RAND EDUCATION

Exploring Robustness of Estimated Teacher Effects to Teacher-by-Student Interactions and Missing Data

J.R. Lockwood & Daniel F. McCaffrey

June 15, 2006

This material is based on work supported by the Department of Education Institute of Education Sciences under Grant No. R305U040005 . Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of these organizations.

Value-Added Models for Teacher Effects Aim to Separate Teacher Inputs from Student Background

- Goal of value-added models (VAM) for teacher effects:
  - Use longitudinal student achievement measures to distinguish teacher effects from the effects of student background variables
- A problem of causal inference with observational data
  - Teachers teach nonequivalent student populations
- Inferences necessarily depend on statistical model assumptions
  - Different modeling approaches have developed rapidly
  - Literature has identified sources of potential bias in estimated effects and has begun exploring robustness to modeling assumptions

This Work Focuses on Testing Robustness to Two Assumptions

1. A single, scalar teacher effect rather than one that varies across students is sufficient for inference

2. Missing test score data are “missing at random”

- Issues are seemingly unrelated

- But a single modeling strategy is capable of addressing both of them

History of Achievement Outcomes Contains Substantial Information About Individual Students

- **The goal of VAM is to separate this achievement profile into what is due to the student and what is due to schooling**

- **The part due to the student is the best indicator the data can provide of how the student would perform regardless of context**

- **Our models try to capture this "general achievement" information about each student and use it elsewhere in the model:**

  1. **Letting the effect of each teacher on each student depend on that student's general level of achievement**
  2. **Letting the probability of a student missing test administrations depend on that student's general level of achievement**

A Value-Added Model that Parameterizes the General Level of Achievement

- **Presented here for simple case of a single cohort of students tested in consecutive grades (years) in one subject**
- Y_{it} **is scaled score for student** i **in year** t
- δ_i **is the latent measure of student** i**'s general level of achievement**

$$
\begin{aligned}
Y_{i1} &= \mu_1 + \theta_{1(i)} + \delta_i + \epsilon_{i1} \\
Y_{i2} &= \mu_2 + \alpha_{21}\theta_{1(i)} + \theta_{2(i)} + \delta_i + \epsilon_{i2} \\
Y_{i3} &= \mu_3 + \alpha_{31}\theta_{1(i)} + \alpha_{32}\theta_{2(i)} + \theta_{3(i)} + \delta_i + \epsilon_{i3} \\
Y_{i4} &= \mu_4 + \alpha_{41}\theta_{1(i)} + \alpha_{42}\theta_{2(i)} + \alpha_{43}\theta_{3(i)} + \theta_{4(i)} + \delta_i + \epsilon_{i4} \\
&\ldots
\end{aligned}
$$

$$
\theta_{tj} \sim N(0, \tau_t^2) \qquad \delta_i \sim N(0, \gamma^2) \qquad \epsilon_{it} \sim N(0, \sigma_t^2)
$$

Analyses Based on Data from a Large Urban School District

- One of the nation's largest urban school districts (about 75,000 students)
- We analyze one cohort of 9,295 students from 1998 to 2002
- Students in grades 1 to 5 for these years
- Math and reading scaled scores from annual spring testing
- Links students across years, to teachers and to schools
- Focus on grade 1 teachers in 1998 to grade 5 teachers in 2002
  - About 1,500 teachers in total

Investigation 1: Estimating Student-by-Teacher Interactions

- To date, the array of different VAMs specify a single, scalar estimate (per subject) of the effect of each teacher
- Experience suggests that teachers might be *differentially* effective with different types of students
- General level of achievement (δ_i) is intuitive as a characteristic on which interactions might manifest
- Primary motivations for investigating interactions:
  - Quantifying interactions might inform targeted interventions or even class assignments ("increasing everyone's value-added")
  - Strong interactions would have critical consequences for the types of inferences supportable by VAM estimates

We Extended the Basic Model to Allow Teacher Effects to Vary Across Students

- **Teacher effect on student i in year t assuming no interaction:**

$$\theta_{t(i)}$$

- **Teacher effect on student i in year t allowing interaction:**

$$\theta_{0,t(i)} + \theta_{1,t(i)}\delta_i$$

(average effect $+$ adjustment for each student)

- **Effects modeled as bivariate normal, with variances and correlation estimated from data**
- **Embedding the interactions into the model was challenging**
  - **Needed to assume $\alpha_{tt'} \equiv 0$: no persistence of teacher effects**
  - **Not a costly assumption because persistence appears to be weak with these data ($\alpha_{tt'} \approx 0.1$ to 0.2)**

Results Provide Evidence of Interactions

- Model successfully implemented in Bayesian framework using WinBUGS software
- Two preliminary findings:
  1. Model provides evidence of teacher-by-student interactions, with magnitudes corroborated by exploratory analyses of the data
  2. The two effects for each teacher are positively correlated (around 0.5 for math and 0.3 for reading)
     - Literal interpretation: teachers who are effective on average are particularly effective with above average students
     - NOTE: This is preliminary and requires substantial further exploration

The Magnitudes of the Estimated Interactions Are Relatively Small

- Potential variation of teacher effects across students with different δ's accounts for about 9% of the total variance in teacher effects
  - Can lead to correlations as low as about 0.7 between teacher effects on generally very low achieving students (-2 SD(δ)) to teacher effects on very high achieving students (+2 SD(δ))
- However, the impact of interactions on teacher effects depends on the variation in classroom averages of δ
  - In the data, variation between classrooms in δ is small relative to overall variation in δ
- Estimated teacher effects across classrooms with very low and very high average levels of achievement are correlated at least 0.97
- Conclusion: Interactions appear to be present, but bias introduced by failing to account for them is likely to be negligible

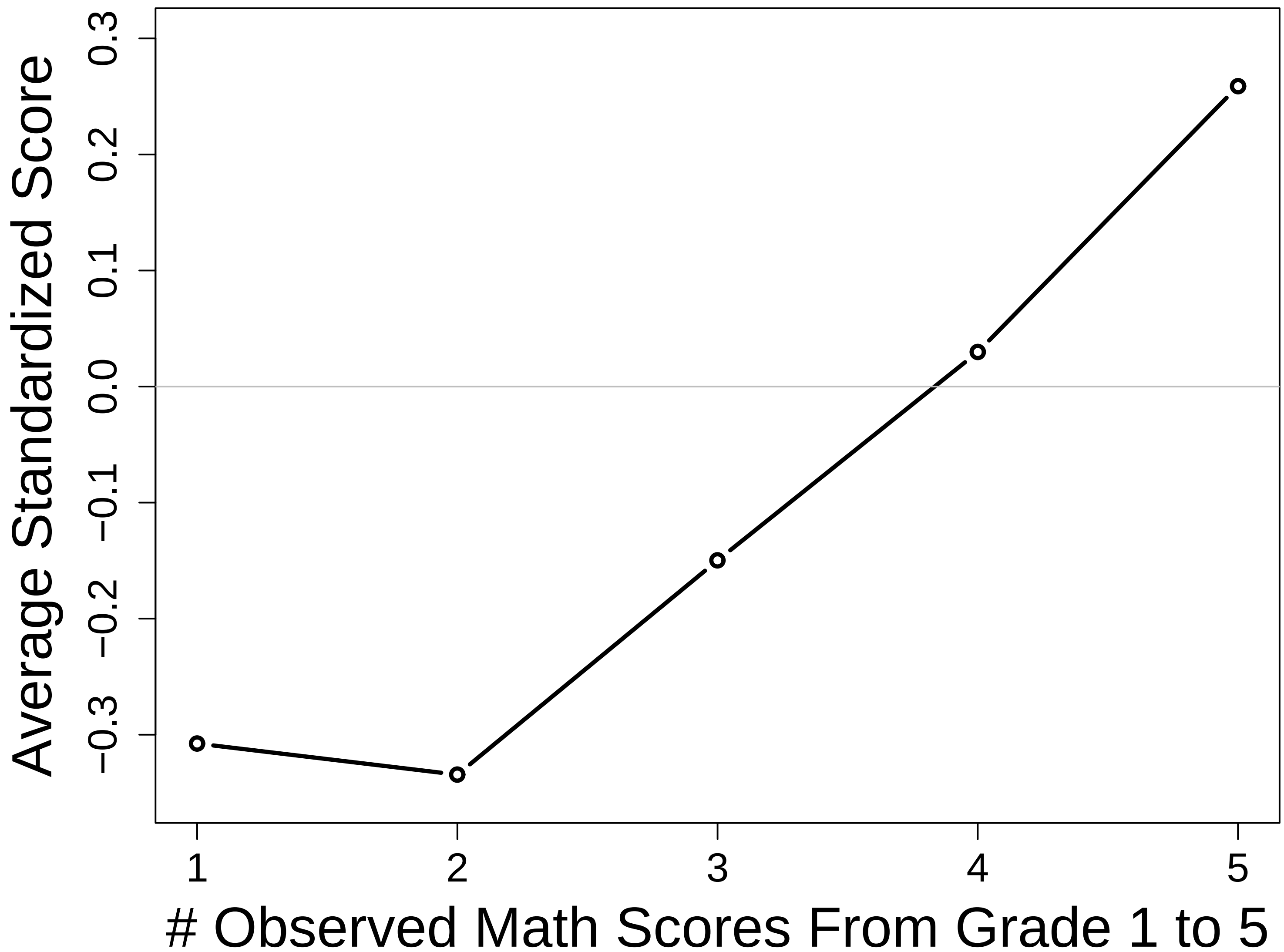
Investigation 2: Relaxing Assumptions About Missing Data

- NCLB testing participation requirements result in nearly all students being tested in many classrooms
- However, all longitudinal student achievement data includes some incomplete records
  - Often large proportion of records are incomplete
  - Only about 20% of students in our data have complete testing histories
- Mobility results in students with incomplete records when they are outside of the data collection unit
  - Missing test scores
  - Missing teacher links

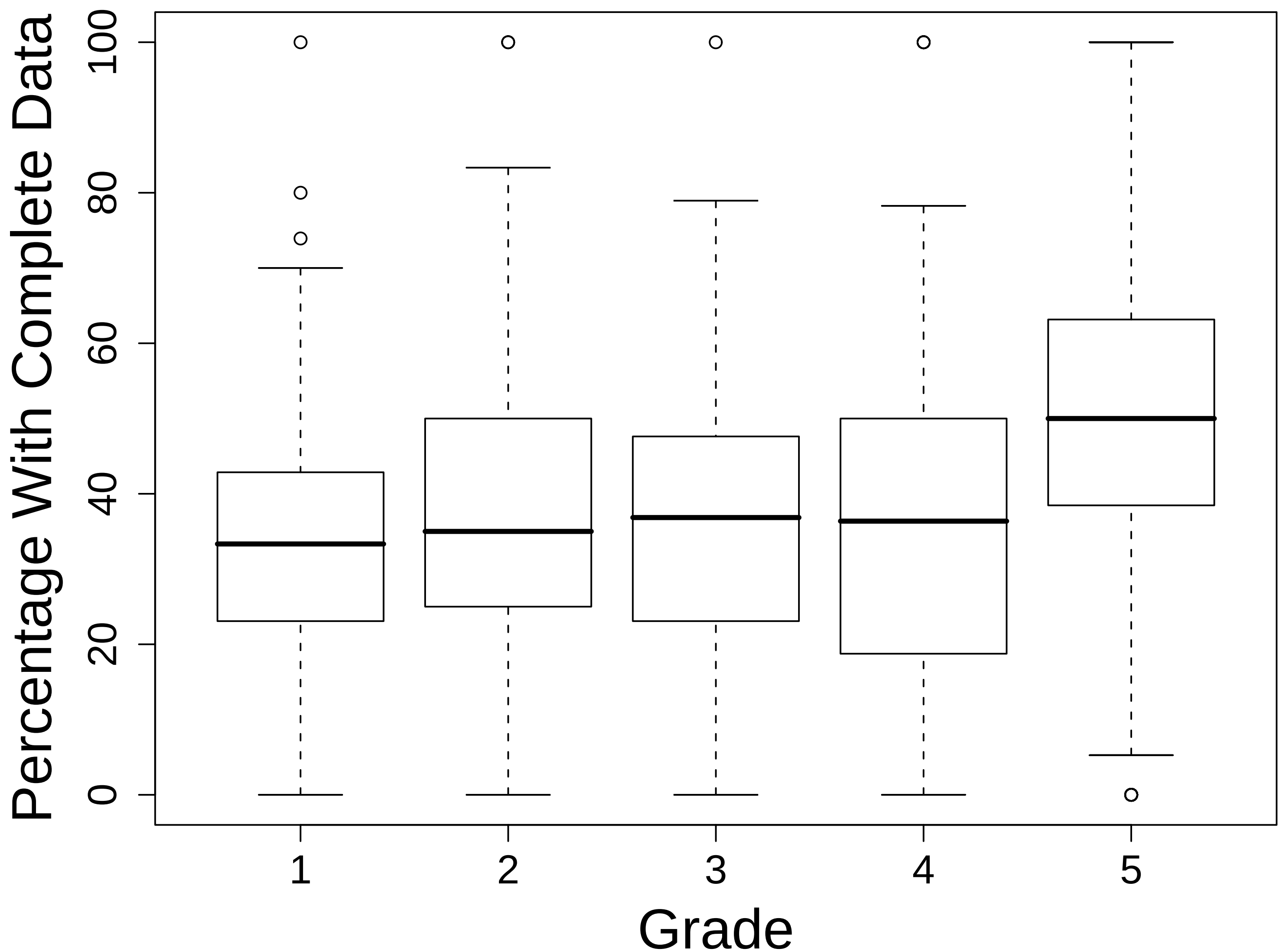
VAM Literature Cites Missing Data as a Potential Source of Bias in Estimated Teacher Effects

- "When a substantial number of records are incomplete. . . there is a concern that not only will the variance of the estimates increase but also that the estimates may be biased." (Braun, 2004)
- Intuitively, missing scores present the potential for bias when:
  - Students with missing scores differ from other students
  - Classrooms differ in the percentage of students with incomplete data

Students with Missing Scores Tend to be Lower Scoring When They Are Observed



Classrooms Differ Greatly in the Percentage of Students with Complete Data



Three Different Classifications of Missing Data

- Missing completely at random (MCAR)
  - Missing scores are a simple random sample of all scores
- Missing at random (MAR)
  - Students with missing data can differ from other students but only in observable ways
  - Conditional on the observed data, the distribution of missing data is same for students with complete and incomplete data
- Missing not at random (MNAR)
  - Missing data differ systematically from observed data even conditional on observed scores
  - MNAR models have not been used with VAM

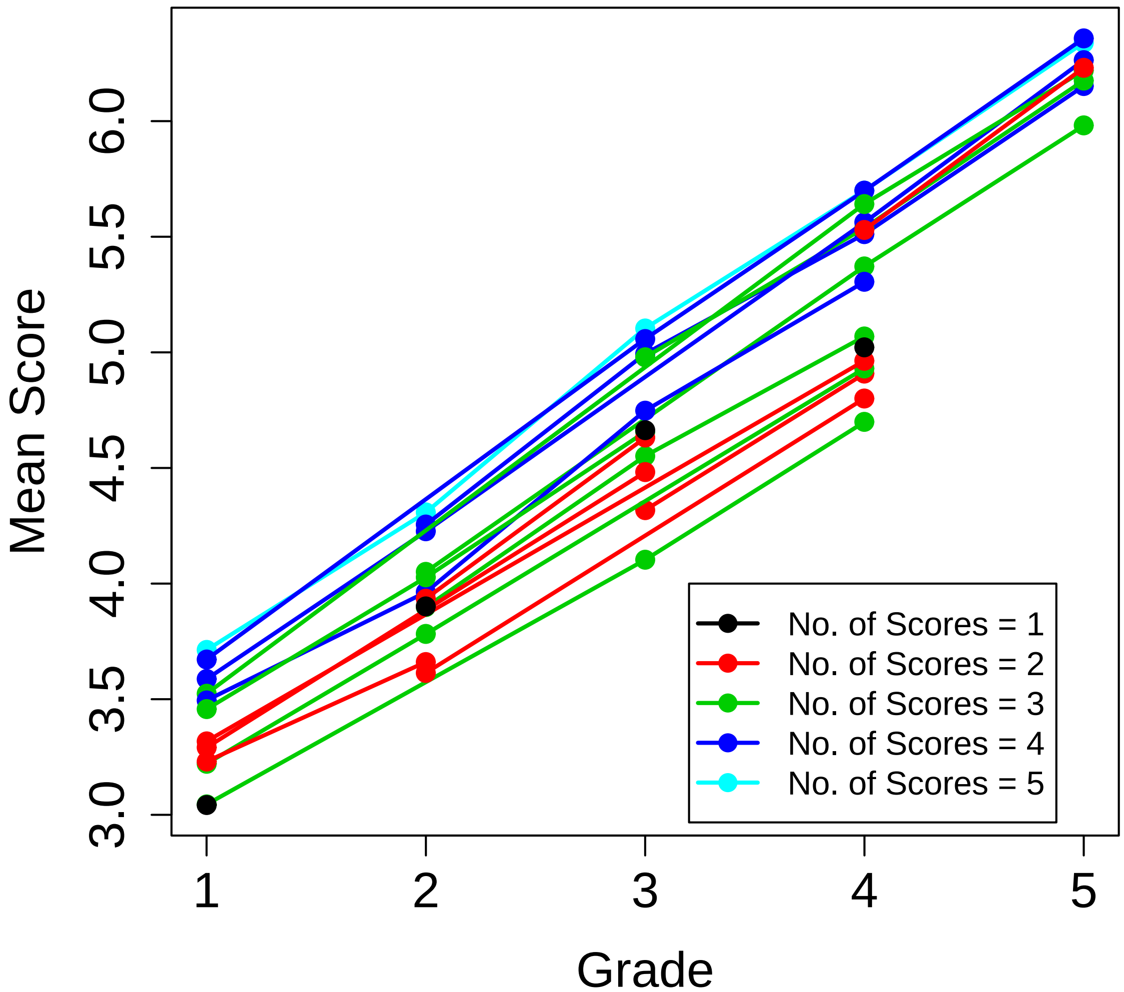
We Embedded Two Different Types of MNAR Models into Our General Model

- **MNAR Selection Model**
  - **Builds on the fact that lower scoring students are more likely to have unobserved scores**
  - **Number of observed scores depends on δ**
  - **Estimate parameters of the selection model from the data**
- **MNAR Pattern Mixture Model**
  - **Allows test scores for students with different missing data patterns to have different means and covariance matrices**
  - **Teacher effects assumed constant across all patterns**

Results from MNAR Models Almost Identical to Standard Results Which Assume MAR

- For now have looked only at mathematics data
- We estimated teacher effects using:
  - The basic model, which assumes MAR
  - The two MNAR models
- The correlations between teacher effects from the two MNAR models and those from the basic model were at least 0.98 for all years
- Behavior of other model parameters suggests that models are functioning reasonably

Estimated Means from Pattern Mixture Model



Robustness of Estimates Results from Two Factors

- Provided our chosen MNAR models are reasonable, results add to growing evidence that missing data are not a large source of bias in estimated effects
- Why might this happen?
  - Analytical work indicates that multivariate random effects models downweight scores from students with incomplete data relative to scores from students with complete data
  - Data suggest that missing scores are not sufficiently extreme to overcome downweighting in the context of specific MNAR models

Conclusions and Next Steps

- Teacher-student interaction models and missing data models suggest that ignoring interactions or assuming MAR is not leading to appreciable bias in estimated teacher effects
- Next steps:
  - Vet current findings, including replication on additional datasets
  - Try to work more classroom contextual information into the models (e.g. classroom heterogeneity on δ)
  - Expand set of MNAR specifications and develop exploratory techniques to assess when assumptions about missing data have the potential to create bias