

*The Inter-temporal Stability of Teacher Effect Estimates*

Daniel F. McCafferty, PhD – RAND Corp.

J.R. Lockwood, PhD – RAND Corp.

Tim R. Sass, PhD – Florida State University

Thursday, April 24

10:30am - Noon

---

At both the national and state levels, there are initiatives to link teacher pay to teacher performance. These plans hinge on the ability to measure a teacher's contribution to desired academic outcomes. There is also a small but rapidly growing research literature which attempts to measure teacher productivity by estimating "value added" models of student achievement. In a number of these studies researchers have found the inter-temporal variability of measured teacher performance to be large relative to the variability among teachers within a given year. Such instability would obviously be problematic for any teacher incentive system, since performance and reward must be consistently linked in order to provide the proper incentives.

The potential sources of temporal instability are many, including sampling errors, true variation in teacher performance, student-teacher interactions, interactions among students, properties of the achievement assessments, and artifacts of the estimation methods due to stratification of the student population into disjoint subsets that share no teachers. To date, little is known about the relative contributions of these sources and how different methods may be more or less sensitive to them. Without a clear understanding of the relative contributions of the sources of temporal instability in different types of performance measures, it is impossible to develop methods that effectively balance the competing goals of bias reduction and variance control to achieve measures that will function well for performance pay initiatives. To date, little is known about the relative contributions of these sources and how different methods may be more or less sensitive to them.

In this paper we take a first step in addressing these issues by carefully documenting the extent of inter-temporal instability in teacher estimates and how this variability depends on the both the nature of the data and model specification. We employ data from the Florida K-20 Education Data Warehouse (FLEDW), a rich longitudinal data base that allows us compare the performance of individual students and their teachers throughout the State of Florida over a six-year span. We will use the FLEDW data to estimate teacher effects based a teacher's average contribution to student achievement, holding constant student, peer and school characteristics. We will explore how the number of observations per teacher and the number of observations per student affect the inter-temporal stability of the teacher estimates. We will also investigate how alternative methods of controlling for student heterogeneity impact the variability in teacher effects over time.