Is There Always a Direct Link between Classroom Observations and Student Achievement?
An Examination of Non-linear Relationships

Valeriy Lazarev & Denis Newman (Empirical Education Inc.)

Objective.

Objective
Explore the patterns of relationship between observational scores and value-added measures of teacher performance.

Hypothesis
Pairwise correlations between scores (ratings) given to teachers by observers using a particular protocol and value-added scores calculated from student achievement gains do not convey complete information about the relationship between the two types of indicators of teaching quality.

Goal
Help develop a more rigorous approach to comparing different metrics of teacher performance and their correct interpretation and use in a policy context.

Perspective.

Growing interest in using VAM to evaluate school/teacher performance
- Renewed interest in traditional measures of teacher effectiveness
- Research into relationship between different metrics of teacher performance varies
- Recent studies focus on accuracy of VAM estimates, little attention to the methodology of comparisons between VAM and observation scores

Relationship between value-added and observation scores
- Association reported in terms of the linear correlation coefficient
- Observation scores are subjective estimates that use ordinal scale
- Observation scores are therefore not necessarily linear functions of underlying value-added

Methods.

Initial analysis
- Two protocols used in two large urban school systems
- Each system assigned raters to score videos of lessons taught in other school districts
- Raters were trained on the observation protocol and scored the lessons on a number of indicators
- Value-added scores for teachers were known

Subsequent analysis
- Relationships between observation scores and value-added scores performed separately for each domain (38 total)
- Nonparametric regression approach
- Optimal degree of smoothing is set based on the best linear unbiased predictor criterion

$$y_i = f(x_i) + \varepsilon_i$$

y = observation score

x = student achievement gains

$f(x) = E(y|x)$ is a smooth function and ε_i is the zero-mean random error.

Data.

The analyses reported here were conducted as part of the development of a web-based tool for use by school systems to validate their observational protocols through correlations with VAM scores. The tool was field tested in a number of school systems using the district's local protocol and district-trained observers.

The tool used an extensive library of high-quality video recordings of upper-elementary and middle-school lessons in ELA and math made in several urban school districts. Videos were assigned randomly to raters who trained to perform classroom observation across subjects and grade levels.Both protocols were local adaptations of widely-used general (not subject-specific) protocols and were. Lessons represented a wide range of value-added scores, which were calculated on the basis of student achievement on two assessments, one year apart.

The total number of observations (scores) per domain was between 201 and 228. The response rate was therefore very similar across domains.

Results.

Figure 1. Relationship between observation scores and value-added scores: An example of "Π-relationship"
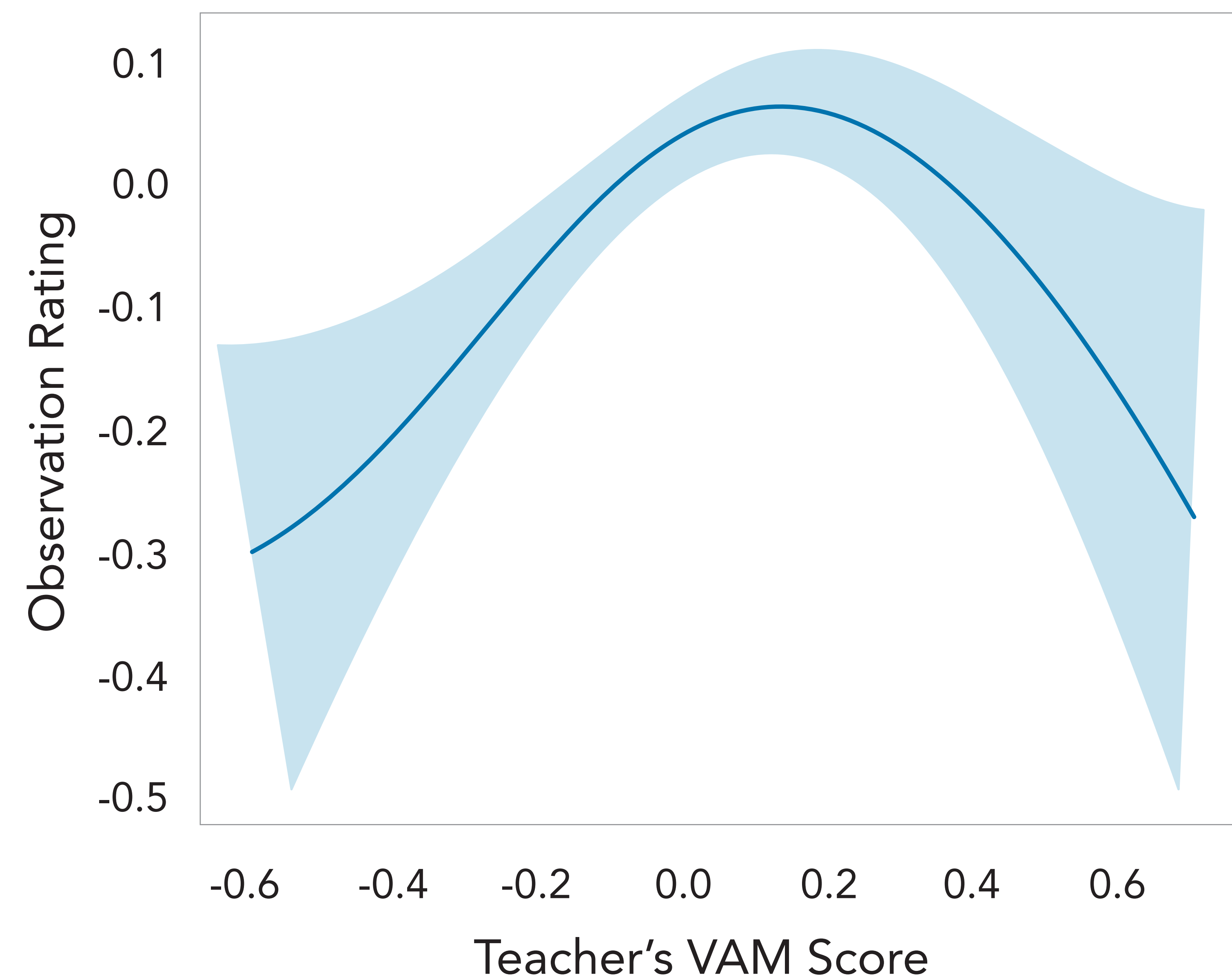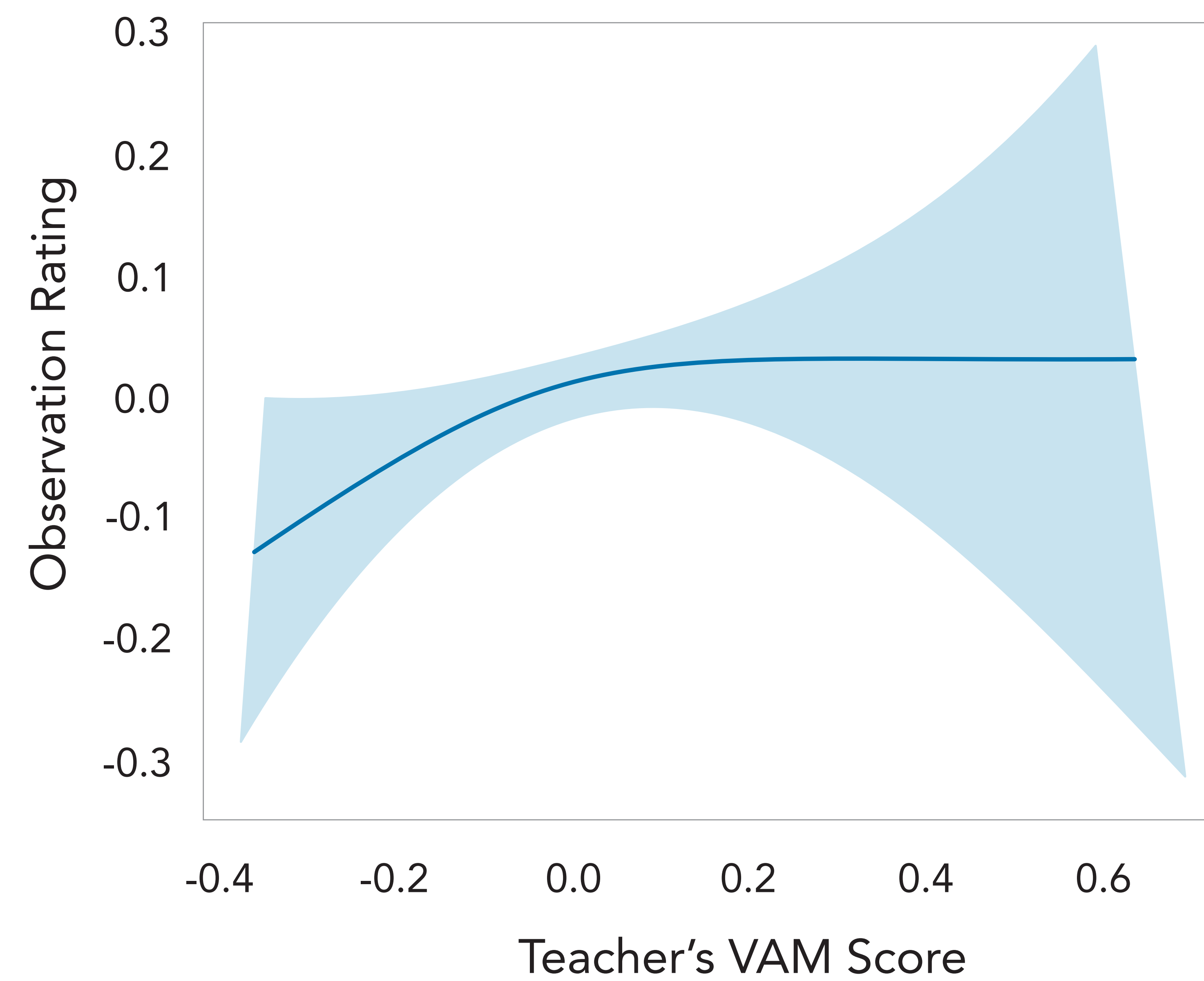

Figure 2. Relationship between observation scores and value-added scores: An example of "Γ-relationship"

Distribution of relationships between observation scores and value-added scores by type

| | Elementary | | | | Middle | | | |
| | State test | | Common core-like test | | State test | | Common core-like test | |
Statistical relationship	Math	ELA	Math	ELA	Math	ELA	Math	ELA
Linear	5	1	0	0	2	3	0	4
Non-Linear, Γ	0	0	0	0	4	0	4	4
Non-Linear, Π	1	2	1	2	1	0	0	0
No significant relationship	2	5	7	6	1	5	4	0

Significance.

We take the analysis of relationships between different metrics of teacher performance one step further focusing on the shape of such relationships and find that the latter are substantially non-linear. This suggests that future analysis of this kind should not limit itself to reporting pairwise correlations between value-added and observation scores but have to present a broader picture and/or apply an appropriate non-linear transformation to an observation scale prior to generating summary statistics. It also suggest that creating a composite observation score or a comprehensive measure of teacher effectiveness as a linear combination of domain scores may not be a valid procedure because it implies a mechanical summation of parts that may or may not be related to each other in an unambiguous manner. Our preliminary results suggest new avenues for quantitative analysis of measures of teacher effectiveness.