This study examined the impact of modeling decisions made in implementing value-added teacher evaluation; such choices include the growth model itself, whether to control for student characteristics, how many years of scores to use, and the subject tested.

Estimates of teacher effectiveness were derived from five models, which were a covariate adjustment model that conditioned on prior achievement only, a covariate adjustment model that conditioned on certain student characteristics as well as prior achievement, a gain score model, the growth model underlying the assessment, and student growth percentiles.

The resulting rank orderings of the teachers were examined and found to be highly consistent with one another using scores for either one or three classroom years. When the movement of teachers between quartiles of the rank orderings was examined, a difference in performance between some models did emerge. The covariate adjustment models were highly consistent, suggesting that control for student-level characteristics was unnecessary. Using three years of data rather than one did not significantly change model performance, and comparison of rank orderings based on reading scores versus mathematics scores gave mixed results.

The year-to-year inconsistency in rank orderings was striking. Movement of teachers between quartiles from one year to the next was far greater than that observed when comparing modeling conditions. Under a rating scheme in which teachers were distinguished from average effectiveness if they appeared in the extremes of the rankings, nearly half of teachers changed ratings from one year to the next.