Value of Value-Added Models Based on Student Outcomes to Evaluate Teaching

Ronald A. Berk*

WARNING: *In order to satisfy FDA regulations for all articles produced in nature, you are forewarned that different sections of this article can contain tree nuts, soy, hormones, antibiotics, dairy, gluten, pesticides, preservatives, extra lard and sodium, high glycemic, MSG, GMO, MIC, KEY, and MOUSE. All ingredients are unnatural and processed in a laboratory under duress and false pretenses. Document should be read online to minimize contamination and allergic reactions. Enjoy!*

Prologue

From the title above, you're probably expecting a spectacularly explosive treatment of the topic with bare-knuckle, mano-a-mano fights, motorcycle and car crashes, SWAT teams smashing through doors, and acrobatic high-speed chases on the Las Vegas Strip. Wait! That sounds more like a day in the life of Jason Bourne. That's not exactly journal fare. What were you thinking?

Bubbling Student Outcomes

Instead, this article will bludgeon you with pointed statistical issues and psychometric standards introduced in my JFD article two years ago, give or take a day or two (Berk, 2014). That prequel dealt with whether student outcomes should be used as one among 15 possible sources of evidence in formative (teaching and course improvement), summative (annual review, contract renewal, promotion & tenure, teaching awards), and program evaluation decisions (accreditation and accountability). I forgot the title. It was something like "Should Student Outcomes Be Used to Evaluate Teaching?"

Recently, student outcomes have bubbled to the top of debates about how to evaluate teaching in community and liberal arts colleges, universities, and professional schools, but even more international attention has been riveted on how outcomes are being used to evaluate teachers and administrators K–12 (Harris, 2012; Rowen & Raudenbush, 2016; Sawchuk, 2014). It remains a bubbling cauldron of controversy. Since 2014, the bubbling continued with some clarification of the issues, numerous research studies in the public sector, and several significant publications and policy statements that can guide current practices.

The Wonder Years

If you have wondered what has happened over the past few years, or maybe not, this article is intended to update you on those "wonder years." It will not repeat information from the prequel because that would upset my editor and waste your time. This one will be more narrowly focused on the technical options available. It will also unspool the requirements that an institution can use to measure student outcomes and extract the teacher's contribution to those outcomes as one source of evidence to evaluate teaching and the instructional program. In other words, this article is the sequel to the prequel. (*QUESTION:* Don't they sound more like cold remedies than articles? Take 2 tablespoons of Sequel® every 6 hours.)

[*SUGGESTION:* I recommend you peruse the prequel to gain a perspective on the policy, psychometric, and legal issues involved with student outcomes. As a dedicated educator, it's what you do. If you want to save 15% or more on car insurance, switch to GEICO®. It's what you do. We now resume this prologue already in progress.]

The prequel ended with recommendations for using student outcomes in formative, summative, and

*The author is extremely grateful to *John R. Lockwood* (Principal Research Scientist) of Educational Testing Service and *Steve Benton* (Senior Research Officer) and *Jason Barr* (Researcher) of The IDEA Center for reviewing the accuracy and appropriateness of the content, thoughtful technical comments, and laughing at a few of the jokes on an earlier draft of this article. None of the reviewers should be held responsible for the recommendations and opinions expressed herein or any mistakes in the text. There is only one person who should be held accountable, and that person, of course, is: "The Donald."

program decisions. They were not exactly what proponents of student outcomes wanted to read. Then why open up old wounds and start the bleeding again? *ANSWER: The employment decisions being made about faculty based on inappropriate, unreliable, and/or invalid data can ruin their careers.* These high-stakes decisions are still being made at hundreds of institutions internationally. The following sections are designed to cauterize the sources of that bleeding.

The purposes of this article are to synthesize what we've learned in order to furnish guidelines for practice. It covers the following contentious territory: (1) push for student outcomes, (2) pushback against student outcomes, (3) definition of value-added models (VAMs), (4) validity and reliability of VAM scores, (5) standards for using VAM, and (6) conclusions and recommendations. So what is the…

Push for Student Outcomes

"Race to the Top" in 42 States

Although the use of student test scores to measure teaching performance dates back more than a century (Reese, 2013) to *Jurassic World* with its star "Indominus Rex," the most recent push came from the K–12 sector. The federal "Race to the Top" (RTTT) competition stipulated that states and school districts applying for grants must incorporate measures of student achievement into teacher evaluations and, "at a minimum," use those evaluations to make "high stakes" summative decisions about compensating, promoting and retaining teachers (U.S. Department of Education [USDOE], 2009, 2010). At the beginning of 2016, student outcomes were embedded in law or regulations in 42 states and the District of Columbia (Sawchuk, 2016).

USDOE Requirements

The USDOE federal review required districts: (1) to measure student growth, (2) to design and implement evaluation systems that include multiple rating categories that take into account data on student growth, (3) to evaluate teachers and principals annually and provide feedback, including student growth data, and (4) to use these evaluations to inform decisions about professional development, compensation, promotion, retention, tenure, and certification (U.S. Department of Education, 2009). It is significant that these requirements specify *growth* of student outcomes, captured by standardized achievement tests, instead of *status* which indicates performance at one point in time with a threshold or cut score. The former is regarded as superior to the latter because of the greater information it provides.

Institutional Accountability and Accreditation

In higher education, there are mandates for student outcome data by state legislatures and accreditation review boards. Such data are so tantalizing and intuitively appealing for accountability and accreditation that student outcomes are being adopted or given serious attention at hundreds of institutions. More than 1500 administer the *National Survey of Student Engagement*, while others use the *Student Assessment of Their Learning Gains*, *Transparency in Learning and Teaching Survey*, and *ETS Proficiency Profile* (see Berk, 2014, for details). Student outcomes have even been recommended in place of student ratings to evaluate teaching (Weinberg, Fleisher, & Hashimoto, 2007). That takes us smoothly into the …

Pushback against Student Outcomes

Policy Makers & Educators

There is considerable disagreement, skepticism, and conflict among policy makers and educators, especially statisticians, about whether the research evidence on value-added models (VAMs) that measure student growth is sufficient to justify their systematic use for high-stakes decisions to evaluate teaching (American Statistical Association, 2014; Amrein-Beardsley, 2014; Ballou & Springer, 2015; Chetty, Friedman, & Rockoff, 2014; Darling-Hammond, 2015; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Glazerman et al., 2010; Guarino, Rechase, & Wooldridge, 2015; Haertel, 2013; Harris & Herrington, 2015; Hill, Kapitula, & Umland, 2011; Johnson, 2015; National Research Council, 2010; Raudenbush, 2013; Ravitch, 2009; Rothstein, 2010). The primary target of resistance has been the use of VAMs for summative decisions and the consequences of those decisions on teachers and students rather than for formative and program decisions.

AFT, NEA, & Teacher Resistance

In response to the federal requirements of RTTT, American Federation of Teachers President Randi Weingarten (Weingarten, 2014) called for an end to using VAM as a component of teacher evaluations (Sawchuk, 2014). The National Education Association (Walker, 2014) questioned the relationships between teaching quality and VAMs. Several studies have indicated that teachers distrust VAM in favor of classroom observations (Eger, 2014; Goldring et al., 2015; Jiang, Sporte, & Luppescu, 2015). This resistance also prompted more than a dozen lawsuits (Sawchuk, 2016). The problems associated with the technical issues and statistical inferences about teacher performance from student performance K–College have taken on increasingly intense interest.

Impact of the Pushback

As educators at all levels were being drawn into the vortex of this pushback and the stakes increased, the support for VAM continued to fall (Harris & Herrington, 2015). The cumulative impact of the resistance from so many credible sources contributed to the passage of the Every Student Succeeds Act (ESSA) at the end of 2015 (U.S. Department of Education, 2015), which stated that *student outcomes are no longer required in the evaluation of teachers* (Klein, 2016). Outcome measures are not a necessary ingredient to receive federal funds in the design and implementation of a teacher evaluation system. ESSA now allows states to decide whether and how to evaluate teachers with or without test scores (Amrein-Beardsley, 2015).

With 42 states already committed to student outcomes tethered to their evaluation systems (National Council on Teacher Quality, 2015), what will happen next? It is totally at the states' discretion. With the previous federal mandate lifted, ESSA will affect the use of VAMs in K–12 (Amrein-Beardsley, 2016) and, probably, in higher education.

Given that status report, is VAM redeemable in those applications? What does the most recent research reveal about what VAM can and cannot do to improve our knowledge about student growth and its use for personnel decisions about teachers? Let's examine the definition of VAM and then the evidence on the validity and reliability of VAM scores. Is there a comprehensible…

Definition of Value-Added Models (VAMs)

Just how do you cleanly extract the teacher's contributions to his or her students' outcomes? That's the *raison d'être* (a Greek phrase meaning literally, "there's a raisin in your nose") for this article. *Value-added models* (*VAMs*) are intended to predict the "*value a teacher adds to student achievement growth.*" In other words, how much do teachers influence their students' outcome performance beyond the factors that are not within their control?

VAM is the rubric for a variety of true value-added and student growth models. It does not provide information on why students grew or how they could improve their growth. [*NOTE:* VAM is actually a family of statistical models (Braun, 2015). I bet you thought that statistics don't have families with children, grandchildren, mothers-in-law, and pets. They do. They just keep their personal lives private and don't share on Facebook. You can find them on LinkedIn.]

The technical definition of *VAM* in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME Joint Committee, 2014) is as follows:

> Estimating the contribution of individual schools or teachers to student performance by means of complex statistical techniques [that would make nonstatisticians throw up] that use multiple years of student outcome data, which typically are standardized test scores. (p. 225)

I hope that clears things up. Kidding. Based on the most recent research, there are several VAMs that aggregate students' achievement growth to measure pre-post change: (1) gain-score based models which aggregate pre-post difference scores (Berk, 2014), (2) transition-based models which compute aggregate changes in performance over two or more years (Braun, 2005), (3) student growth percentiles-based models which compute the mean or median percentiles aggregated across students (Betebenner, 2008, 2011), and (4) value-added measures-based models which establish current test scores based on previous test scores along with demographic characteristics to isolate the teacher's impact (Harris, 2011, 2013b). The last-named has received the most attention in the literature; but it is also the most difficult to apply in practice. VAM was briefly described in the prequel. Now it's time to examine the…

Validity and Reliability of VAM Scores

Extraneous Factors Contaminating VAM Scores

The validity depends on how well the particular VAM adjusts for other factors that might systematically affect or bias the teacher's VAM score. Those factors create a statistical quagmire. The factors that vary across classrooms in each application are the most problematic. If they are not controlled, they will be systematic errors in the VAM score. Those that do not vary are not sources of bias; they are random errors that would average out.

Our hero identified more than 50 potential extraneous factors in K–12 (Berk, 1988, 1990) and 40 at the college level (Berk, 2014) related to characteristics of students, courses, and outcome measures. Darling-Hammond (2015, p. 133) compiled a list of more than 20 factors related to the school, prior teachers, peer culture and achievement, differential summer learning gains and losses, the home environment, and individual student needs, health, and attendance.

Teacher's Contribution to VAM Score

The key question is: "When is this article going to end? BOOOORING!" Wait! There's another question: With all of the factors that can affect student performance, can we pinpoint how much the teacher contributes? Yes. There are a few recent estimates based on

research in grade levels K–12. The American Statistical Association (2014) reported that "Most VAM studies find that *teachers account for about 1% to 14% of the variability in test scores*" (p. 2). YIIIKES! Can you believe that? Me neither. That means that the majority of variation in scores is attributable to the numerous factors outside of the teacher's control.

Ironically, although the teacher contributes a smidgen of the variance, the *VAM score typically is weighted 30%–50%* in his or her evaluation, with equal or greater weight given to classroom observations (Goldhaber, 2015; Harris, 2014; Johnson, 2015). Those weights attached to a single score are disproportionately too high compared to the teacher's contribution and the decisions being made (American Statistical Association, 2014).

Technical Issues

But the statistical plot thickens when we consider the following characteristics that affect VAM scores:

1. There are large standard errors (aka low reliability) associated with VAM scores (Ballou & Springer, 2015; Harris, 2011; Harris & Herrington, 2015; Raudenbush & Jean, 2012).
2. Different achievement tests yield different VAM scores (Lockwood et al., 2007; Papay, 2011; Rothstein, 2011).
3. Different schools, courses, classes, and sections yield different VAM scores Goldhaber & Hansen, 2013; Haertel, 2013; Kane, McCaffrey, Miller, & Staiger, 2013; Loeb, 2013; Raudenbush, 2014a).
4. Different time increments yield different longitudinal VAM scores (Goldhaber, 2015; Goldhaber & Hansen, 2013; Kane et al., 2013; Loeb, 2013).
5. Different statistical models yield different VAM scores for up to 50% of the teachers (Briggs & Domingue, 2011; Goldhaber, Gabele, & Walch, 2013; Newton, Darling-Hammond, Haertel, & Thomas, 2010; Sass, Semykina, & Harris, 2014).
6. VAMs rarely adjust for more than a few nonteacher factors that could bias the score, e.g., prior achievement and demographic characteristics of students (Ballou, Mokher, & Cavalluzzo, 2012; Ballou, Sanders, & Wright, 2004; Goldhaber, 2015; Johnson, 2015).
7. There are unintended consequences of using VAMs to make personnel decisions about teachers, e.g., pernicious teacher practices, such as teaching to the test, and undermining collaboration among teachers (American Statistical Association, 2014; Johnson, 2015; Raudenbush, 2014b; Raudenbush & Jean, 2012).

To what extent do these 48 technical issues (*WAKE-UP CALL:* Just checking if you're still awake.), I mean seven issues generalize to the college classroom and decisions about faculty? The statistical issues and psychometric limitations are virtually identical. The actual effects could be different. There is greater variability across K–12 students and teachers and the factors that can affect their performance than among college students and faculty.

However, when VAM is applied to either K–12 or higher education, there are two indisputable technical conclusions: (1) *VAM scores are inaccurate, biased, and unstable due to several intractable problems,* and (2) the *"teacher effect" usually inferred from the scores is actually more of a "nonteacher effect."* High-stakes decisions will result in errors in teacher performance classification and may inappropriately focus on the value-added measure to the exclusion of others (Goldhaber, 2015; Klees, 2016; Raudenbush & Jean, 2012).

Should You Use VAM?

With all of these weaknesses of VAMs and the cautions about their application to evaluate teaching, how do you navigate through this maelstrom of confusion? Should you use VAM? The answer is teetering precariously between "no" and "under certain conditions." There is consensus that *VAM scores should not be used alone to make any high-stakes decisions about teachers*. If you want to travel the bumpy, serpentine road of student outcomes, there are professional standards and requirements you can follow to proceed with VAM scores as one source of evidence in combination with other sources to evaluate teachers. Let's review the…

Standards for Using VAM

There are three earth-shattering, mind-numbing, Pulitzer-prize potential publications that provide a thick layer of understanding, caution, and guidelines on the use of VAM beyond that already conveyed in the prequel: (1) the most recent revision of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME Joint Committee, 2014), (2) an executive summary policy statement by the American Statistical Association (2014), and (3) a policy statement by the American Educational Research Association Council (2015) on the scientific and technical requirements for using VAM to evaluate educators. I read them so you don't have to. A *CliffsNotes*® version will be presented here.

These documents synthesized the research, statistical and psychometric issues, and policy arguments into standards and statements to guide practices in K–College applications. They lend *gravitas* (*DERIVATION:* a French word derived from "*grav,*" meaning "order,"

and "*itas*," meaning "the zucchini.") to the significance of VAM in educational practice. In fact, most of the technical problems and standards described in Berk (2014) have been codified into formal specific requirements by two national professional associations: American Statistical Association and American Educational Research Association.

Standards for Educational and Psychological Testing

In the seventh edition of the *Standards* (AERA, APA, & NCME Joint Committee, 2014), the critical standard for the use of a test for high-stakes employment decisions (Standard 11.3) still specifies that *"a close link be demonstrated between test content and the job or professional/occupational requirements"* (p. 178). Student outcome measures do not satisfy that standard or those in the *Personnel Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 2009).

There is also a new standard explicitly addressing the use of VAM (Standard 13.2). In the context of evaluation and accountability, it states that statistical models, such as VAMs, should be described and justified with evidence of their appropriateness for the inference that a teacher's effectiveness improves student achievement. Decisions regarding what variables to include in such models should be informed by empirical evidence regarding the effects of their inclusion or exclusion (p. 213). This standard specifies the technical requirements, use, and reporting of VAM and any scores or indices that result.

American Statistical Association (ASA) Statement

The American Statistical Association (2014) issued a set of recommendations for the use of VAMs:

1. The ASA endorses wise use of data, statistical models, and designed experiments for improving the quality of education.
2. VAMs are complex statistical models, and high-level statistical expertise is needed to develop the models and interpret their results.
3. Estimates from VAMs should always be accompanied by measures of precision and a discussion of the assumptions and possible limitations of the model. These limitations are particularly relevant if VAMs are used for high-stakes purposes.
    a. VAMs are generally based on standardized test scores, and do not directly measure potential teacher contributions toward other student outcomes.
    b. VAMs typically measure correlation, not causation: Effects—positive or negative—attributed to a teacher may actually be caused by other factors that are not captured in the model.
    c. Under some conditions, VAM scores and rankings can change substantially when a different model or test is used, and a thorough analysis should be undertaken to evaluate the sensitivity of estimates to different models.
4. VAMs should be viewed within the context of quality improvement, which distinguishes aspects of quality that can be attributed to the system from those that can be attributed to individual teachers, teacher preparation programs, or schools. The majority of opportunities for quality improvement are found in the system-level conditions. Ranking teachers by their VAM scores can have unintended consequences that reduce quality.

American Education Research Association (AERA) Council Statement

Following the ASA Statement (2014) and the revision of the *Standards* (AERA, APA, & NCME Joint Committee, 2014), AERA (2015, pp. 449–451) published their own statement of eight technical requirements that must be met for the use of VAM.

VAM scores:

1. must only be derived from students' scores on assessments that meet professional standards of reliability and validity for the purpose to be served;
2. must be accompanied by separate lines of evidence of reliability and validity that support each claim and interpretative argument;
3. must be based on multiple years of data from sufficient numbers of students;
4. must only be calculated from scores on tests that are comparable over time;
5. must not be calculated in years or for subjects where there are not standardized assessments that are accompanied by evidence of their reliability and validity;
6. must never be used alone or in isolation in educator or program evaluation systems;
7. and the evaluation systems in which they are included must have ongoing monitoring for technical quality and validity of use;
8. and accompanying evaluation reports and determinations based on VAM must include statistical estimates of error associated with student growth measures and any ratings or measures derived from them.

Conclusions and Recommendations

What have we learned? Are value-added models a flash in the VAM? Maybe we're being VAMboozled (Amrein-Beardsley, 2016). After unspooling the story of VAM, there can't be a happy ending. It's like the end-of-season finales of TV drama series. Of course, those seasons are only three weeks long now so the shocking final episode isn't really that shocking. VAM's season is a little longer between the pre- and posttesting to estimate gain.

Conclusions

The conclusions from the prequel still hold for the sequel. What's changed is the track record of VAM applications and research and the professional recommendations for executing them. Even statistically rigorous uses of VAM can't control for (1) standardized achievement tests in college courses that are not sensitive to course content and teaching methods, (2) no available standardized tests for most courses to compute VAM scores, and (3) intractable student, course, and measurement factors that affect achievement (Berk, 2014; Klees, 2016).

Ironically, some of the most serious problems are not attributable to the statistical model per se, but, instead, to the quality of the data input and the data not input. This is the equivalent of the "whodunit" in a murder mystery. Here it's the data that committed the crime; not the statistic. There are both crimes of commission and omission that ultimately weaken the validity and reliability of the VAM scores or render them impossible to estimate.

Top 10 Recommendations

Based on what we know now about VAMs, how should you proceed if your institution wants to use student outcomes to evaluate faculty and your instructional program? Here are my top 10 recommendations:

1. Specify your decisions (formative, summative, and program) and then pick the best sources of evidence for those decisions (Berk, 2006, 2013b).
2. DO NOT use just one source for any decisions—all sources are fallible in different ways, so any single source may yield inaccurate and biased information.
3. Use multiple sources among the 15 strategies previously identified, including student ratings, classroom observation, self-ratings, and others, to render personnel decisions about faculty (Braun, 2015; Darling-Hammond, 2013, 2015; Grissom & Youngs, in press; Harris, 2013a, 2013b, 2014; Goldring et al., 2015; Lyde, Grieshaber, & Byrns, 2016; Measures of Effective Teaching, 2012, 2013; van der Lans, van de Grift, & van Veen, 2015).
4. Triangulate VAM scores with the ratings from other sources to provide a more complete and accurate evidence-based profile of teaching performance than any one source (Berk, 2005, 2006, 2009, 2013a, 2013b).
5. Use only measures (tests, scales, questionnaires) that meet technical and legal standards for appropriate levels of reliability and validity.
6. Standardized tests, such as *ETS Proficiency Profile* and professional licensure and certification tests, used as achievement outcome measures are unrelated to the curriculum and instruction in any single college course (Berk, 2014); consequently, they may be applicable for program decisions only.
7. VAM scores cannot be estimated in courses for which there are no appropriate standardized tests available.
8. Follow the specific Standards (AERA, APA, & NCME Joint Committee, 2014) and ASA (2014) and AERA Council (2015) requirements for designing and executing VAMs and computing and interpreting their scores.
9. Student outcomes (teacher-made tests, standardized tests, perceived learning measures) using VAM scores or other growth measures may be used with *CAUTION* by faculty for formative decisions, faculty developers for professional development programs, and administrators for program evaluation and accreditation *in conjunction with other sources*.
10. Student outcomes should NOT be used for high-stakes employment decisions about part-time and full-time faculty and administrators, such as hiring, firing, contract renewal, promotion, demotion, tenure, merit pay, and teaching awards.

BOTTOM LINES: As Cassius once said before eating a Caesar salad, "*The fault, dear educators, is not in our VAM, but in ourselves, that we are not able to input the right data*" (*Julius Caesar*, Act 1, Scene 2, Lines 140–141).

References

AERA (American Educational Research Association), APA (American Psychological Association), & NCME (National Council on Measurement in Education) Joint Committee on Standards. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

American Educational Research Association Council. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher, 44*(8), 448–452.

American Statistical Association. (2014). *ASA statement on using value-added models for educational assessment*. Retrieved on May 1, 2016, from http://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf

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

Amrein-Beardsley, A. (2015, December 10). *The "Every Student Succeeds Act" (ESSA) to replace "No Child Left Behind" (NCLB)*. Retrieved on June 7, 2016 from http://vamboozled.com/the-every-student-succeeds-act-essa-replaces-no-child-left-behind-nclb/

Amrein-Beardsley, A. (2016, April 4). *What ESSA means for teacher evaluation and VAMs*. Retrieved on June 1, 2016, from http://vamboozled.com/what-essa-means-for-teacher-evaluation-and-vams/

Ballou, D. Mokher, C. G., & Cavalluzzo, L. (2012). *Using value-added assessment for personnel decisions: How omitted variables and model specification influence teachers' outcomes*. Retrieved on May 12, 2016, from http://www.aefpweb.org/sites/default/files/webform/AEFP-Using VAM for personnel decisions_02-29-12.pdf

Ballou, D., Sanders, W. L., & Wright, P. S. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Education and Behavioral Statistics, 29*(1), 37–65.

Ballou, D, & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation sysytems. *Educational Researcher, 44*(2), 77–86.

Berk, R. A. (1988). Fifty reasons why student achievement gain does not mean teacher effectiveness. *Journal of Personnel Evaluation in Education, 1*, 345–363.

Berk, R. A. (1990). Limitations of using student achievement data for career ladder promotions and merit pay decisions. In J. V. Mitchell, Jr., S. L. Wise, & B. S. Plake (Eds.), *Assessment of teaching: Purposes, practices, and implications for the profession* (pp. 261–306). Hillsdale, NJ: Erlbaum. Retrieved on February 2, 2014, from http://digitalcommons.unl.edu/burosassessteaching/10/

Berk, R. A. (2005). Survey of 12 strategies for measuring teaching effectiveness. *International Journal on Teaching and Learning in Higher Education, 17*(1), 48–62. Retrieved on January 28, 2014, from http://www.isetl.org/ijtlhe/pdf/IJTLHE8.pdf

Berk , R. A. (2006). *Thirteen strategies to measure college teaching: A consumer's guide to rating scale construction, assessment, and decision making for faculty, administrators, and clinicians*. Sterling, VA: Stylus Publishing.

Berk, R. A. (2009). Using the 360° multisource feedback model to evaluate teaching and professionalism. *Medical Teacher, 31*(12), 1073–1080. DOI: 10.3109/01421590802572775

Berk, R. A. (2013a). Top 5 flashpoints in the assessment of teaching effectiveness. *Medical Teacher, 35*, 15–26. DOI: 10.3109/0142159X.2012.732247 (http://informahealthcare.com/doi/abs/10.3109/0142159X.2012.732247)

Berk, R. A. (2013b). *Top 10 flashpoints in student ratings and the evaluation of teaching: What faculty and administrators must know to protect themselves in employment decisions*. Sterling, VA: Stylus Publishing.

Berk, R. A. (2014). Should student outcomes be used to evaluate teaching? *Journal of Faculty Development, 28*(2), 87–96.

Betebenner, D. W. (2008). *A primer on student growth percentiles*. Dover, DE: National Center for the Improvement of Educational Assessment.

Betebenner, D. W. (2011). *A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories*. Dover, DE: National Center for the Improvement of Educational Assessment.

Braun, H. (2005). *Using student progress to evaluate teaching: A primer on value-added models*. Princeton, NJ: Educational Testing Service.

Braun, H. (2015). The value in value added depends on the ecology. *Educational Researcher*, *44*(2), 127–131.

Briggs, D., & Domingue, B. (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.

Chetty, R., Friedman, J., & Rockoff, J. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*. Retrieved on May 9, 2016, from http://obs.rc.fas.harvard.edu/cherry/va_exec_summ.pdf

Darling-Hammond, L. (2013). *Getting teacher evaluation right*. New York: Teachers College Press.

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, *44*(2), 132–137.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, R. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 93*(6), 8–15.

Eger, W. (2014, August 27). Education is not "Moneyball": Why teachers can't trust value-added evaluations yet. *Education Week Teacher*. Retrieved on May 23, 2016, from http://www.edweek.org/tm/articles/2014/08/27/fp_eger_valueadded.html

Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D., & Whitehurst, G. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: Brookings Institute.

Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher*, *44*(2), 87–95.

Goldhaber, D., Gabele, B., Walch, J. (2013). Does the model matter? Exploring the relationship between student achievement-based teacher assessments. *Statistics and Public Policy, 1*(10), 28–39.

Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica, 80*(319), 589–612.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher, 44*(2), 96–104.

Grissom, J. A., & Youngs, P. (Eds.). (in press). *Making the most of multiple measures : The impacts and challenges of implementing rigorous teacher evaluation systems*. New York: Teachers College Press.

Guarino, C., Rechase, M, & Wooldridge, J. (2015). Can value-added measures of teacher performance be trusted? *Education Finance and Policy, 10*(1), 117–156.

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

Harris, D. (2012, December 4). Creating a valid process for using teacher value-added measures. *Washington Post*. Retrieved on May 12, 2016, from http://www.washingtonpost.com/blogs/answer-sheet/wp/2012/12/04/the-future-of_teacher-evaluation-how-value-added-could-be-used/

Harris, D. (2013a). *How might we use multiple measures of teacher performance?* Palo Alto, CA: Carnegie Foundation for the Advancement of Teaching.

Harris, D. (2013b). *How do value-added indicators compare to other measures of teacher effectiveness?* Palo Alto, CA: Carnegie Foundation for the Advancement of Teaching.

Harris, D. (2014). *How might we use multiple measures of accountability?* Carnegie Knowledge Network Policy Brief. Retrieved on May 10, 2016, from http://www.carnegieknowledgenetwork.org/briefs/multiple_measures/

Harris, D. N., & Herrington, C. D. (2015). Editors' introduction: The use of teacher value-added measures in schools: New evidence, unanswered questions, and future prospects. *Educational Researcher*, *44*(2), 71–76.

Haertel, E. (2013). *Reliability and validity of inferences about teachers based on student test scores.* Princeton, NJ: Educational Testing Service.

Hill, H., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal, 48*(3), 794–831.

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher, 44*(2), 105–116.

Johnson, S. M. (2015). Will VAMS reinforce the walls of the egg-crate school? *Educational Researcher, 44*(2), 117–126.

Joint Committee on Standards for Educational Evaluation. (2009). *The personnel evaluation standards: How to assess systems for evaluating educators* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Kane, T. J., McCaffrey, D., Miller, T., & Staiger, D. (2013, January). *Have we identified effective teachers? Validating measures of effective teaching using random assignment* (MET Project Research Paper). Seattle, WA: Bill & Melinda Gates Foundation.

Klees, S. J. (2016). VAMs are never "accurate, reliable, and valid." *Educational Researcher, 45*(4), 267.

Klein, A. (2016, March 31). The every student succeeds act: An ESSA overview. *Education Week*. Retrieved on June 7, 2016, from http://www.edweek.org/ew/issues/every-student-succeeds-act/

Lockwood, J.R., McCaffrey, D.F., Hamilton, L.S., Stecher, B., Le, V., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement, 44*(1), 47–67.

Loeb, S. (2013). *How stable are value-added estimates across years, subjects and student groups?* Retrieved on May 15, 2016, from http://www.carnegieknowledgenetwork.org/briefs/value-added/value-added-stability/

Lyde, A. R., Grieshaber, D. C. & Byrns, G. (2016). Faculty teaching performance: Perceptions of a multi-source method for evaluation. *Journal of the Scholarship of Teaching and Learning, 16*(3), 82–94. DOI: 10.14434/josotl.v16i3.18145

Measures of Effective Teaching. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Seattle, WA: Bill & Melinda Gates Foundation.

Measures of Effective Teaching. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study* [Policy and practice brief]. Seattle, WA: Bill & Melinda Gates Foundation.

National Council on Teacher Quality. (2015, November). *State of the states 2015: Evaluating teaching, leading and learning.* Retrieved on June 8, 2016, from http://www.nctq.org/dmsStage/StateofStates2015

National Research Council. (2010). *Getting value out of value-added.* Washington, DC: National Academy Press.

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).

Papay, J. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal, 48*(1), 163–193.

Raudenbush, S. W. (2014a). *What do we know about using value added to compare teachers who work in different schools?* (Carnegie Foundation for the Advancement of Teaching Knowledge Brief). Retrieved on May 20, 2016, from http://www.carnegieknowledgenetwork.org/briefs/value-added/

Raudenbush, S. W. (2014b). *What do we know about the long-term impacts of teacher value added?* (Carnegie Foundation for the Advancement of Teaching Knowledge Brief). Retrieved on May 20, 2016, from http://www.carnegieknowledgenetwork.org/briefs/value-added/

Raudenbush, S. W., & Jean, M. (2012). *How should educators interpret value-added scores?* (What We Know Series: Value-Added Methods and Applications; Knowledge Brief). Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

Ravitch, D. (2009, September 9). The start of an interesting and dangerous school year. *Education Week*. Retrieved on May 12, 2016, from http://blogs.edweek.org/edweek/Bridging-Differences/2009/09/editors_note_bridging_differen.html

Reese, W. (2013). *Testing wars in the public schools: A forgotten history*. Cambridge, MA: Harvard University Press.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics, 125*(1), 175–214.

Rothstein, J. (2011). *Review of "Learning about teaching: Initial findings from the Measures of Effective Teaching Project."* Boulder, CO: National Education Policy Center.

Rowen, B., & Raudenbush, S. W. (2016). Teacher evaluation in American schools. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 1159–1216). Washington, DC: American Educational Research Association.

Sass, T. R., Semykina, A., & Harris, D. (2014). Value-added models and the measurement of teacher quality. *Economics of Education Review, 38*, 9–23.

Sawchuk, S. (2014, January 7). AFT's Weingarten backtracks on using value-added measures for evaluation. *Education Week*. Retrieved on May 5, 2016, from http://blogs.edweek.org/edweek/teacherbeat/2014/01/weingartens_retrenchment_on_va.html

Sawchuk, S. (2016, January 6). ESSA loosens reins on teacher evaluations, qualifications. *Education Week*. Retrieved on June 5, 2016, from http://www.edweek.org/ew/articles/2016/01/06/essa-loosens-reins-on-teacher-evaluations-qualifications.html

U.S. Department of Education. (2009). Race to the Top program. Washington, DC: Author. Retrieved on April 30, 2016, from http://www.whitehouse.gov/issues/education/k-12/race-to-the-top

U.S. Department of Education. (2010). Race to the Top scoring rubric. Washington, DC: Author. Retrieved on April 30, 2016, from http://www2.edgov/programs/racetothetop/scoringrubric

U.S. Department of Education. (2015). Every Student Succeeds Act (ESSA). Retrieved on June 5, 2016, from http://www.ed.gov/essa?src=rn

van der Lans, R. M., van de Grift, W. J. C. M., & van Veen, K. (2015). Developing a teacher evaluation instrument to provide formative feedback using student ratings of teaching acts. *Educational Measurement: Issues & Practice, 34*(3), 18–27.

Walker, T. (2014, May 30). New study strikes latest blow against "value added" teacher evaluation. *NEA Today*. Retrieved on May 11, 2016, from htpp://neatoday.org/2014/05/30/new-study-srtikdes-latest-blow-against-value-added-teacher-evaluation-2/

Weinberg, B. A., Fleisher, B. M., & Hashimoto, M. (2007). *Evaluating methods for evaluating instruction: The case of higher education* (Working Paper 12844). Cambridge, MA: National Bureau of Economic Research. Retrieved on June 7, 2016, from http://www.nber.org/papers/w12844

Weingarten, R. (2014). Where we stand. *American Educator, 38*(1), 1.

Ronald A. Berk, Ph.D., is professor emeritus, biostatistics and measurement, and former assistant dean for teaching at The Johns Hopkins University. Now he is a speaker, writer, PowerPoint coach, and jester-in-residence. He can be contacted at rberk1@jhu.edu, www.ronberk.com, www.pptdoctor.net, or www.linkedin.com/in/ronberk/, and blogs at http://ronberk.blogspot.com.

The Active Learning Classroom: Strategies for Practical Educators

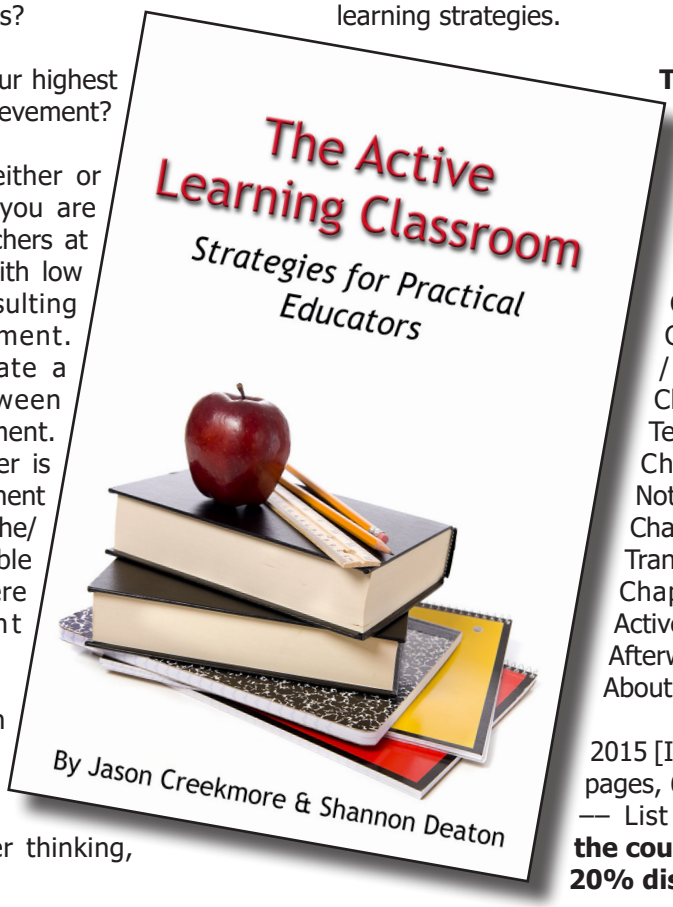
The goal of the book is simple: To improve student achievement by helping teachers implement active learning strategies in the classroom. To begin, consider the following two questions in relation to your own classroom:

1. Are your students actively engaged throughout the entirety of your daily lessons?

2. Are students meeting your highest expectations regarding achievement?

If you answered 'no' to either or both of these questions, you are not alone. Classroom teachers at all levels are challenged with low student engagement, resulting in low student achievement. Numerous studies indicate a positive correlation between engagement and achievement. For this reason, the teacher is the most important component of the learning process, as he/she is ultimately responsible for creating an atmosphere conducive to student achievement.

Active Learning has proven to be one of the most important tools for engaging students, promoting skills in motivation, higher-order thinking, communication, creative thinking, and problem-solving. Most teachers agree that these skills are essential for increasing student achievement; however, these skills are difficult to foster in the traditional 'sage on a stage' model. Educators must learn to adopt a new 'guide on the side' teaching paradigm whereby traditional instruction is supplemented by active learning strategies.



The Contents

2015 [ISBN: 1-58107-287-2; 92 pages, 6 x 9 inches, soft cover] –– List price $15.95 –– **use the coupon code below for a 20% discount.**

20% Discount on Our Titles

For a 20% discount on titles featured in this issue, call 405-372-6158 and provide code JFD001 with your order; or, fax an order to 405-377-2237, giving the code with your order. You may also order from our online store at http://store.newforums.com, and enter the discount code in the coupon section of your order.

NEW FORUMS PRESS