education review // reseñas educativas
editors: david j. blacker / gustavo e. fischman / melissa cast-brede

a multi-lingual journal of book reviews

September 22, 2012 ISSN 1094-5296
*Education Review/Reseñas Educativas* is a project of the National Education Policy Center http://nepc.colorado.edu
Follow *Education Review* on Facebook and on Twitter: #EducReview

Harris, Douglas N. (2011) *Value-Added Measures in Education: What Every Educator Needs To Know*. Cambridge, MA: Harvard Education Press.

Pp. ix + 276 ISBN 978-1-61250-000-3

Reviewed by Amanda Slattery & Edward Garcia Fierros
Villanova University

Methods of evaluating teacher effectiveness have become a prominent topic of debate in education forums. In the past years, value-added measures have recently gained attention as a viable alternative to "snapshot" measures produced through student proficiency marks on standardized state tests.

Many scholars agree that current methods of quantifying teacher and school effectiveness (including annual administrative reviews and student test scores) are not working (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2011). Limitations of the current system of teacher evaluation include: the failure to provide a comprehensive review of teacher instructional practices, a

Citation: Slattery, Amanda & Edward Garcia Fierros. (2012 September 22) Review of *Value-added Measures in Education* by Douglas N. Harris. *Education Review, 15.* Retrieved [Date] from http://www.edrev.info/reviews/rev1181.pdf

This review was accepted under the previous Editor for English, Gene V Glass.

failure to provide specific feedback on teacher strengths and areas for improvement, instability of value-added statistical models, and a failure to distinguish superior teachers from weak teachers (Darling-Hammond, et al., 2011). Many argue that the system is failing our students as the age of standardized testing has led to test-driven instruction as teachers experience increasing pressure to "teach to the test" so that their school (and in some cases, their own professional reputation) receives high marks (Darling-Hammond *et al*., 2011). Although these shortcomings are well recognized, the solution to these issues is hotly debated. Out of these conversations, value-added measures of teacher and school effectiveness have gained attention in recent years as viable measurement tools. Furthermore, the value-added measures as part of teacher evaluation has become an essential element of President Obama's educational policy which seeks to "hold individual teachers accountable for the performance of their students on standardized tests" (Klein, 2012, p. 26).

Harris's book, *Value-Added Measures in Education*, enters the arena in the midst of this debate. Harris provides an organized, clear and formulaic argument for using value-added measures as a component of school-level evaluations. The book follows a point-counterpoint model in which Harris describes value-added measures and how they are calculated, he then addresses some of the drawbacks of value-added, finally concluding that we cannot afford to keep our educational evaluation methods as they are, and that of the options currently available, value-added measures are the best alternative. Those who disagree with Harris claim that alternate evaluation tools, such as student portfolios and classroom observations (which Harris discounts as unfeasible) are stronger tools for evaluation (Darling-Hammond *et al*., 2011; Amrein-Beardsley, 2008).

Throughout the book, Harris uses metaphors to make statistical concepts accessible to the average reader. He also includes a district-level case study, referencing this district throughout the work as a baseline to demonstrate how various methods of evaluation would work with real data. The book is admittedly written for teachers and administrators, as Harris seeks to elucidate the mystery

surrounding value-added measures in education throughout the work.

To frame his argument, Harris begins by describing and defining key terms that will be used throughout the book in reference to testing methods, measurements, and reporting measures. Harris presents a strong critique of the "snapshot" measure of school and teacher evaluations currently employed to measure effectiveness. He suggests that the proficiency standards used in accordance with the Elementary Secondary Education Act (ESEA) or No Child Left Behind (NCLB) are ineffective because the proficiency line is arbitrary and NAEP scores demonstrate that proficiency standards differ state-to-state. In addition, snapshot measures fail to account for factors outside of a school's control, allow for complacency in high-performing schools, drive schools to exclude students who may not perform well on tests (such as special needs students), push good teachers out of "failing" schools, and create immense frustration and reduced motivation in educators. For all of these reasons, Harris argues, a new method of evaluation is needed.

Harris concludes that a measure of student growth would be a more appropriate evaluation tool than the existing status model. He systematically covers various measures that have been used to determine student growth, including: growth to proficiency, cohort-to-cohort growth, and individual growth measures. He argues that individual growth using a value-added model is the best method available for wide-scale use, contending that the use of value-added measures eliminate each of the problems with the snapshot measure previously described; that with value-added measures, factors outside of school control are taken into consideration, high-performing schools would have to prove they are, indeed, high performing, there is no incentive to exclude students, teachers would have a motivation to stay in "failing" schools, and Harris claims, administrators have rated value-added measures as more accurate assessments of their performance than AYP scores. Other researchers contend this final thought, reporting that teachers and administrators do not see value-added as an accurate reflection of their annual progress (Darling-Hammond *et al*., 2011).

In chapter four, Harris turns to his goal of clarifying the calculation and mathematical principles underlying value-added measures. It is at this point that the book moves away from a theoretical framing of value-added and into the statistical components of computing a value-added scores. Harris conceptually describes how basic value-added and advanced value-added scores are produced. Basic value-added scores simply compare school growth to overall state growth, while advanced value-added attempt to control for school-level factors through the statistical computation. These specific "school-level factors" could vary from class size, to school resources, to student demographics. Harris discusses the wide array of school-level factors that could be considered when determining an advanced value-added score. Although he does a good job of putting "all of the cards on the table," these multiple methods of computing value-added scores is one of the critiques of the value-added system; where do we draw the line between factors that influence student learning "enough" to be calculated into an assessment that could have significant impacts on students, teachers, and schools. Harris proposes three rules in response to this dilemma: 1) Hold people accountable for what they can control, 2) Hold people at least party accountable for important factors they partly control, 3) Attach stakes to performance measures that are proportional to the quality of those measures; that is, inversely related to the degree of statistical error.

Harris then provides a run through of the various types of statistical errors for his readers who are unfamiliar with statistics. The math concepts and research fundamentals referenced throughout the book are well explained with both graphic and metaphorical representations of the concepts. Harris explains that value-added measures increase the chance for random statistical error, while decreasing the risk of systematic error.

Systematic error still exists, however, in value-added measures. Many critics of value-added methods of evaluation feel that these limitations of value-added measures negate their value (Amrein-Beardsley, 2008; Darling-Hammond, *et al*., 2011; Ladd & Walsh, 2000; Rothstein, 2009). For instance, value-added scores cannot account for tracking in schools (in which students are purposefully placed with certain teachers), and honors

students who score in the top percentiles on state tests have nowhere to go but down, placing their teachers in a position in which it is actually impossible to receive a high value-added score (Amrein-Beardsley, 2008; Rothstein, 2009). Darling-Hammond and others maintain that "teachers' [highly unstable] value-added ratings are significantly affected by differences in the students who assigned to them" (2011, p. ii). This instability was brought to light in Brigg and Domingue's (2011) reexamination of value-added analysis teacher effectiveness of Los Angeles teachers. The flawed value-added analysis conducted for the Los Angeles Times report of teacher effectiveness was presented as a causal model when few experts agree that value-added models can deliver rankings with such certainty. For example, student mobility and test-alignment to state standards are uncontrolled influences on value-added scores that were not considered by the *LA Times* teacher effectiveness rankings, but are import to consider.

Given these errors, Harris continues to support the use of value-added measures to evaluate schools, but not to evaluate teachers. Harris recognizes that errors are more likely to be magnified with teacher value-added scores. In addition, there is a large variance in year-to-year teacher value-added scores. This high variance in teacher value-added results is one of the strongest critiques for value added measures. (Darling-Hammond, *et al.*, 2011) Yet Harris still supports the use of value-added measures to evaluate schools, noting that statistical error is reduced given the large population of schools. However, there are still limitations with school value-added measures, including the fundamental question of to what extent do we statistically account for school resources when determining school-level value-added scores?

Harris writes that the most appropriate measure of value added methods is one in which the value-added scores are calculated under two conditions: 1) adjustments are made for sampling error, and 2) the score combines value-added measures over two-to-three years. This is not how the majority of districts are currently calculating value-added scores. This model would only be useful for school-level evaluations because teachers would not receive value-added feedback during their first two-to-three years of teaching. Further, if value-added measures were used to

evaluate teachers, in this model, the value-added score would be irrelevant because many teachers would have received tenure by the time their value-added results would be available.

Harris purports that value-added measures can improve teaching and learning and he makes several recommendations in order that they do so. To begin, he reiterates that value-added measures should be used only for school-level evaluation and accountability. This process, he warns, should be implemented carefully after pilot programs are run and meticulous consideration is made the ways in which value-added scores are used. Furthermore, value-added measures, he suggests, should be used to determine how well schools are serving certain groups of marginalized students while also identifying effective practices and programs that can be replicated elsewhere. Harris proposes that teacher performance measures be created through a combination of value-added scores and other measures more closely tied to teaching practices. He recommends using the National Board of Professional Teaching Standards, which is consistent with Darling-Hammond *et al.*'s (2011) proposal for teacher evaluations. Furthermore, Harris acknowledges that individual student scores should to be tied to individual teachers and individual schools in order to monitor growth, even if these scores are not linked to teacher evaluations.

To demonstrate these concepts in action, Harris describes some of the current value-added practices employed by school districts or states that are used in the teacher and/or school evaluation process. Harris discusses the Tennessee Educational Value-Added Assessment System (TEVAAS), one of the oldest and most well known value-added approaches in educational assessment. Yet, the program used (EVAAS), developed by Bill Sanders, remains Sanders' intellectual property and as such, an audit of the statistical procedures employed by the system is not possible (Amrein-Beardsley, 2008). Thus, one of the oldest value-added evaluation measures in the country does not have public methodological accountability.

Harris supports the TVAAS system because teacher-TVAAS scores are used in low-stakes ways, however, up to 50 percent of a teacher performance rating can be based

on his or her TVAAS score and teacher development plans are created from these measures. Several of the other value-added programs Harris describes also use value-added measures to evaluate teachers, some even linking teacher value-added scores to merit pay programs. As he discusses these programs, Harris acknowledges their strengths and weaknesses but rests his argument on the potential that value-added measures have to improve teaching and learning.

Overall, this book is useful for anyone seeking to understand the statistical methods and computation factors associated with value-added measures. Harris provides a clear overview that is accessible to the non-statistician. However, his overarching message that we should use value added measures because they are the best evaluation method available is too limiting. Harris writes, "We do not just need better performance measures, but a new way of thinking. The essence of education is about taking students where they start, even far down on the ladder of readiness, and helping them grow as much as they can" (p. 229). And yet Harris's "new way of thinking," namely, using value added measures, relies heavily on the old system of student test scores. We find ourselves at a crossroads in which we can continue to use test-based evaluations or we can work to create evaluation tools that are currently only imagined, tools which provide our students with feedback on their learning progress, teachers with feedback on the effectiveness of their teaching practices, and schools with feedback on their ability to develop student learners. Value added measures may be one piece of such an evaluation, but we should not settle for value added simply because it is all we have right now.

## References

Amrein-Beardsley, Audrey. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher, 37*(2), 65-75.

Briggs, Derek & Domingue, Ben. (2011) Due Diligence and the Evaluation of Teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles Unified School District teachers by the Los Angeles Times. Boulder, CO: *National Education Policy Center*. Retrieved June 18,

2012 from http://nepc.colorado.edu/publication/due-diligence.

Collins, Clarin & Amrein-Beardsley, Audrey. (2011 November 23) Review of Value-added measures in education by Douglas N. Harris. Education Review, 14. Retrieved September 21, 2012 from http://www.edrev.info/reviews/rev1126.pdf

Darling-Hammond, Linda; Amrein-Beardsley, Audrey; Haertel, Edward H., & Rothstein, J. (2011) Getting teacher evaluation right: A background paper for policy makers. *Research Briefing.* Washington DC: American Educational Research Association, National Academy of Education.

Klein, A. (2012). Obama uses funding, executive muscle to make often-divisive agenda a reality. *Education Week, 31,* (35). p. 26.

Ladd, H. F., & Walsh, R. P. (2002). Implementing value-added measures of school effectiveness: Getting the incentives right. *Economics of Education Review, 21*, 1-17.

Rothstein, J. (2009) Student sorting and bias in value added estimation: Selection on observables and unobservables. *Education Finance and Policy, 4,* (4) p. 537-571.

Rothstein, J. (2011). Review of "Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project." Boulder, CO: *National Education Policy Center*. Retrieved 6/18/2012 from http://nepc.colorado.edu/thinktank/review-learning-about-teaching.

## About the Reviewers

Amanda Slattery is a graduate student in the Department of Education and Counseling at Villanova University, where Edward Garcia Fierros is an Associate Professor and the Department Chairperson.

Copyright is retained by the first or sole author, who grants right of first publication to the *Education Review*.

*Education Review/Reseñas Educativas* is a project of the National Education Policy Center http://nepc.colorado.edu

Editors

David J. Blacker

blacker@edrev.info

Gustavo Fischman

fischman@edrev.info

Melissa Cast-Brede

cast-brede@edrev.info

Gene V Glass

glass@edrev.info

Follow *Education Review* on Facebook

and on Twitter: #EducReview