Domain III (Instruction) was

Domain I: Planning

1. Plans lessons that demonstrate knowledge of content and pedagogy
2. Plans lessons that meet the diversity of students' backgrounds, cultures, skills, learning levels, language proficiencies, interests, and special needs
3. Selects instructional goals that incorporate higher level learning for all students
4. Plans units of instruction that align with Mississippi Curriculum Frameworks or, when applicable, Mississippi College- and Career-Readiness Standards

Domain II: Assessment

5. Collects and organizes data from assessments to provide feedback to students and adjusts lessons and instruction as necessary
6. Incorporates assessments into instructional planning that demonstrates high expectations for all students

Domain III: Instruction

7. Demonstrates deep knowledge of content during instruction
8. Actively engages students in the learning process
9. Uses questioning and discussion techniques to promote higher order thinking skills
10. Brings multiple perspectives to the delivery of content
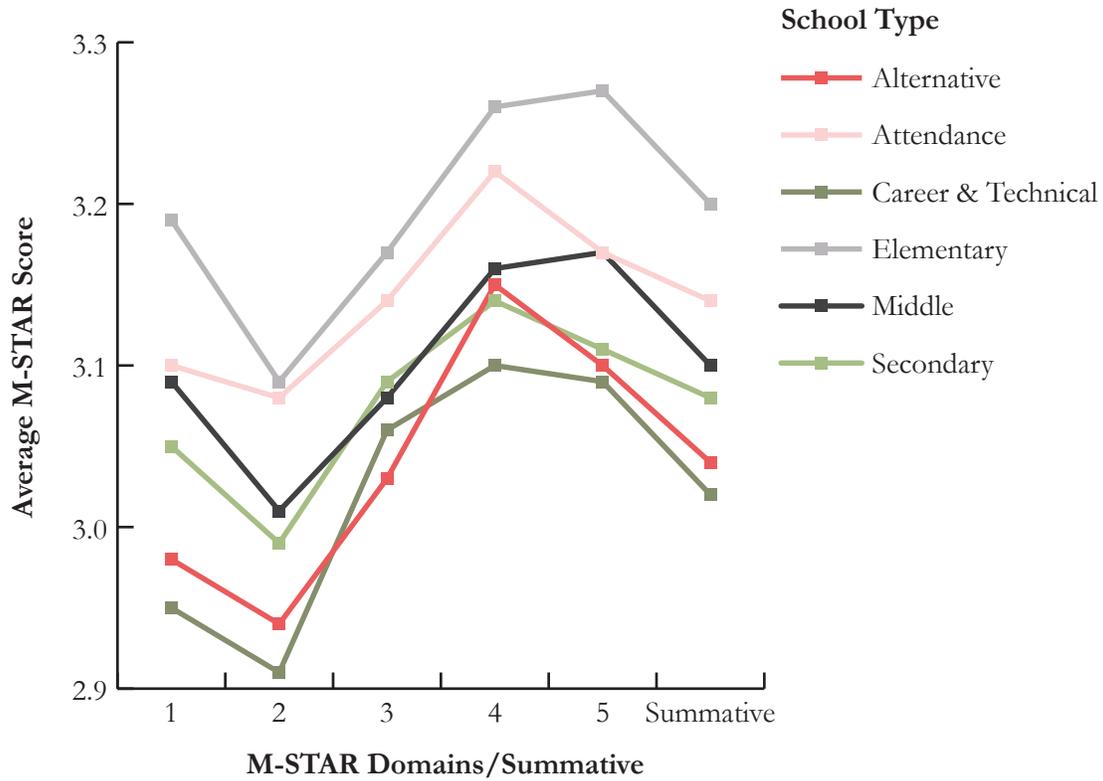11. Communicates clearly and effectively

Domain IV: Learning Environment

12. Manages classroom space and resources effectively for student learning
13. Creates and maintains a climate of safety, respect, and support for all students
14. Maximizes time available for instruction
15. Establishes and maintains a culture of learning to high expectations
16. Manages student behavior to provide productive learning opportunities for all students

Domain V: Professional Responsibilities

17. Engages in continuous professional development and applies new information learned in the classroom
18. Demonstrates professionalism and high ethical standards; acts in alignment with Mississippi Code of Ethics
19. Establishes and maintains effective communication with families
20. Collaborates with colleagues and is an active member of a professional learning community in the school

found to have the strongest correlation of all the domains with school accountability grades, while Domain V (Professional Responsibilities) had the weakest correlation.

Across all school levels, teachers scored highest on Standard 18–Demonstrates professionalism and high ethical standards. The three standards with the lowest scores were

• Standard 10–Brings multiple perspectives to delivery of content

• Standard 9–Uses question and discussion techniques to promote higher order thinking skills

• Standard 2–Plans lessons that meet the diversity of students' backgrounds, cultures, skills, learning levels, language proficiencies, interests, and special needs

Factors Linked to M-STAR Scores

The strongest predictor of both teacher and school performance was student socioeconomic status. In schools where many students receive FRL, teachers tend to receive lower M-STAR scores. Similarly, schools where lots of students receive FRL tend to receive lower school accountability grades.

A number of other factors were also found to have an influence on M-STAR scores. The higher a school's accountability grade, the higher the M-STAR scores of teachers at that school were likely to be. In addition, teachers' years of experience (the more the better) and school location in an urban setting were also found to have a positive effect on both M-STAR scores and school accountability grades. Grade level was found

FACTORS LINKED TO RISING M-STAR SCORES

🔍 Higher school accountability ratings

💵 Higher student socioeconomic status

📅 More years of teacher experience

🏙 Urban schools

⬇ Lower grade levels

to have a negative relationship with M-STAR scores, with teachers in elementary grades generally receiving higher M-STAR scores than teachers in higher grades.

Variation in M-STAR Scores

Although the HLM indicated some variation in M-STAR scores, it was minimal. M-STAR scores did not vary widely, which indicates problems with principals differentiating among teachers. Principals tended to differentiate very little between teachers, rating them primarily with 3s. However, they tended to keep teachers ranked in the same order, so even the slight differentiation was reliable and systematic, not random. This supports a conclusion that higher-rated teachers can be judged as more effective than lower-rated teachers, even if the differences in their scores are small.

The HLM found more variation in teacher effectiveness within schools than across schools, which indicates a range of teacher performance levels at a given school regardless of its accountability grade.

Because teacher ratings were positively linked to schoolwide accountability grades, M-STAR appears to be reliably measuring aspects of teacher performance related to student growth on the standardized assessments used to calculate school accountability grades. Additionally, M-STAR's internal consistency and reliability were found to be high.

Limitations

This study sought to determine correlations between a number of variables. While many interesting correlations were discovered, it is important to keep in mind that correlation does not imply causation. The relationships that were uncovered by this study may be caused by other factors not included in our model.
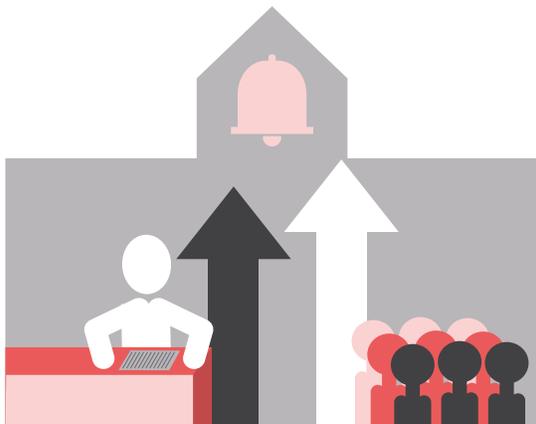
How can the study results be translated into action?

- The teacher-evaluation process is meant to meaningfully differentiate performance among teachers, yet most teachers were rated at the same performance level. The low variation in M-STAR scores may indicate that administrators need additional training to accurately score teachers so that scores better reflect actual differences in teacher effectiveness.

- The negative correlation between FRL percentage and M-STAR scores may indicate that effective teachers are not equitably distributed across schools and/or that student characteristics directly affect their teachers' ratings. Further investigation into this relationship is needed.

> Teachers may benefit from more focused training in areas where they received lower M-STAR scores.

- Elementary school teachers were more likely than teachers at other levels to receive high M-STAR scores, although elementary schools in general had lower school accountability grades than high schools or middle schools. Additional research could help shed light on this paradox.

- The three M-STAR standards for which teachers received the lowest average scores (Standard 10–Brings multiple perspectives to delivery of content; Standard 9–Uses question and discussion techniques to promote higher order thinking skills; Standard 2–Plans lessons that meet the diversity of students' backgrounds, cultures, skills, learning levels, language proficiencies, interests, and special needs) are areas where teachers could benefit from more focused training, professional development, and curricular resources to promote their growth and effectiveness.

- School accountability grades were found to predict M-STAR scores, indicating that the M-STAR reliably measures aspects of teachers' practices that are related to growth in student achievement on standardized assessment. However, schoolwide accountability grades cannot explain differences in student achievement

on standardized tests between teachers within schools and do not account for other ways in which teachers foster student achievement, such as boosting students' aspirations, collaboration skills, or creativity. Linking teachers with the achievement of the specific students in their classroom is necessary to explain variability in teachers' contributions to student growth within schools.

- Continuing analysis of M-STAR results is needed. Ideally, individual teacher effectiveness should be measured over at least three years to allow for more accurate judgment of the long-term relationships between teacher effectiveness and student achievement.

- Additional research that includes other factors that have been reported to explain variation in teacher effectiveness and student achievement may be beneficial. These variables of interest include:
  o Educational attainment of students' parents
  o Parental involvement and level of expectations
  o School-climate indicators (e.g., safety, high expectations)
  o School resources (e.g., facilities, instructional materials)
  o Teacher knowledge, education, and professional development
  o Administrator leadership
  o Class curriculum

SUGGESTED ACTIONS

- ✔ Provide additional scoring training to administrators

- ✔ Investigate possible correlations between teacher distribution and student characteristics

- ✔ Research inconsistencies with elementary teacher scores

- ✔ Provide teachers more focused training, professional development, and curricular resources

- ✔ Link teachers with the achievement of specific students

- ✔ Continue to analyze M-STAR results

- ✔ Research additional factors that may explain variation in scores

M-STAR is considered to be reliably measuring aspects of teacher performance related to student growth on standardized assessments

MISSISSIPPI STATE UNIVERSITY™
RESEARCH & CURRICULUM UNIT