|  |  |
| --- | --- |
| http://www.tcrecord.org/Images/logo_print.gif  **Chetty, et al. on the American Statistical Association’s Recent Position Statement on Value-Added Models (VAMs): Five Points of Contention**  by [Margarita Pivovarova](http://www.tcrecord.org/AuthorDisplay.asp?aid=22743), [Jennifer Broatch](http://www.tcrecord.org/AuthorDisplay.asp?aid=22744) & [Audrey Amrein-Beardsley](http://www.tcrecord.org/AuthorDisplay.asp?aid=20670) — August 01, 2014  *Over the last decade, teacher evaluation based on value-added models (VAMs) has become central to the public debate over education policy. In this commentary, we critique and deconstruct the arguments proposed by the authors of a highly publicized study that linked teacher value-added to students’ long-run outcomes, Chetty et al. (2014, forthcoming), in their response to the American Statistical Association statement on VAMs. We draw on recent academic literature to support our counter-arguments along main points of contention: causality of VAM estimates, transparency of VAMs, effect of non-random sorting of students on VAM estimates and sensitivity of VAMs to model specification.*  INTRODUCTION  Recently, the authors of a highly publicized and cited study that linked teacher value-added estimates to the long-run outcomes of their students (Chetty, Friedman, & Rockoff, 2011; see also Chetty, et al., in press I, in press II) published a “point-by-point” discussion of the “Statement on Using Value-Added Models for Educational Assessment” released by the American Statistical Association (ASA, 2014). This once again brought the value-added model (VAM) and its use for increased teacher and school accountability to the forefront of heated policy debate.  In this commentary we elaborate on some of the statements made by Chetty, et al. (2014). We position both the ASA’s statement and Chetty, et al.’s (2014) response within the current academic literature. As well, we deconstruct the critiques and assertions advanced by Chetty, et al. (2014) by providing counter-arguments and supporting them by the scholarly research on this topic.  In doing so, we rely on the current research literature that has *really* been done on this subject over the past ten years. This more representative literature was completely overlooked by Chetty, et al. (2014), even though, paradoxically, they criticize the ASA for not citing the “recent” literature appropriately themselves (p. 1). With this being our first point of contention, we also discuss four additional points of dispute within the commentary.  POINT 1: MISSING LITERATURES  In their critique of the ASA statement, posted on a university-sponsored website, Chetty, et al. (2014) marginalize the current literature published in scholarly journals on the issues surrounding VAMs and their uses for measuring teacher effectiveness. Rather, Chetty et al. cite only works representing econometrician’s scholarly pieces, apparently in support of their a priori arguments and ideas. Hence, it is important to make explicit the rather odd and extremely selective literature Chetty, et al. included in the reference section of their critique, on which Chetty, et al. relied "to prove" some of the ASA’s statements incorrect. The whole set of peer-reviewed articles that counter Chetty, et al.’s arguments and ideas are completely left out of their discussion.  A search on the Educational Resources Information Center (ERIC) with “value-added” as key words for the same last five years yields 406 entries, and a similar search in Journal Storage (JSTOR, a shared digital library) returns 495. Chetty, et al., however, only cite 13 references to critique the ASA’s statement, one of which was the actual statement itself, leaving 12 external citations in total and in support of their critique. Of these 12 external citations, three are references to their two forthcoming studies and a replication of these studies’ methods; three have thus far been published in peer-reviewed academic journals, six were written by their colleagues at Harvard University; and 11 were written by teams of scholars with economics professors/econometricians as lead authors.  POINT 2: CORRELATION VERSUS CAUSATION  The second point of contention surrounds whether the users of VAMs should be aware of the fact that VAMs typically measure correlation, not causation. According to the ASA, as pointed out by Chetty, et al. (2014), effects “positive or negative—attributed to a teacher may actually be caused by other factors that are not captured in by the model” (p. 2). This is an important point with major policy implications. Seminal publications on the topic, Rubin, Stuart and Zanutto (2004) and Wainer (2004) who positioned their discussion within the Rubin Causal Model framework (Rubin, 1978; Rosenbaum and Rubin, 1983; Holland, 1986), clearly communicated, and evidenced, that value-added estimates cannot be considered causal unless a set of "heroic assumptions" are agreed to and imposed. Moreover, “anyone familiar with education will realize that this [is]...fairly unrealistic” (Rubin, et al. 2004, p. 108). Instead, Rubin, et al. suggested, given these issues with confounded causation, we should switch gears and evaluate interventions and reward incentives as based on the descriptive qualities of the indicators and estimates derived via VAMs. This point has since gained increased consensus among other scholars conducting research in these areas (Amrein-Beardsley, 2008; Baker, et al., 2010; Betebenner, 2009; Braun, 2008; Briggs & Domingue, 2011; Harris, 2011; Reardon & Raudenbush, 2009; Scherrer, 2011).  POINT 3: THE NON-RANDOM ASSIGNMENT OF STUDENTS INTO CLASSROOMS  The third point of contention pertains to Chetty, et al.’s statement that recent experimental and quasi-experimental studies have already solved the “causation versus correlation” issue. This claim is made despite the substantive research that evidences how the non-random assignment of students constrains VAM users’ capacities to make causal claims.  The authors of the Measures of Effective Teaching (MET) study cited by Chetty, et al. in their critique, clearly state, “we cannot say whether the measures perform as well when comparing the average effectiveness of teachers in different schools…given the obvious difficulties in randomly assigning teachers or students to different schools” (Kane, McCaffrey, Miller & Staiger, 2013, p. 38). VAM estimates were found to be biased for teachers who taught more relatively homogenous sets of students with lower levels of prior achievement, despite the levels of sophistication in the statistical controls used (Hermann, Walsh, Isenberg, & Resch, 2013; see also Ehlert, Koedel, Parsons, & Podgursky, 2014; Guarino et al., 2012).  Researchers repeatedly demonstrated that non-random assignment confounds value-added estimates independent of how many sophisticated controls are added to the model (Corcoran, 2010; Goldhaber, Walch, & Gabele, 2012; Guarino, Maxfield, Reckase, Thompson, & Wooldridge, 2012; Newton, Darling-Hammond, Haertel, & Thomas, 2010; Paufler & Amrein-Beardsley, 2014; Rothstein, 2009, 2010).  Even in experimental settings, it is still not possible to distinguish between the effects of school practice, which is of interest to policy-makers, and the effects of school and home context. There are many factors at the student, classroom, school, home, and neighborhood levels that would confound causal estimates that are beyond researchers’ control. Thus, the four experimental studies cited by Chetty, et al. (2014) do not provide ample evidence to refute the ASA on this point.  POINT 4: ISSUES WITH LARGE-SCALE STANDARDIZED TEST SCORES  In their position statement, ASA authors (2014) rightfully state that the standardized test scores used in VAMs should not be the only outcomes of interest for policy makers and stakeholders. Indeed, current agreement is that test scores might not even be one of the most important outcomes capturing a student’s educated self. Also, if value-added estimates from standardized test scores cannot be interpreted as causal, then the effect of “high value-added” teachers on college attendance, earnings, and reduced teenage birth rates cannot be considered causal either as opposed to what is implied by Chetty, et al. (2011; see also Chetty, et al., in press I, in press II).  Ironically, Chetty, et al. (2014) cite Jackson’s (2013) study to confirm their point that high value-added teachers also improve long-run outcomes of their students. Jackson (2013), however, actually found that teachers who are good at boosting test scores are not always the same teachers who have positive and long-lasting outcomes on non-cognitive skills acquisition. Moreover, value-added as related to test scores and non-cognitive outcomes for the same teachers were then, and have since been shown to be, weakly correlated with one another.  POINT 5: MODEL SPECIFICITY  Lastly, ASA (2014) expressed concerns about the sensitivity of value-added estimates to model specifications. Recently, researchers have found that value-added estimates are highly sensitive to the tests being used, even within the same subject areas (Papay, 2011) and the different subject areas taught by the same teachers given different student compositions (Loeb & Candelaria, 2012; Newton, et al., 2010; Rothstein, 2009, 2010). While Chetty, et al. rightfully noted that different VAMs typically yield correlations around *r* = 0.9, this is typical with most “garbage in, garbage out” models. These models are too often used, too often without question, to process questionable input and produce questionable output (Banchero & Kesmodel, 2011; Gabriel & Lester, 2012, 2013; Harris, 2011).  What Chetty, et al. overlooked, though, are the repeatedly demonstrated weak correlations between value-added estimates and other indicators of teacher quality, on average between *r =* 0.3 and 0.5 (see also Corcoran, 2010, Goldhaber et al., 2012; McCaffrey, Sass, Lockwood, & Mihaly, 2009; Broatch and Lohr, 2012; Mihaly, McCaffrey, Staiger, & Lockwood, 2013).  CONCLUSION  In sum, these are only a few “points” from this “point-by-point discussion” that would strike anyone even fairly familiar with the debate over the use and abuse of VAMs. These “points” are especially striking given the impact Chetty, et al.’s original (2011) study and now forthcoming studies (Chetty, et al., in press I, in press II) have already had on actual policy and the policy debates surrounding VAMs. Chetty, et al.’s (2014) discussion of the ASA statement, however, should cause others pause in terms of whether in fact Chetty, et al. are indeed experts in the field, or not. What certainly has become evident is that they do not have their minds wrapped around the extensive set of literature or knowledge on this topic. If they had, they may not have come off as so selective, as well as biased, citing only those representing certain disciplines and certain studies to support certain assumptions and “facts” upon which their criticisms of the ASA statement were based.  *References*  American Statistical Association. (2014). *ASA Statement on using value-added models for educational assessment.* Retrieved fromhttp://www.amstat.org/policy/pdfs/ASA\_VAM\_Statement.pdf  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., & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, D.C.: Economic Policy Institute. Retrieved from http://www.epi.org/publications/entry/bp278  Banchero, S. & Kesmodel, D. (2011, September 13).  Teachers are put to the test: More states tie tenure, bonuses to new formulas for measuring test scores. *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB10001424053111903895904576544523666669018.html  Betebenner, D. W. (2009b). Norm- and criterion-referenced student growth. *Education Measurement: Issues and Practice, 28*(4), 42-51. doi:10.1111/j.1745-3992.2009.00161.x  Braun, H. I. (2008). Viccissitudes of the validators. Presentation made at the 2008 Reidy Interactive Lecture Series, Portsmouth, NH. Retrieved from www.cde.state.co.us/cdedocs/OPP/HenryBraunLectureReidy2008.ppt  Briggs, D. & Domingue, B. (2011, February). 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 from nepc.colorado.edu/publication/due-diligence  Broatch, J. and Lohr, S. (2012) “Multidimensional Assessment of Value Added by Teachers to Real-World Outcomes”, *Journal of Educational and Behavioral Statistics*, April 2012; vol. 37, 2: pp. 256–277.  Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood.* Cambridge, MA: National Bureau of Economic Research (NBER), Working Paper No. 17699. Retrieved from http://www.nber.org/papers/w17699  Chetty, R., Friedman, J. N., & Rockoff, J. (2014). *Discussion of the American Statistical Association’s Statement (2014) on using value-added models for educational assessment.* Retrieved from http://obs.rc.fas.harvard.edu/chetty/ASA\_discussion.pdf  Chetty, R., Friedman, J. N., & Rockoff, J. E. (in press I). Measuring the impact of teachers I: Teacher value-added and student outcomes in adulthood. *American Economic Review*.  Chetty, R., Friedman, J. N., & Rockoff, J. E. (in press II). Measuring the impact of teachers II: Evaluating bias in teacher value-added estimates. *American Economic Review*.  Corcoran, S. (2010). *Can teachers be evaluated by their students’ test scores? Should they be? The use of value added measures of teacher effectiveness in policy and practice.* Educational Policy for Action Series. Retrieved from: http://files.eric.ed.gov/fulltext/ED522163.pdf  Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. J. (2014). The sensitivity of value-added estimates to specification adjustments: Evidence from school- and teacher-level models in Missouri. *Statistics and Public Policy. 1*(1), 19–27.  Gabriel, R., & Lester, J. (2012). Constructions of value-added measurement and teacher effectiveness in the *Los Angeles Times:* A discourse analysis of the talk of surrounding measures of teacher effectiveness. Paper presented at the Annual Conference of the American Educational Research Association (AERA), Vancouver, Canada.  Gabriel, R. & Lester, J. N. (2013). Sentinels guarding the grail: Value-added measurement and the quest for education reform. *Education Policy Analysis Archives, 21*(9), 1–30. Retrieved from http://epaa.asu.edu/ojs/article/view/1165  Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica,* 80, 589–612.  Goldhaber, D., Walch, J., & Gabele, B. (2012). Does the model matter? Exploring the relationships between different student achievement-based teacher assessments. S*tatistics and Public Policy, 1*(1), 28–39.  Guarino, C. M., Maxfield, M., Reckase, M. D., Thompson, P., & Wooldridge, J.M. (2012, March 1). *An evaluation of Empirical Bayes’ estimation of value-added teacher performance measures*. East Lansing, MI: Education Policy Center at Michigan State University. Retrieved from http://www.aefpweb.org/sites/default/files/webform/empirical\_bayes\_20120301\_AEFP.pdf  Harris, D. N. (2011). *Value-added measures in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.  Hermann, M., Walsh, E., Isenberg, E., & Resch, A. (2013). *Shrinkage of value-added estimates and characteristics of students with hard-to-predict achievement levels.* Princeton, NJ: Mathematica Policy Research. Retrieved form http://www.mathematica-mpr.com/publications/PDFs/education/value-added\_shrinkage\_wp.pdf  Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945–960.  Jackson, K. C. (2012). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina.Cambridge, MA: National Bureau of Economic Research (NBER), Working Paper No. 18624. Retrieved from http://www.nber.org/papers/w18624  Kane, T., McCaffrey, D., Miller, T. & Staiger, D. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. Bill and Melinda Gates Foundation. Retrieved from http://www.metproject.org/downloads/MET\_Validating\_Using\_Random\_Assignment\_Research\_Paper.pdf  Loeb, S., & Candelaria, C. (2013). How stable are value-added estimates across  years, subjects and student groups? Carnegie Knowledge Network. Retrieved from http://carnegieknowledgenetwork.org/briefs/value‐added/value‐added‐stability  McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy, 4*, 572–606.  Mihaly, K., McCaffrey, D., Staiger, D. O., & Lockwood, J.R. (2013). A  composite estimator of effective teaching. Seattle, WA: Bill and Melinda Gates Foundation. Retrieved from: http://www.metproject.org/downloads/MET\_Composite\_Estimator\_of\_Effective\_Teaching\_Research\_Paper.pdf  Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives, 18*(23). Retrieved from: epaa.asu.edu/ojs/article/view/810.  Papay, J. P. (2010). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal, 48*(1), 163–193.  Paufler, N. A., & Amrein-Beardsley, A. (2014). The random assignment of students into elementary classrooms: Implications for value-added analyses and interpretations. *American Educational Research Journal.*  Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy, 4*(4), 492–519. doi:10.1162/edfp.2009.4.4.492  Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 17*, 41–55.  Rothstein, J. (2009). *Student sorting and bias in value-added estimation: Selection on observables and unobservables*. *Education Finance and Policy, (4)*4, 537–571. doi:http://dx.doi.org/10.1162/edfp.2009.4.4.537  Rothstein, J. (2010, February). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*. 175–214. doi:10.1162/qjec.2010.125.1.175  Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics, 6*, 34–58  Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics, 29*(1), 103–116.  Scherrer, J. (2011). Measuring teaching using value-added modeling: The imperfect panacea. *NASSP Bulletin, 95*(2), 122–140. doi:10.1177/0192636511410052  Wainer, H. (2004). Introduction to a special issue of the *Journal of Educational and Behavioral Statistics* on value-added assessment. *Journal of Educational and Behavioral Statistics, 29*(1), 1–3. doi:10.3102/10769986029001001   |  | | --- | | **Cite This Article as:** *Teachers College Record*, Date Published: August 01, 2014 [http://www.tcrecord.org](http://www.tcrecord.org/Home.asp) ID Number: 17633, Date Accessed: 8/7/2014 1:03:08 PM  [Purchase Reprint Rights for this article or review](https://www.tcrecord.org/Permissions.asp?ContentId=17633) | |