



Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. New York: Routledge.

Pp. 272

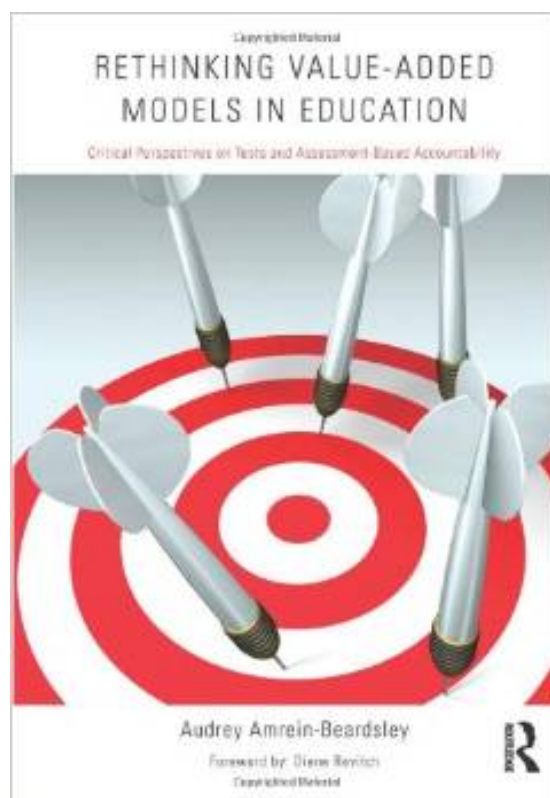
ISBN: 978-0415820127

Reviewed by *Rachael Gabriel*

University of Connecticut
United States

Dr. Amrein-Beardsley's recent book, *Rethinking Value-Added Models in Education* (Routledge, 2014), is the single most comprehensive resource on the uses and abuses of Value-added Measurement (VAM) in U.S. education policy. As the centerpiece of several new generation teacher evaluation policies, VAM has been the subject of a firestorm of media attention, position statements, journal articles, policy briefs and a wide range of scholarly debate.

Oddly, or perhaps thankfully, for all its dramatic heralding, only twelve states have recently adopted VAM as part of their teacher evaluation or compensation policies - due in no small part to the criticism Amrein-Beardsley and others have made available to the voting public (e.g., Amrein-Beardsley & Collins, 2012; Baker et al., 2010; Baker, Oduwale, Greene, 2013). Though only twelve states have VAM written into their teacher evaluation policies, others use it at the district or school level, for purposes as varied as program evaluation, teacher compensation, educational research and legal challenges. Therefore, this book



isn't just for the hundreds of thousands of teachers who live and work in those twelve unlucky states (e.g., New York, Florida, Tennessee, Ohio, etc.), it's about a measurement tool that captured America's imagination - convincing us that we could accurately and reliably measure a teachers' impact on their students - against her better judgment.

There is something for everyone in this text: from definitions for those with no background in statistics to thorough discussions of the reining positions and debates between researchers. It therefore serves as a primer on education policy as well as a deep dive into the specifics of VAM. The text rises far above other options for understanding VAM because it successfully combines technical detail with social activism. This not only displays the range of Amrein-Beardsley's thinking about the subject, but also the depth of her understanding of VAM's technical merits and social implications.

For example, Amerin-Beardsley uses the introductory chapter to establish the place of VAM in the history of social policy and recent education policy. VAM is framed as one example in a long history of social engineering experiments, in line with what she calls the "Measure & Punish (M&P) Theory of Change" that characterizes most contemporary education reform efforts. Within the introduction, Amrein-Beardsley also crafts a refreshingly unapologetic statement of her own position on the issue - linking it to her background as a mathematics teacher and educational researcher. This transparency and context allows her to relinquish claims to neutral objectivity and replace it with clearly argued, but passionate outrage, which makes the story of VAM both compelling and urgent.

Historicizing VAM and its surrounding controversies allows the reader to assume some analytic distance, though these debates are still very real and very raw. Within the 211 pages, she addresses VAM as both a statistical tool with specific features and assumptions, and a policy tool with socially constructed meanings and implications. This dual treatment, and the belief that data does not and never could "speak for itself" constructs a version of VAM in which the tool takes

on a life of its own, with people and policies positioned and defined in response to its divisive construction (Gabriel & Lester, 2013). This characterization of VAM is also what makes this text so readable. VAM itself becomes a fascinating character in a larger story about education policy. Its role in recent policies, public debates and lawsuits is nothing short of operatic in quality. The sweeping nature of the story of VAM mixed with details on its technical merits creates a text that is encyclopedic in scope: with entire chapters devoted to discussions of VAM's assumptions, reliability and biases.

Amrein-Beardsley's emotion as a writer (sometimes anger, sometimes passion) simmers tangibly below the surface of her prose, and explodes once in a while in a long, complex, exclamation-pointed sentence. It seems that when you know as much as she does about VAM, and have to face the general ignorance and perverted rhetoric on the subject, neutrality isn't an option. This treatment of the topic as one of urgency and consequence is not only rooted in a moral sense of outrage, but a deeply personal sense that the public and its teachers are being misled and mischaracterized. As Amrein-Beardsley points out, the fact that the public has been "VAMboozled" is not unique in the history of education or other social policy. It is, however, avoidable, if the public has access to a fuller understanding of what's involved and what's at stake when policies include VAM. This book represents one of many steps towards such access.

As such, each chapter read like transcripts of a Ted Talk, with multi-part lists (e.g., "the eight things that are now evident about VAM"), explained with tight examples and logic checks along the way. The use of multi-part lists, extensive citations and the added feature of a list of "top ten assertions" at the end of each chapter allow readers to pick and choose the sections of greatest interest. This is important because the combined depth and breadth of information about VAM can make some chapters feel skimmable. For example, those who are most interested in how sorting bias occurs, are not likely to need the explicit definitions of terms like

reliability and validity. Those who are interested in the history of VAM may not need the context in the larger scheme of education reform, or the history of government involvement in social services. Still, since so many arguments about VAM are based on popular, vague and/or mixed understandings of its foundational concepts and tools of logic, this thorough treatment is a welcome addition to the literature. The end-of-chapter assertion boxes also demonstrate Amrein-Beardsley's unique ability to lay it *all* out there: the good, bad and the normally obfuscated "truth" about VAM.

The crowning feature of this book is its final chapter, in which alternatives and solutions to VAM are presented. Though Amrein-Beardsley buys into the current fascination with using "multiple measures" for determining teacher effectiveness, she is quick to point out that combining the strengths and weaknesses of imperfect indicators "does not increase overall levels of reliability and validity" (p. 211). Instead, she advocates for the strategic pursuit of face validity, in the absence of any plausible construct validity, by suggesting that individuals' professional judgment and local definitions of effectiveness be weighed as heavily as statistical markers. Her unique solution, one that has not yet been empirically validated or even widely discussed in policy circles, is essentially a panel of supervisors and peers evaluating and rating based on local definitions of effectiveness, and informed by local and/or research-based measures. She argues that multiple local stakeholders are the most powerful tool for interpreting observable indicators effectiveness.

For those things that we cannot see - the outcomes or outputs of effective teaching (e.g., test scores, graduation rates, etc.) - both local and externally validated tools should be used. This proposal is unique in its insistence on multiplicity and its faith in the importance of human judgment. People (not a single representative of states or districts as collectives) should be responsible for selecting *and* designing (not one or the other) the measures of effectiveness. And people, (not observation tools or even single observers) should be responsible for discussing and analyzing the

inputs and processes of effectiveness in teaching.

By way of a summary of this suggested solution, Amrein-Beardsley writes:

This solution does not rely solely on mathematics and the allure of numbers or grandeur of objectivity that too often comes along with numerical representation, especially in the social sciences.

This solution does not trust the test scores too often (and wrongly) used to assess teacher quality, simply because the test output is already available (and paid for) and these data can be represented numerically, mathematically, and hence objectively.

This solution does not marginalize human judgment, but rather embraces human judgment for what it is worth, as positioned and operationalized within a more professional, democratically-based, and sound system of judgment, decision-making, and support.¹

Current teacher evaluation policies that claim to rely on "multiple measures of effectiveness" may still only require one observer's opinion, and the combination of one-of-each type of measure for student achievement, student growth and/or other outcomes. This combination of single probes stands in contrast Amrein-Beardsley's proposed combination of multiple probes for each aspect of an evaluation (inputs, processes, outputs).

Also unique to Amrein-Beardsley's solution is the faith in individual teachers and supervisors to act as professionals - a faith that many researchers question as study after study points out the weak correlations between principal ratings and student achievement scores. For the most part, media outlets and even researchers have chalked up discrepancies to human error - blaming individual judgment for any discrepancy between observation scores and VAM ratings. Far from being surprised or discouraged by these discrepancies, Amrein-Beardsley suggests just the opposite: that VAM ratings are the data point that needs revision and oversight, and the professional judgment of educators is what should be held as the gold standard.

This faith that individuals act in the best

¹ <http://vamboozled.com/author/audrey-amrein-beardsley/>

interest of students and the profession is also reflected in how Amrein-Beardsley recounts the stories of Houston area teachers who were fired based on VAM ratings. Their value as professionals is never called into question. Rather the value of VAM as a policy tool is put on trial, and it loses miserably. This respect for teachers and belief in people-not-numbers pervades other examples and arguments across the eight chapters. It is clear that Amrein-Beardsley is writing to an audience she believes in, and one that deserves to be equipped with all the facts.

The depth and breadth that characterizes this text is in itself a testament to Amrein-Beardsley's belief in the power of people in plural. She sets out to define and explain a topic that is known for being shrouded in political spectacle (Gabriel & Allington, 2011). Despite the incredible volume of reporting and editorializing VAM has seen in the news, she examines still-unexamined assumptions at length, assuming that there are those in the world who want and need to have them explained. In other words, she dares to believe that people want and need to know the whole story, and that knowing this could change something about how we view teachers, testing, measurement and policy.

Finally, this belief in people is underscored by the book's dedication page. A page in the front matter shows a picture of the Cambodian orphanage to which proceeds from the book will be donated. If nothing else good comes of VAM's ignominious presence in education research literature, it is good that we have the whole story laid out here. And it is good that what we invest in reading and sharing this text will mean something beyond its audience of US readers. From Cambodia to Houston, this text represents one more step in a mission to bring a collective intelligence, compassion and truth to bear when issues policy become threats to equity, despite our better judgment.

References

Amrein-Beardsley, A. & Collins, C. (2012). *The SAS Education Value-Added AssessmentSystem (SAS® EVAAS®) in the Houston Independent School District*

- (HISD): Intended and Unintended Consequences. Education Policy Analysis Archives, 20 (12). Retrieved from <http://epaa.asu.edu/ojs/article/view/1096>.
- Baker, E., Barton, P., Darling-Hammond, L., Haertel, E., Ladd, H., Linn, R., et al. (2010). *Problems with the Use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute.
- Baker, B.; Oluwole, J.; Greene, P. (2013). The legal consequences of mandating high stakes decisions based on low quality information: teacher evaluation in the race-to-the-top era. *Education Policy Analysis Archives*. 21. Retrieved from: <http://epaa.asu.edu/ojs/article/view/1298>
- Gabriel, R., & Lester, J. N. (2013) The romance quest of education reform: A discourse analysis of The LA Times' reports on value-added measurement teacher effectiveness. *Teacher's College Record*. 115(12).
- Gabriel, R. & Allington, R (April, 2011). *Teacher Effectiveness Research and the Spectacle of Effectiveness Policy* American Educational Research Association, New Orleans, LA.

About the Reviewer

Rachael Gabriel

Assistant Professor of Reading Education
University of Connecticut
rachael.gabriel@uconn.edu



education review // reseñas educativas
a multi-lingual journal of book reviews

editors: **gustavo e. fischman /**
melissa cast-brede / gene v glass

Supported by the Mary Lou Fulton Teachers College, Arizona State University

 *Education Review/Reseñas Educativas/Resenhas Educativas* is supported by the edXchange initiative's Scholarly Communications Group at the Mary Lou Fulton Teachers College, Arizona State University. Copyright is retained by the first or sole author, who grants right of first publication to the *Education Review*. Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and ***Education Review***, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-sa/3.0/>. All other uses must be approved by the author(s) or ***Education Review***. ***Education Review*** is published by the Scholarly Communications Group of the Mary Lou Fulton Teachers College, Arizona State University.

Please contribute reviews at <http://www.edrev.info/contribute.html>.

Connect with *Education Review* on Facebook
(<https://www.facebook.com/pages/Education-Review/178358222192644>) and on Twitter
@EducReview