# SAS® EVAAS

POLICY BRIEF

# Summary of RAND Research on Value-Added Modeling

## Introduction

According to the RAND Corporation's website, it is a "nonprofit institution that helps to improve policy and decision making through research and analysis."[i] Over the past decade, several of its researchers have focused on measuring schooling and teaching effectiveness, most notably Daniel McCaffrey (now at ETS), J.R. Lockwood and Bing Han. RAND does not provide value-added modeling for profit nor does the organization have an agenda regarding the use of value-added modeling and standardized testing, and its researchers are well respected with regards to their statistical expertise. As a result, the research by RAND is generally considered to be well done, and its conclusions to be technically sound.

Over the years, these researchers have replicated many value-added and growth models, including EVAAS models on a small scale, and their research has provided insight into issues related to both the statistical modeling and the practical application of the modeling. In the early years, RAND researchers raised a variety of concerns regarding value-added modeling, such as whether school and teacher effects were correlated with student characteristics or other sources of bias.

Many of these concerns were subsequently addressed through further research and improved computational ability to use large data sets with complex statistical models. This document summarizes the early concerns as well as the more recent findings by RAND (with emphasis added by SAS on particularly relevant conclusions).

## Early RAND Findings

There are two RAND papers in particular that are often cited by critics of value-added modeling. These papers, from 2003 and 2004, raise a number of potential concerns related to value-added modeling and ultimately conclude that further research is needed to assess the validity of those concerns. The key concerns raised in these papers are discussed below, and the following section illustrates how later RAND research assuaged the researchers' initial concerns.

The two hundred page monograph published in 2003 (and also summarized as a 2004 research brief) covered a number of issues related to value-added modeling, which was new to many in educational policy and research. In the monograph, RAND researchers indicated that "[t]he research base is currently insufficient for us to recommend the use of VAM for high-stakes decisions."[ii] However, this statement is misleading out of context of the monograph. In the same section, the researchers noted that the other, more traditional measures used to evaluate teachers (such as observations) were also susceptible to the same concerns they raised for value-added, and the researchers noted that "VAM might actually provide less-biased and more-precise assessments of teacher effects."[iii] Thus, while RAND researchers approached value-added modeling with their usual caution, their initial assessment indicated that this statistical approach could provide better information than other measures of teacher effectiveness.

A 2004 paper further explored potential sources of bias in teacher effects, and it concluded that:

> *[M]ixed models that account for student correlation over time are reasonably robust to such misspecifications when all the schools in the sample serve similar student populations. However, student characteristics are likely to confound estimated teacher effects when schools serve distinctly different populations.* [iv]

In other words, the researchers had eliminated one potential source of bias (among schools that serve similar populations) but they still had reservations regarding schools that serve very different populations. However, it should be noted that, at this point, researchers were still limited in their computational capacity and the research was based on a very small number of elementary schools (4 or 5) rather than a statewide sample. The paper notes a number of complexities regarding the interpretation of their findings and ultimately suggests that further research is needed.

## Recent RAND Findings

In the past decade, RAND researchers have addressed many of their recommended next steps for research, and their conclusions indicate further understanding and support of value-added modeling as initial concerns have been mitigated.

**Despite RAND researchers' initial concerns about value-added modeling, a 2008 paper summarized the benefits as follows, illustrating just how much had changed since the early research and early concerns about the impact of student characteristics on teacher effects:**

> *There are many studies that support the use of value-added for assessing teachers. Kane and Staiger (2008) find value-added estimates and estimates of teacher performance on randomly assigned classrooms provide similar evaluations of teachers. Cantrell et al (2007) also find that value added scores are the best predictors of teachers' future students' outcomes even among pairs of teachers on randomly assigned classes. <u>This result echoes earlier findings by Sanders and colleagues [SAS EVAAS modelers] showing value-added measures are correlated with students' future outcomes</u>. <u>Many studies show estimated teacher effects are often not correlated with aggregate classroom demographics</u> and a limited number of studies have found them to be related to principal evaluations, teacher observation protocols, and tests of teacher knowledge. Moreover, the <u>implications of violations of assumptions in terms of how large the bias may be and the prevalence of the conditions that lead to violations is not known and may not be severe</u>.* [v]

**The same 2008 paper continues, noting how value-added measure may improve educational practices and be used in high stakes decisions, such as evaluation and merit pay.**

> *Given the weaknesses in the current evaluations, measures of teacher effects on student outcomes could benefit the teaching profession by guiding effective teachers to continue best practices and directing others to seek professional development. Accurate measures of teacher effects on student learning could also serve as the basis for compensation and accountability systems which might motivate teachers who are struggling to produce student learning gains to seek the advancement opportunities they require, reward the most effective teachers in response to younger teachers' and young potential teachers' demands for being rewarded on the basis of their merits, and attract a broader range of professionals to the profession.* [vi]

**This 2008 paper concludes in noting that while robust value-added measures <u>may</u> not be perfect, they are preferable to other measures used in teacher evaluations, which is similar to the 2003 assessment but more strongly worded.**

*Other uses of student test scores suffer from all the same shortcomings of value-added and more. Thus, even though value-added is potentially less than ideal, it may have significant advantages over other test-based accountability and evaluation systems.*[vii]

**Furthermore, RAND researchers emphasize that complex statistical models are necessary to provide reliable estimates of teacher effects.**

*Currently there is limited consensus in the VAM literature about the best analytic approaches. Advocates of ad hoc methods argue that transparency is the primary concern for providing useful VAM-A estimates. However, quantitative researchers from both the statistical and economic modeling traditions argue that complex methods are likely to be necessary for accurate estimation of teacher effects and that accountability or compensation systems based on performance measures with weak statistical properties will fail to provide educators with useful information to guide their practice and could eventually erode their confidence in such systems.*[viii]

**In another paper from 2007, RAND specifically advocated a longitudinal, mixed model approach, which is used in all SAS EVAAS models (note the emphasis on unbiased estimates in context to the 2004 paper).**

*When applied to longitudinal data involving a large number of correlated measurements on individuals, [mixed model methods] can provide nearly unbiased estimates even under relatively complex heterogeneity models involving multiple, unobserved individual-specific attributes whose relationship to the observed measurements varies across those measurements.*

*The ability of the mixed model approach to perform well simultaneously for a number of parameters that grows as the number of test scores grows is particularly relevant for estimating individual teacher effects.*[ix]

**A 2012 paper from RAND specifically noted the importance of using all available testing data for students:**

*Simple value-added models that control for just a few tests scores (or only one score) and no other variables produce measures that underestimate teachers with low-achieving students and overestimate teachers with high-achieving students.*[x]

**A 2011 paper compared many models used for teacher effects, and it concluded that the statistical models similar to EVAAS were best suited for teacher evaluation. In fact, not only did RAND support the use of complex statistical models like EVAAS; its researchers explained the technical problems inherent in simple value-added models.**

*Gain scores and ANCOVA are simple to implement, transparent, and widely used. In practice many people equate these methods with value-added modeling. Simple gain scores are more widely used by practitioners and many economic applications have used simple ANCOVA-like approaches (for example, Aaronson, Barrow, and Sander, 2003, Kane, Rockoff, and Staiger, 2006). Both these methods are relatively poor performing… the gain score measures appear to have most noise relative to other sources of variability in the estimator and the ANCOVA measures appear to have the largest noise component.*[xi]

**Even a relatively complex econometric fixed-effects model is not sufficiently sophisticated to overcome common statistical problems.**

> *Fixed effect methods also are challenging to compute especially for large samples with large numbers of teachers and students and special care is required to account for the over identification of the model that includes effects for all students and teachers.* [xii]

**The fact that students have missing data, and that these data are not missing at random, is one of the primary reasons why a complex statistical model is required for value-added estimates. The 2007 paper by RAND noted how the mixed model approach is preferred among others in how to address students with missing test scores.**

> *This suggests that our general findings about the bias compression of the mixed models approach are not invalidated by the complexities of missing data, but it is likely that incompleteness in the test score data will in general degrade the bias compression to some extent. On the other hand, the mixed models approach makes use of all of the information available for each student in estimating the unknown parameters … and so might lead to particular efficiency gains relative to other approaches when missing data are substantial.* [xiii]

**The RAND validation of how EVAAS models address students with missing data (using a "missing at random" approach) was reinforced in a 2011 paper as well:**

> *We find that allowing the data to be missing not at random has little impact on estimated teacher effects. The robustness of estimated teacher effects to the missing data assumptions appears to result from both the relatively small impact of model specification on estimated student effects compared with the large variability in teacher effects and the down weighting of scores from students with incomplete data.* [xiv]

## Conclusion

Throughout the past decade, the RAND studies have introduced and answered a variety of questions related to value-added modeling. By nature, the researchers take a fairly academic approach, where potential limitations and/or concerns are often noted. Despite this conservative stance, there is a clear evolution in the understanding and support of the RAND researchers towards statistically robust and reliable value-added models.

---

i www.rand.org; accessed March 2013.

ii McCaffrey, D.F., J.R. Lockwood, D.M. Koretz, and L.S. Hamilton. (2003). Evaluating Value-Added Models for Teacher Accountability. RAND Monograph No. 158 (quote from p. 119).

iii McCaffrey, D.F., J.R. Lockwood, D.M. Koretz, and L.S. Hamilton. (2003). Evaluating Value-Added Models for Teacher Accountability. RAND Monograph No. 158 (quote from p. 121).

iv McCaffrey, D.F., J.R. Lockwood, D.M. Koretz, T.A. Louis and L.S. Hamilton. (2004). Models for Value-Added Modeling of Teacher Effects. Journal of Educational and Behavioral Statistics. 29(1): 67-101 (quote from Abstract).

v McCaffrey, D.F. and J.R. Lockwood (2008). Value-Added Models: Analytic Issues. Presented at the Workshop on Value-Added Modeling sponsored by the National Research Council and the National Academy of Education, Board on Testing and Accountability, Washington, DC, Nov. 13-14, 2008 (quote from p. 12).

vi McCaffrey, D.F. and J.R. Lockwood (2008). Value-Added Models: Analytic Issues. Presented at the Workshop on Value-Added Modeling sponsored by the National Research Council and the National Academy of Education, Board on Testing and Accountability, Washington, DC, Nov. 13-14, 2008 (quote from p. 12).

vii McCaffrey, D.F. and J.R. Lockwood (2008). Value-Added Models: Analytic Issues. Presented at the Workshop on Value-Added Modeling sponsored by the National Research Council and the National Academy of Education, Board on Testing and Accountability, Washington, DC, Nov. 13-14, 2008 (quote from p. 12).

viii McCaffrey, D.F. and J.R. Lockwood (2008). Value-Added Models: Analytic Issues. Presented at the Workshop on Value-Added Modeling sponsored by the National Research Council and the National Academy of Education, Board on Testing and Accountability, Washington, DC, Nov. 13-14, 2008 (quote from p. 10).

ix Lockwood J.R. and D.F. McCaffrey (2007). Controlling for individual heterogeneity in longitudinal models, with applications to student achievement. Electronic Journal of Statistics, 1: 223-252 (quote from p. 27).

x McCaffrey, D.F. (2012). Do Value-Added Methods Level the Playing Field for Teachers? What We Know Series. Carnegie Knowledge Network (quote from p. 2).

xi McCaffrey, D. F., B. Han, and J.R. Lockwood. (2008). From Data to Bonuses: A Case Study of the Issues Related to Awarding Teachers Pay on the Basis of the Students' Progress. Presented at the Conference on Performance Incentives: Their Growing Impact on American K-12 Education, Feb. 28-29, National Center on Performance Incentives at Vanderbilt University (quote from p. 28).

xii McCaffrey, D. F., B. Han, and J.R. Lockwood. (2008). From Data to Bonuses: A Case Study of the Issues Related to Awarding Teachers Pay on the Basis of the Students' Progress. Presented at the Conference on Performance Incentives: Their Growing Impact on American K-12 Education, Feb. 28-29, National Center on Performance Incentives at Vanderbilt University (quote from p. 28).

[xiii] Lockwood J.R. and D.F. McCaffrey (2007). Controlling for individual heterogeneity in longitudinal models, with applications to student achievement. *Electronic Journal of Statistics*, 1: 223-252 (quote from p. 29).

[xiv] Lockwood J.R. and D.F. McCaffrey (2011). Missing Data in Value-Added Modeling of Teacher Effects. *The Annuals of Applied Statistics*. 5 (2A): 773-797 (quote from Abstract).