

Finding from the research on using assessment data for teacher evaluation
Brian A. McNulty Ph.D.

- **There are 2 critical findings:**
 - **Teacher behavior is not stable across lessons or years**
 - **Teacher effects on achievement are not stable across lessons or years. (Berliner 2014)**
- **The key question is whether assessment measures “can accurately identify individual teachers’ contributions to student learning and offer a credible measure of teacher ‘effectiveness.’”**

“For this to occur 3 conditions need to be in place:

- **The right assessments** – Individual students’ actual achievement is accurately measured by tests that reflect valuable learning on a vertical scale with equal-interval units.
- **Random student assignment** – Classroom conditions and group traits don’t vary substantially from one classroom to another.
- **Ability to tease out the value each educator has added** – Teachers’ individual contributions can be isolated from other factors that affect student learning.

None of these assumptions holds true, in fact, “the violations of these assumptions are considerable.”

- **Assessments- standardized tests focus mainly on basic skills and, because they are grade-specific, don’t accurately measure the achievement of students working above or below grade level.**
- **This means that these are particularly inaccurate for teachers working with low- and high-achieving students**
- **Random student assignment- Teachers working in low-income communities have far more students with educational, psychological, health, and social needs**

“It is clear that the assumption of equivalence among classrooms is far from reality.” (Darling-Hammond (2015)

- **Additional measurement issues pertain especially to floor and ceiling effects. Teachers who work with students who are out of the range of the test’s precision will find there are few or no items on the test to measure growth. (Kersting, et al. 2013)**
- **The only way to know if measures of teaching truly identify effective teaching and not some unmeasured student characteristics is by randomly assigning teachers to students. (Gates Foundation, 2013)**

- “Value-Added measures are generally based on standardized test scores, and do not directly measure potential teacher contributions toward student outcomes.” (American Statistical Association 2014)
- **When three years of rankings are used, reliability is even worse:**
 - 74% of teachers ranked in the bottom quintile shift out of that quintile
 - 56% of teachers ranked in the top quintile shift out of that quintile. (Yeh, 2013).
- The collective set of stable evidence should convince policymakers that the assumptions imbedded in Race to the Top about **the easy use of observational and achievement data to evaluate teaching are faulty.** (Good, & Lavigne, 2014).
- Because of this (variability), you simply cannot ethically and legally use these data for high-stakes decisions about teacher competence. (Berliner 2014)

Table 2. Instability of Value-Added Teacher Rankings in San Diego and 5 Florida Counties

Locale	Teacher Rankings	
	Bottom 20% in Year t; Top 80% in Year t+1	Top 20% in Year t; Bottom 80% in Year t+1
San Diego, CA	65%	71%
Dade County, FL	70%	67%
Duval County, FL	67%	61%
Hillsborough County, FL	67%	67%
Orange County, FL	59%	65%
Palm Beach County, FL	69%	68%

Yeh (2013)

Teacher Evaluation and Student Assessment References

- American Statistical Association. (2014). ASA Statement on Using Value-Added Models for Educational Assessment. Retrieved from https://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf
- Amrein-Beardsley, A., & Collins, C. (2012). The SAS Education Value Added Assessment System (SAS[®] EVAAS[®]) in the Houston Independent School District (HISD): Intended and Unintended Consequences. *Education Policy Analysis Archives*, 20 (12). Retrieved [date], from <http://epaa.asu.edu/ojs/article/view/1096>
- Amrein-Beardsley, A., (2014). Rethinking Value-Added Models in Education: Critical Perspectives on Tests and Assessment-Based Accountability Routledge. NY.NY
- Baker, B.D., Oluwole, J., Green, P.C. III (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(5).
- Berliner, D., 2014. Exogenous Variables and Value-Added Assessments: A Fatal Flaw. *Teachers College Record* Volume 116, 010307, January 2014, 31 pages
- Darling-Hammond, L. (2015). Can Value Added Add Value to Teacher Evaluation? *Educational Researcher*. March 2015 vol. 44 no. 2 132-137
- Darling-Hammond, L., & Snyder, J. (2015b). Meaningful Learning in a New Paradigm for Educational Accountability: An Introduction. *Education Policy Analysis Archives Vol. 23 No. 7*
- Gates Foundation (2013). Feedback for Better Teaching: Nine Principles for Using Measures of Effective Teaching. Bill and Melinda Gates Foundation.
- Gates Foundation (2013). Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study. Bill and Melinda Gates Foundation
- Good, T. L., & Lavigne, A. L. (2014). Issues of teacher performance stability are not new: Limitations and possibilities. *Education Policy Analysis Archives*, 23(2). <http://dx.doi.org/10.14507/epaa.v23.1916>
- Holloway, J., & Amrein-Beardsley, A. (2015). "Truths" Devoid of Empirical Proof: Underlying Assumptions Surrounding Value-Added Models in Teacher Evaluation. *Teachers College Record*, Date Published: June 29, 2015. <http://www.tcrecord.org> ID Number: 18008, Date Accessed: 7/8/2015 5:48:27 PM
- Kersting, N. B., Chen, M., & Stigler, J. W. (2013). Value-added Teacher Estimates as Part of Teacher Evaluations: Exploring the Effects of Data and Model Specifications on the Stability of Teacher Value-added Scores. *Education Policy Analysis Archives*,

21(7). <http://dx.doi.org/10.14507/epaa.v21n7.2013>

Konstantopoullos, S. (2014). Teacher Effects, Value-Added Models, and Accountability. *Teachers College Record* Volume 116, 010304, January 2014, 21 pages

Shepard (2013). *Teachers College Record* Volume 115 Number 9, 2013, p. 1-12
<http://www.tcrecord.org> ID Number: 17116

Welner, K., G. (2013). Consequential Validity and the Transformation of Tests from Measurement Tools to Policy Tools. *Teachers College Record* Volume 115 Number 9, 2013, p. 1-6
<http://www.tcrecord.org> ID Number: 17115, Date Accessed: 5/9/2015 5:07:28 PM

Yeh, S.S., (2013). A Reanalysis of the Effects of Teacher Replacement Using Value-Added Modeling. *Teachers College Record* Volume 115, 120305, December 2013.

Dr. Brian McNulty is a Founding Partner with Creative Leadership Solutions

Prior to this he was the Vice President for Leadership Development at the Leadership and Learning Center. Before that he served as the Vice President for Field Services at the Mid-continent Research for Education and Learning (McREL). Before coming to McREL, he was an Assistant Superintendent for Adams County School District 14, and the Assistant Commissioner of Education, for the Colorado Department of Education.

Brian's work and writing have been featured in books, scholarly journals and periodicals throughout the world. An author of more than 40 publications, Brian's most recent books include, *Leaders Make It Happen (2011)* with Laura Besser (an AASA member book) and *School Leadership that Works: from Research to Results*, an ASCD best selling publication co-authored with Robert Marzano and Tim Waters. His most recent chapter *Students with Disabilities: Moving Your Numbers (2014)* with Deborah Telfer in *Navigating Achievement for Struggling Students with the Common Core State Standards*. (eds.) Englewood CO: Lean + Learn Press.