



Value-Added Measures in Education: The Best of the Alternatives is Simply Not Good Enough

by Audrey Amrein-Beardsley – January 12, 2012

*On September 8, 2011 Teachers College Record published a book review of Douglas N. Harris's recent book *Value-Added Measures in Education*. In this commentary the author takes issue with not necessarily the book's *What Every Educator Needs to Know* content but the author's overall endorsement of value-added, and his and others' imprudent adoption of some highly complex assumptions.*

On September 8, 2011 *Teachers College Record* published a book review written by Senior Researcher Michael Strong who reviewed and endorsed economist Douglas N. Harris's recent book *Value-Added Measures in Education: What Every Educator Needs to Know*. I would like to point readers to Strong's book review to read about the general and specific contents of the book, and also point readers to the actual book to read a well-packed, solid, one-stop resource including pretty much everything there is to know about value-added. This includes practical information about value-added statistics and models, systems and policies, summative and formative (mis)uses, and also some value-added fallacies and loose certainties. Of issue in this commentary is Harris's overall and Strong's secondary endorsement of value-added.

I dutifully respect the right to disagree with both of my esteemed colleagues about their general tenets and beliefs, and more specifically about their overall message. I write here a commentary declaring, first, that there should always be two reviews to every book. More importantly, I write here about the fundamental assumptions Strong, Harris, and many others continue to make and marginalize when writing about value-added, especially when writing on its behalf. These assumptions are too important to be taken for granted (see also Scherrer, 2011), especially as this topic is potentially, and in reality, becoming increasingly more consequential.

The conundrum: "Do we have to prove beyond a reasonable doubt that value-added is better than the status quo and viable alternatives? Or, do we think there are enough problems with the current systems that the burden of proof is really on the credentialing/checklist status quo?" (Harris, 2011, p. 169).

Depending on where you stand, you too will devalue or appreciate *Value-Added Measures in Education*. While throughout the book Harris takes a reasonable, cautious, and at times overly compromising (and confusing) approach, the bottom line is this: Because value-added measures are better than the status, snapshot measures used in decades past (with this assertion most topical researchers agree), value-added even given its considerable errors and imperfections is good enough for wide-scaled adoption, hyper-implementation, and (ab)use. This is what Harris defines as his "more productive middle ground" (p. 3). However, this negotiated stance is inherently flawed given the fundamental assumptions Harris and others make, and presumptuously acknowledge as "givens."

Assumption #1: Value-added should be used to measure teacher and school quality. While it stands to reason that value-added measures provide one indicator of change in student achievement as students' test scores are used to measure value-added, whether aggregating these scores at the teacher and school levels yields adequately valid inferences about teacher and school quality is still grossly uncertain. Harris does not reference at least the teacher attribution criticisms and debates about the proportion of value-added credit teachers deserve, outside of using basic proportions to account for things like team teaching (p. 115).

Assumption #2: Value-added should be used to hold teachers and schools accountable for what "they do control and contribute to student learning" (p. 4; see also p. 52). Harris's cardinal rule is that people should be held accountable for what they control. This makes sense, and espouses value-added as the superior choice, primarily because value-added accounts for students' entering achievement when measuring change. However, Harris changes his cardinal rule from holding people accountable for what they can control to holding people accountable "even for factors only partly within their control" (p. 82). Harris attempts to defend this stance, assembling rules about controllable, partially controllable, and uncontrollable variables, yet the only thing clear and consistent is Harris's inbuilt logic that when complicated, protest the subpar alternative to advance the superior value-added.

Assumption #3: For what teachers and schools do not control, "advanced value-added" (p. 87) technicians should statistically reduce the sizeable errors to acceptable and tolerable levels (assuming such levels exist). There are, and may forever be, too many errors thwarting the applicability of value-added (Darling-Hammond, 2010; Ravitch, 2010; Rothstein, 2010). These systematic and random errors are caused by things like the issues already inherent in standardized tests, students who differentially care about their test performance and how their scores might influence their teachers, missing longitudinal data, prior teachers' residual effects, summer school augmentation and summer learning decay, the (non) random assignment of students into classrooms (although "tracking" rarely occurs in elementary schools, p. 114), peer and

teacher interaction effects, home and other construct (ir)relevant variables, and the like. Value-added is far from where it needs to be for much, and in particular high-stakes decision-making (Capitol Hill Briefing, 2011; Baker, Barton, Darling-Hammond, Haertel, Ladd, Linn et al., 2010; Scherrer, 2011).

Assumption #4: Value-added output can help schools and teachers diagnose needs, and accordingly reform and address change for the better. Evidence suggests that thus far teachers and administrators do not seem to understand the value-added models being used to evaluate them and, accordingly, are less likely than believed to formatively use value-added data to reform and improve curriculum and instruction. This is especially true if professional development is not provided (Amrein-Beardsley, 2008; Capitol Hill Briefing, 2011; Eckert & Dabrowski, 2010). While the Value-Added Research Center (VARC) at Harris's home institution is conducting the most comprehensive research on value-added data use, results are still very preliminary and have not yet indicated any such truths, especially as cited (p. 163).

Assumption #5: "Arguments against value-added-based accountability have been rooted in a researcher perspective that is really not appropriate for policy decisions" (p. 8). The assumption here is that things like reliability and validity, both of which value-added is deeply lacking, are not to matter much in the policy arena.

It may sound strange - even self-defeating - for a college professor to tell policy makers not to look at research the way researchers do. But that is my key message here. I look at value-added not only from my current, official perspective as a university researcher, but as an educator, parent, and adviser to local, state, and national policy makers, as well as former school board member and legislative aide (p. 168).

Other Unchecked Assumptions: Value added will also induce people to work harder (p. 13), attract new hires to the profession (p. 13); lessen the (apparent) "incentive for low-scoring schools to continue the endless cycle of reform in curricula, policies, and personnel" (p. 66); and awaken the traditionally high-scoring schools and teachers resting on their high score laurels (p. 66). In addition, the unintended consequences associated with traditional snapshot measures will diminish if not disappear with the adoption of value-added (p. 67). The side-effects of dishonorable mention include the inordinate focus on moving "bubble kids" above minimum cut-scores (p. 57); the exclusion of disadvantageous students from tests (p. 66); and the attrition of teachers from high-needs schools (p. 66). The lesson here is that research does matter, especially when unchecked assumptions are delivered as unconditional truths.

While no educational accountability system is flawless, nor will such a system ever probably exist, Harris calls for using multiple measures along with value-added so each inadequate measure can offset the other's shortcomings. This is in line with the professions' *Standards for Educational and Psychological Testing* (AERA, 2000, see also p. 167). However, no research yet supports the implied validity of this value-added approach. It has not yet been established that value-added outcomes correlate with these other measures, or convincingly correlate at the levels necessary for validation (Jacob & Lefgren, 2008; Milanowski, Kimball, & White, 2004; Wilson, Hallman, Pecheone, & Moss, 2007; see also p. 137-138). Additional research suggests that if anything, value-added might be trumping and distorting such other independent measures (Capitol Hill Briefing, 2011; Rockoff, Staiger, Kane, & Taylor, 2010).

In the end, Harris's eager endorsement of value-added is negligently dismissive of the very real and practical issues prohibiting everything value-added has to offer. It is too soon to enthusiastically endorse value-added as it is still uncertain whether value-added will ever be everything it was conceptually cut out for, or everything it is already assumed to be. While most agree that value-added is the best of the alternatives, it is simply not yet good enough.

References

- American Educational Research Association (AERA). (2000). *AERA position statement on high-stakes testing in Pre-K - 12 education*. Retrieved from <http://www.aera.net/?id=378>
- Amrein-Beardsley, A. (2008). Methodological concerns about the Education Value-Added Assessment System (EVAAS). *Educational Researcher*, 37(2), 65-75. doi:10.3102/0013189X08316420
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., F Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute. Retrieved from <http://www.epi.org/publications/entry/bp278>
- Capitol Hill Briefing. (2011, September 14). *Getting teacher evaluation right: A challenge for policy makers*. A briefing by E. Haertel, J. Rothstein, A. Amrein-Beardsley, and L. Darling-Hammond. Washington DC: Dirksen Senate Office Building (research in brief). Retrieved from <http://www.aera.net/Default.aspx?id=12856>
- Darling-Hammond, L. (2011, September 7). Too unreliable. *The New York Times*. Retrieved from <http://www.nytimes.com/roomfordebate/2010/09/06/assessing-a-teachers-value/value-added-assessment-is-too-unreliable-to-be-useful>
- Eckert, J. M., & Dabrowski, J. (2010, May). Should value-added measures be used for performance pay? *Phi Delta Kappan*,

91(8), 88-92.

Harris, D. (2011). *Value-added Measures in Education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136. doi:10.1086/522974

Milanowski, A., Kimball, S. M., & White, B. (2004). *The relationship between standards-based teacher evaluation scores and student achievement: Replication and extensions at three sites*. Madison, WI: Center for Education Research, Consortium for Policy Research in Education. Retrieved from http://cpre.wceruw.org/papers/3site_long_TE_SA_AERA04TE.pdf

Ravitch, D. (2010, September, 7). A big margin for error. *The New York Times*. Retrieved from <http://www.nytimes.com/roomfordebate/2010/09/06/assessing-a-teachers-value/assessing-teachers-by-student-scores-is-too-error-ridden-to-be-effective>

Rockoff, J. E., Staiger, D. O., Kane, T. J., & Taylor, E. S. (2010). *Information and employee evaluation: Evidence from a randomized intervention in public schools*. New York, NY: Columbia Business School. Retrieved from <http://www.nber.org/papers/w16240>

Rothstein, J. (2010, September 7). More harm than good. *The New York Times*. Retrieved from <http://www.nytimes.com/roomfordebate/2010/09/06/assessing-a-teachers-value/dont-be-too-quick-to-embrace-value-added-assessments>

Scherrer, J. (2011). Measuring teaching using value-added modeling: The imperfect panacea. *NASSP Bulletin*, 95(2), 122-140. doi:10.1177/0192636511410052

Wilson, M., Hallman, P. J., Pecheone, R., & Moss, P. (2007, October). *Using student achievement test scores as evidence of external validity for indicators of teacher quality: Connecticut's Beginning Educator Support and Training [BEST] Program*. Stanford, CA: Stanford Center for Opportunity Policy in Education. Retrieved from <http://edpolicy.stanford.edu/publications/pubs/107>

Cite This Article as: *Teachers College Record*, Date Published: January 12, 2012
<http://www.tcrecord.org> ID Number: 16648, Date Accessed: 9/1/2012 11:14:40 PM

[Purchase Reprint Rights for this article or review](#)