

**Session Title:**

An Investigation of the Behavior of Value-Added Models for Estimating MSP Impact

**MSP Project Name:**

Data Connections: Developing a Coherent Picture of Mathematics Teaching and Learning

**Presenters:**

Walter Stroup, University of Nebraska-Lincoln  
Pamela Fellers, University of Nebraska-Lincoln  
Leslie Lukin, Lincoln Public Schools  
John Sutton, RMC Research Corporation  
Xin Wang, RMC Research Corporation

**Authors:**

Pamela Fellers, University of Nebraska-Lincoln (Lead)  
Wendy Smith, University of Nebraska-Lincoln  
Jennifer Green, University of Nebraska-Lincoln

**Project Session****Strand 3****Summary:**

Appropriate evaluation of MSP programs is essential. One tool for measuring the impact of MSP programs on teaching effectiveness is student achievement data. This requires a coherent picture of student progress before, during, and after a project's professional development program. However, the data available to projects often do not meet the technical requirements of current statistical methods including value-added models (VAMs). Specifically, models assume random assignment of students to classes when, in reality, students are *not* randomly assigned to classes for a variety of educationally legitimate reasons. This project is investigating the impact of non-randomization on the ability of VAM to detect impact of an MSP project. The examination of non-randomization will be presented and discussed in this session.

**Section 1: Questions framing the session:**

- How can value-added models (VAMs) be effectively used to estimate MSP program impact when using student achievement data?
- How are students actually assigned to classrooms in school districts?
- What can we learn about the impact of an MSP program under realistic scenarios of non-randomization that reflect actual practices of assigning students to classrooms?
- What should MSP projects and evaluators plan for data collection to make effective use of VAMs?

**Section 2: Conceptual framework:**

NSF's MSP programs provide substantial content-based professional development to teachers. Through the MSP program, NSF is making a significant investment in the education of our nation's mathematics teachers. There is considerable and increasing interest in finding out which MSP projects appear to be effectively improving the quality and impact of mathematics teaching and thus are candidates for scaling up.

Evaluation of MSP projects is one of NSF's 5 key features (evidence based design and outcomes), and thus is a critical component for providing feedback and important information about the effectiveness of the program. Professional development programs focus on improving teachers' abilities to provide quality instruction, but rigorous evaluations are needed to determine whether these programs are actually effective. *Data Connections* aims to identify, explore, understand, and modify or develop new statistical models to be used for evaluation of MSP projects. Providing key insights into these models and their uses for measuring program impact will inform not only current projects but help future projects plan appropriate evaluation strategies.

As we continue to evaluate and adapt our statistical models, the realities of the data commonly available to MSP programs affect our research. We identify assumptions implicit in current statistical models, and how violations of those assumptions affect the accuracy of the estimates of classroom effects and in turn the detection of a program effect. One key assumption of statistical models that conflicts with educationally-sound school district practice is that students are randomly assigned to classes. In reality, assignment must be based on a variety of criteria that viable statistical models must be able to accommodate.

Value-added models (VAMs) have gained recent popularity (and in some cases, notoriety), as *Race to the Top* dollars dictate that states must incorporate student outcomes into teacher evaluations. Multiple VAMs are being used in a variety of ways, with widely varying results. Within the context of MSP work, *Data Connections* is working to develop VAM techniques that can reliably and stably estimate classroom and school effects and the changes to such effects that can be associated with specific interventions (i.e. MSP professional development). VAM methods provide opportunities to estimate the proportion of variability in achievement or student growth attributable to classrooms, as well as estimate an individual classroom's effect on student learning. Of urgent concern is to develop and explore the opportunities and limitations of VAM techniques for use with less-than-ideal (i.g., real) school district data on student achievement.

We note that a single score on a VAM should never be the sole basis for any type of judgment about teaching effectiveness. Since VAMs have been receiving increasing attention by states, it is worth better understanding VAMs and improving VAM techniques. There are many concerns about VAMs but insufficient research data to investigate which are valid. Thus one of our purposes is to use data to examine various concerns of VAMs. As long as the input to VAMs are student scores on state achievement tests (i.e. single point in time for each student, each year), any results must be interpreted with great caution. When determining the quality of MSP projects, comparative evaluations of student outcomes will inevitably use state student achievement tests. In order to understand how these data characterize MSP projects and

teaching effectiveness, we need to better understand the VAM techniques that can be used with such data.

The purpose of *Data Connections* is to develop statistical models to create a coherent picture of teaching and learning. Our initial focus is on using VAMs to estimate classroom effects, and how those models can be adapted for use with real school district data on student achievement. Specifically, we are currently using a simulation study to investigate the effect of non-randomly assigning students to classrooms on estimating classroom effects and program impact. *Our overall goal is to develop, refine, and disseminate statistical models that develop a coherent picture of mathematics teaching and learning, particularly in regard to MSP programs.*

### **Section 3: Explanatory framework:**

Our MSP is directly studying the intersection between K-12 student success and effective STEM teaching by using student achievement data to estimate classroom effects and in turn estimating MSP program impact. Analysis of student achievement data is challenging when high-quality longitudinal data meeting the conditions for modeling are not readily available. Specifically, one complexity exhibited in the data available to MSP projects is that students are not randomized to classrooms.

To adequately provide information to MSP projects about analysis to estimate program impact our models need to be understood and address the common challenges present in real data. It is thus essential to address the concern of non-randomization structures in current school systems.

Other researchers have started to assess the effect of specific types of non-randomization on classroom estimates. Work to date has focused on ways one teacher might try and scheme the system. However, there is limited research that addresses the general way students are placed in classes and how this non-randomization affects the estimates of classroom effects.

In order to investigate how VAMs are affected by non-randomization simulation studies were conducted. These simulation studies allow us to explore and better understand the properties of estimates under departures from assumptions. Within our simulation studies we used a completely randomized assignment of students to teachers as the basis of comparison. To incorporate a realistic structure of student assignment we worked with information provided from Lincoln Public Schools (LPS).

The simulation study was set up in a way that the same students went through the 6<sup>th</sup>-8<sup>th</sup> grades under both the completely randomized design and the non-random structured design producing two sets of data. Both datasets were analyzed using the same model and estimates of classroom effects were returned for comparison.

We are currently in the process of summarizing our simulation studies and the effect of non-randomization on classroom effect estimates. We will then expand our simulation study design to assess the ability to detect program effects under non-randomization. These results will allow us to report on the accuracy of VAMs when a suggested non-randomization pattern is followed.

Some valuable lessons learned have come from our research thus far, specifically relating to the implications for evaluation. Multiple data sets and the appropriate use of VAMs – properly understood – can help evaluators better document impacts and effects of MSP programs. Additionally, efficiency of evaluation can be enhanced by applying

VAMs to data sets, but doing so requires rethinking evaluation design to accommodate different methodologies. For example, effective use of VAMs requires data over longer time-frames, accurate baseline (pre-study) data, and realization that data are likely to be noisy and the “MSP signal” relatively weak. The latter suggest the need for further discussion about how statistical results are interpreted.

#### **Section 4: Discussion:**

Our current efforts to understand the effect of non-randomization on estimates from VAMs are in the beginning stages. As we gain insight from initial simulation studies we will continue to modify the design incorporating other aspects present in current school systems such as adding more classes per teacher, including the possibility of multiple math teachers in a year (e.g. intervention programs), and integrating representative movement between levels of courses. With the help of other MSP programs there is the possibility of identify additional assignment structures of students to classes to incorporate into the simulation studies for a more comprehensive evaluation.

The insights gained from this research will allow us to address a challenge presented when using VAMs to analyze real data. The ultimate goal is to be able to provide information about statistical models to be used to measure the impact of MSP programs through the use of real data. The more we learn how the models being used behave under conditions present in real data, the better we can inform the MSP community and advise MSPs and their evaluators about their use and limitations.

For evaluators, data generated at the individual MSP project level has potential utility for other MSP projects. Projects and evaluators need to think beyond the scope of the individual project. Potential benefit to all MSPs would come from thinking about data management at the outset of the project – e.g. with a view to making the data generally available. The findings from the VAMs, coupled with other data findings can position the evaluator to become more involved in program design, helping to better think about and contribute to theories of action guiding programs.

**Section 5: How will you structure this session? What is your plan for participant interaction?**

Our session will include a presentation of our current research and exploration of the randomization or lack thereof in the data often available to MSP projects.

Our plan for participant interaction is to get feedback on the assignment scheme used in our simulation study adapted from LPS. We would also like to potentially be able to identify other structures of student assignment to classrooms based on information provided from participants and their knowledge of structures currently used in school systems. These identified structures could potentially be used to expand our simulation study of the impact non-randomization has on our models and their ability to detect program impact on student outcomes.

Time	Activity
1 min:	Project introductions
5 min:	Overview and discussion of challenges encountered to date and changes in direction of study
2 min:	Discuss rationale for using simulation study
5 min:	Explain assignment structure built into the simulation based on specific data set
10 min:	Presentation of initial results and external run of simulation
5 min:	Implications for evaluation
2 min:	Summary and next steps
10 min:	Whole group discussion of alternative assignment schemes and related issues, Q & A