Value-Added Imprecision

Lauren A. Menard

laurenannmenard@gmail.com

menardl@nsula.edu.com

ABSTRACT

From employment termination and pay incentives to reflections on teacher preparation programs, federalism has delineated consequences for the academic growth of students— a carrot or stick for the value teaching has added. A position of this article is that value-added measures of teacher effectiveness are less objective than presumed. Quantitative, standardized, and publicly disclosed teacher effectiveness ratings rely on student growth measuring tools with recognized distortions and inconsistencies. Extant literature on the imprecision of measuring student growth was reviewed, and the following six areas were identified: (a) construct shifting, (b) measurement systems, (c) snapshot summative assessments, (d) percentage increases, (e) student characteristics, and (f) prior ineffective teaching. The present discussion opens dialogue into equitable comparisons of effectiveness among teachers and the institutions that have prepared them for the classroom.

Keywords:

INTRODUCTION

Teacher evaluations now play a major role in school improvement (Murphy, Hallinger, & Heck, 2013). School districts in 21 states evaluate teacher effectiveness with a value-added component, or some measure of the academic advancement or worth teaching contributed (The Center for Greater Philadelphia, 2004). Value-added measures (VAM) analyze the test scores of students and estimate a teacher's effect on academic gains (Kupermintz, 2003). Price (2013) described new educational reform that "bind teacher evaluation to student test performance" (p. xv1), and for the purpose of this paper, VAM are considered measures of student academic growth that are used in any way to determine a teacher's effectiveness rating. The trend in current standards-based educational reform is for states to revise teacher evaluation systems to include measures of student academic growth in order to qualify for Race to the Top (U. S. Department of Education, 2009) funds and receive No Child Left Behind (NCLB, 2002) accountability waivers (Education Week, 2013; Herlihy, Karger, Pollard, Hill, Kraft, Williams, & Howard, 2012).

Teacher unions, teacher preparatory institutions, and other stakeholders have heeded a call to arms to secure a voice in revolutionary educational reform initiatives that incentivize highly effective teachers (Price, 2013) and publicly humiliate ineffective teachers (Darling-Hammond, 2012). An example of ardent public interest is the Los Angeles Times (2012) public database, Grading the Teachers: Value-added Analysis, which lists the names and evaluation scores of area teachers. In New York City, the teacher publicly identified as the worst teacher of the year was subjected to hounding by reporters who questioned "her lack of skill and commitment" (Darling-Hammond, 2012, ¶2). This tabloid-type treatment epitomizes Price's (2013) description of a new educational landscape where the teacher is to blame for a student's lack of academic growth. Similarly, teacher education is held responsible for the lack of effective teachers and the resulting failures of public education (Price, 2013). However, teachers and teacher preparatory programs are not the only culprits to consider.

The imprecision of tools and procedures utilized for measuring the academic growth of students distort teacher effectiveness ratings, to some degree. According to Briggs and Domingue (2011):

There is a great irony here that following a decade in which there has been a great push to increase the scientific rigor of educational evaluations of what works through an increasing emphasis on the importance of experimental designs, much of this seems on the verge of being thrown out the window in pursuit of teacher evaluations with non-experimental designs that would not be eligible for review were they to be subject to the standards of the What Works Clearinghouse. (p. 19)

Looking past test scores in New York City, for example, revealed the city's worst teacher was actually an experienced and admired teacher, a teacher with administrator support— not a bad teacher— an English as a Second Language teacher who signed up for the challenge of teaching new immigrant students (Darling-Hammond, 2012).

When student growth is determined with imprecise measurement and unclear procedures and teacher ratings are publicly compared, advantage should go to the classroom teacher whose professional reputation and livelihood are placed in jeopardy by the limitations of high-stakes teacher effectiveness ratings based on measures of student growth.

The purpose of the current position paper was to review extant literature for identified areas of imprecision in the tools and procedures used to measure the academic growth of students. The review began with background into current standard's based educational reform and included a discussion of how federal and political interests in reducing income and racial achievement disparities have increased a classroom teacher's accountability for low student growth. Following these discussions, six areas of imprecision in measuring student academic growth were identified (i.e., construct shifting, measurement system, snapshot summative assessments, percentage increases, student characteristics, and prior ineffective teaching).

BACKGROUND

In the interest of regaining America's reputation as the "best educated nation in the world" (U. S. Department of Education, 2010, p. 1), the Common Core State Standards (CCSS) reform initiative is looked to as a vehicle for improving low-performing schools. President Barak Obama noted America's falling National Assessment of Educational Progress (NAEP) ranking as one justification for the Reauthorization of the Elementary and Secondary Education Act (A Blueprint for Reform): "A generation ago, we led all nations in college completion, but today, 10 countries have passed us" (U. S. Department of Education, 2010, p. 1). Forty-five states and three United States territories (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010) have adopted CCSS. Thus, as teachers in classrooms all across America are adjusting to more rigorous, common-standards aligned instruction and assessment, teacher effectiveness is being simultaneously evaluated with an unprecedented ascendancy of student academic growth (Noell & Burns, 2006).

Systemic Academic Failures

High expectations for the academic growth of all students based on common assessments more closely aligned with the rigor of NAEP assessments would be less controversial if American students were generally performing satisfactory on current standardized assessments. However, Hoff (2009) observed the number of schools failing to make adequate yearly progress increased 28% in the 2007-2008 school year alone. In November 2012, Secretary of Education Arnie Duncan (2012) acknowledged, "Our graduation rate is unacceptable; Our opportunity gap is unacceptable; Our achievement gap is unacceptable" (para. 5). There were thousands of schools in the United States "where as few as 10 percent of students were reading or doing math at grade level" (U.S. Department of Education, June 25, 2013, para. 16). Swanson and Stevenson (2002) gives context to these epidemic failures at demonstrating adequate student academic progress with the observation that most state assessments, the measures used to determine student growth under NCLB guidelines, fell below NAEP proficiency standards.

Confronting widespread failures in meeting outcome-based academic growth requirements, the Obama administration began granting waivers to exempt states from NCLB accountability (Education Week, 2013). The Secretary of Education called the 2014 NCLB deadline for grade level proficiency of all students lofty and not credible:

Faced with meeting a utopian goal, too many states took the easy path. They dummied-down their standards to make it look like more students were proficient. And too many schools—like those that we are honoring tonight—that were successfully educating black, brown, and poor children, and were actually closing achievement gaps, were labeled as failures. (Duncan, 2012, para. 10)

An Urgency for Reducing Achievement Gaps

Reducing racial and income disparities in academic achievement has become the new focus of educational reform: "Contrary to what you may have read, these waivers will push states to dramatically accelerate achievement and attainment for disadvantaged students and students of color" (Duncan, 2012, para. 10). Groups such as African American race, low income (free and reduced lunch status), and special education populations were counted as subgroups in NCLB formulas for determining the adequate yearly progress of —not teachers—but schools, school districts, and states. The new ambitious and achievable goal of reducing racial and income disparities by 50% "takes account of the fact that some students start far behind their more privileged peers" (Duncan, 2012, para. 19). According to Duncan (2012), the NCLB held only 20% of schools accountable for the academic growth of African American students, in comparison to the 99% of schools held accountable under new legislation. Duncan's (2012) accountability percentage is based on signature educational reform legislation of the current administration (U. S. Department of Education, 2009; 2010) that require effective teachers to evidence academic growth for all students in their classrooms.

Do new educational reform initiatives actually call for greater accountability? If it was so easy for states to

game the system (Duncan, 2012) and show acceptable student growth under previous legislation, why would NCLB goals be seen as so lofty and unattainable that 41 states requested waivers from NCLB accountability targets calling for the grade-level proficiency of all students (Education Week, 2013)? A more harmful utopian notion than NCLB's 2014 goal of 100% of students being proficient in math and reading (Education Week, 2013) may be the assumption that no racial or economic academic achievement gaps would exist in America if all classroom teachers were highly effective.

Blaming the Teacher

Whereas the NCLB held state, district, and school consequences for low and falling school performance scores, teacher evaluation systems based on measures of student growth have increased the classroom teacher's burden of accountability for low-performing students (e.g., at-risk, English-learners, minority, special education, and low-income). Teachers have been terminated for one year of low student growth, even when they evidenced student growth in previous years (Darling-Hammond et al., 2013). In Good Teachers can be Fired for Bad Math, Downey (March 23, 2012) observed Florida, for example, will fire and revoke the teaching licenses of "teachers with two years of ineffective rankings...thus banning them from teaching in any other public school in the United States" (para.7).

The problem of low-performing subgroups, evidenced under NCLB guidelines as low school and district performance scores, have persisted through years of concentrated state, district, and school efforts, and this problem is unlikely to be solved by effective teachers alone. Why is it assumed that concentrating on the administrative task of teacher evaluations will "accomplish what reformers hope" (Murphy, Hallinger, & Heck, p. 1)? Numbers, proportions, and ratings may change, but the academic growth of low-performing subgroups will not improve simply by holding teachers, rather than districts or states primarily responsible. Teachers teach standards, "standards will not teach themselves" (Klein & Rice, 2012, p. 62), and effective interventions for unsatisfactory academic progress require considerably more support than simply raising the academic bar.

Oversimplification of the multi-leveled, cross-disciplinary, complex societal challenge of the unsatisfactory academic growth of low-performing subgroups epitomizes a silver bullet mentality of educational reform— pre-packaged, top-down solutions that detract from authentic, sustainable progress (Klein & Rice, 2012). Narrowly focusing on teacher effectiveness as a root cause of systemic academic failure inhibits development of appropriate interventions and intense instructional supports, such as increased one-on-one and small group tutoring; social work and behavioral interventions; high-yield materials, including computers and software; specialized programming (e.g., phonics-based reading instruction for older students); and supports and supplementary aids for special needs.

Blaming teachers for low student growth is inherent in evaluation systems where a teacher's contribution to student learning is represented by one rating label that is used to predict and compare teacher effectiveness. Not all teachers have the same assignment and what is needed with this year's students is not the same as what was needed last year and what will be needed next year. Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2013) observed variability in year-to-year teacher effectiveness ratings and questioned the premise of "a stable teacher effect that's a function of the teacher's teaching ability" (p. 12). Similarly, in observations of conditional teacher effectiveness and the student's role in an interactive learning process, Ding and Sherman (2006) questioned "the search for one-size-fit-all kind of teaching effectiveness" (p. 46). In the following section, six areas of imprecision in the tools and procedures used for measuring the student growth that determines teacher effectiveness ratings in VAM evaluations were identified.

VAM IMPRECISION

CONSTRUCT SHIFTING

Students often repeatedly cover the same course or content area, but they are learning different constructs in different grade levels. Construct shifting refers to changes in content focus across grade levels, "such as when mathematics assessments move from testing arithmetic skills in third grade to testing pre-algebra and geometry skills in later grades" (O'Malley, Murphy, McClarty, Murphy, & McBride, 2011, p. 5). Martineau (2006) observed, "If value-added accountability models are to be used on vertically scaled data, new methods of assuring that minimal construct shift occurs are imperative" (p. 58). Student learning varies because learning rate is not necessarily similar (even for the same student) across constructs because students do not repeatedly learn the same constructs or skills. Value-added evaluations that compare a student's learning rate with one teacher to a normed rate by content or compare it to the student's past growth rate assume similar learning rates across constructs. According to Martineau (2006), approaches for assuring minimal construct shifting include, "reporting student scores on the various score scales within subject matter rather than reporting only a single combined general math, general reading, and/or general science scale" (p. 58). Tools for measuring student achievement, however, are typically by the more general grade-level content.

MEASUREMENT SYSTEM

Teacher effectiveness ratings differed for the same teacher by the sensitivity of analyses (Briggs & Domingue, 2011) and by the model selected (Di Carlo, 2012; Harris, Sass, & Semykina, 2010; McCaffrey, Lockwood, Koretz, & Hamilton, 2004). A critical review of a Los Angeles Unified School District's (LAUSD) analysis of teacher effects on student performance was conducted (Briggs & Domingue, 2011). A guiding question was whether teacher effects were successfully isolated. When the LAUD data were re-analyzed with measures for greater sensitivity (i.e., student scores over a longer period of time, the influence of peers, and school-level variables), "only 46. 4% of teachers would retain the same effectiveness rating" (Briggs & Domingue, p. 5). Teacher evaluation ratings also varied according to which tests were included in a selected model (Di Carlo, 2012; Papay, 2011). Student growth scores differ even among instruments aligned to CCSS because (a) tests assign different weights to constructs, (b) larger and smaller proportions of specific constructs (e.g., computation or higher order prompts) are included and excluded differently among assessments, and (c) tests combine constructs in varying ways within testing domains.

Common Core State Standards are standard across states, but how much weight the student academic growth component is given in teacher evaluation processes differs among states. Hypothetically, then, a teacher with the exact same student characteristics and student growth scores— gathered with the same instruments and analyzed exactly the same— could receive different effectiveness ratings in different states.

STUDENT CHARACTERISTICS

The academic growth of students varies by student characteristics and by grade level (Briggs & Domingue, 2011; Capraro, Young, Lewis, Yetkiner, & Woods, 2009; Deno, Fuchs, Marston, & Jongho, 2001). Darling-Hammond (2012) found teacher effectiveness, as determined by student growth, continued to vary according to student characteristics even when controls were applied for student differences. In particular, teacher gains were lower with new English-learners and special education students than with typical students (Darling-Hammond, 2012). Deno et al. (2001) determined that students with disabilities learned new words at less than half the growth rate of typical students. According to Harris et al. (2010), teacher effects are highly sensitive to the mis-specifications and assumptions regarding student/family effects (i.e., time-constant versus trending). With recognition of the instability of effectiveness ratings, Darling-Hammond (2012), once a bullish supporter of VAM, observed, "Test scores reflect whom a teacher teaches, not how well they teach" (¶14).

Teacher effectiveness ratings are compromised by stacking a class with harder to teach, harder to reach, lower performing students because teacher effectiveness ratings vary according to student characteristics. The biggest educational risk of determining teacher effectiveness by student growth is penalizing and discouraging teachers from taking those difficult assignments where student growth is predictably lower than that of average students (Darling-Hammond, 2012).

SNAPSHOT SUMMATIVE ASSESSMENT

Evaluating the effectiveness of every teacher with a one size fits all summative measure of student performance (i.e., proportion of students scoring proficient on state assessment, one standard percentage-point growth for all students) is similar to a medical analogy of evaluating all hospitals on one patient outcome (i.e., survival). Trauma and high-risk specializing centers, analogous to instructional settings that serve students with special learning needs, can be labeled unacceptable in evaluation systems based on a single-snapshot, outcome-based measure because variations in the conditions of the patient upon arrival is not a factor. Teachers should be expected to improve student skills from individualized, pre-instructional performance levels— not from a theoretical grade level standard of where the student should be functioning. Teachers should not be penalized for failing to bring students multiple grade levels below grade level proficiency at the beginning of the year to a summative, standardized proficiency level by year's end.

According to O'Malley et al. (2011), inaccurate student growth models measure change in proportions of students meeting proficiency compared to a set proficiency level or proficiency levels of different cohorts; whereas, accurate growth models compare score changes from the same student over time. Teachers are best able to evidence the value teaching has added to student learning by taking the academic pulse of individual students when they enter the classroom and by projecting learning targets on the same pre-test constructs from those initial baselines. Yet, as noted earlier, comparing a student's learning rate this year to his or her learning rate last year or projecting an individualized learning rate for next year is not precise because different constructs are being learned at each grade level.

PERCENTAGE INCREASE

Teacher performance ratings based on objective percent point increases in academic growth are not always fair**.** Increasing the percentage growth required for all students across the board or individualizing growth rates do not

remedy all concerns, however, because the fundamental question of percentage increase remains. Adequate yearly progress and school performance scores held similar dilemmas— a greater difficulty to sustain and improve growth at highest levels and a distortion in meaningful growth at lowest levels. One option is to project growth with larger percentages for higher achieving students. Yet, according to Darling-Hammond et al. (2013), "Teachers of gifted students show little value-added because their students are already near the top of the test score range" (p. 12). Another option for making standardized percentage increases more equitable is to project individualized percentage-point increases determined from a rate of growth for individual students gathered from the student's previous assessment history. However, is learning 100% more words, for example, from one word to two words, as meaningful as jumping from 100 to 200 words, and is equating these two examples of 100% increases in academic growth fair?

PRIOR INEFFECTIVE TEACHING

The long-term residual effects of prior ineffective teaching play a part in subsequent student gains. In other words, this year's teacher can be penalized by the ineffectiveness of last year's teacher. According to Sanders and Rivers (1996), "Teacher effects are both additive and cumulative with little evidence of compensatory effects of more effective teachers in later grades" (p. 6). How can Sanders and Rivers (1996) classic study on the importance of effective teaching also support the imprecision of determining teacher effectiveness by student growth? Effective teaching is vitally important because a "near-permanent retardation of academic achievement" (Sanders & Rivers, 1996, p. 7) was determined to be present in students who experienced a multiple year sequence of ineffective teaching. At the same time, prior ineffective teaching jeopardizes the effectiveness rating of this year's teacher because the residual effects of ineffective teaching remain "measurable two years later, regardless of the effectiveness of teachers in later grades" (Sanders & Rivers, 1996, p. 6).

DISCUSSION AND CONCLUSION

Utilizing one-size fits all measurements of student growth as dominate factors in teacher evaluations has issues yet to be resolved, but it is not the intent to imply value-added measures are meaningless in determining teacher effectiveness (DiCarlo, 2012). Long-standing educational research have made it clear teachers do make a difference (Sanders & Horn, 1998; Sanders, Wright, & Horn, 1997). According to Sanders, Wright, and Horn (1997), "the most important factor affecting student learning is the teacher" (p. 63). However, finding teacher effectiveness the strongest predictor of student growth is not the same thing as assuming a teacher is not effective if he or she did not exert a given amount of academic growth for all students— in all grade levels assigned, for all skills taught, and in all teacher evaluation models— every year.

The principal supervises teacher evaluations in the scaffold hierarchy of performance management systems (Murphy, Hallinger, & Heck, 2013), but principals would do better to focus on other factors within their control (e.g., academic vision and mission, organization goals, student opportunity, school culture) for improving instructional quality. Optimistic conclusions regarding a teacher's influence on student learning have been made and there is a growing emergence of teacher evaluation models with measures of student growth, but caution "in accepting claims about the ability of teacher evaluation to power significant school improvement" (Murphy et al., 2013, p.3) is recommended. Good questions for principals of schools in need of turn around to ask are, "What is to happen with low-quality teachers? … Where is the reservoir of high-quality teachers waiting in the wings?" (Briggs & Domingue, 2011, p. 20).

Existence of educational improvement's central player on the federal stage— a consistent, standard measure of teacher effectiveness— has been legitimately questioned (Darling-Hammond et al., 2013; Ding & Sherman, 2006; Harris et al., 2010; McCaffrey et al., 2004; McCaffrey, Lockwood, Mihaly, & Sass, 2010). Classic educational research has established that the effects of ineffective teaching lasts at least two years (Sanders & Rivers, 1996), but teachers are being terminated, rather than trained, after two years of failing to demonstrated acceptable student growth (Downey, March 23, 2012). Low-achieving students have been shown to affect the performance ratings of teachers, even with controls for student differences (Darling-Hammond, 2012). The assertion that teacher effectiveness ratings are not linked to "classroom demographic variables such as gifted and talented status, special needs, ELL status and poverty" (Briggs & Domingue, 2011, p.19) was not supported by sensitivity analysis. It follows, with the imprecision involved in determining teacher effectiveness, the ability of the effectiveness of teachers to close achievement gaps is questionable.

An aim of the current review was to open dialogue into the standardized quantifying of teacher effectiveness. Practical implications are held for informing equitable comparisons of effectiveness among teachers and the teacher preparatory institutions that trained them.

REFERENCES

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. *National Education Policy Center*. Retrieved from http://files.eric.ed.gov/fulltext/ED516008.pdf

Capraro, R. M. , Young, J. R. , Lewis, C. W. , Yetkiner, Z. E. , & Woods, M. N. (2009). An examination of mathematics achievement and growth in a Midwestern urban school district: Implications for teachers and administrators. *Journal of Urban Mathematics Education, 2*(2), 46-65.

Darling-Hammond, L. (2012). Value-added evaluation hurts teaching. *Education Week*. Retrieved from http://www.edweek.org/ew/articles/2012/03/05/24darlinghammond_ep.h31.html

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2013). Evaluating teacher evaluation. *Phi Delta Kappan*, *93*(6), 8-15.

Deno, S. L., Fuchs, L. S., Marston, D., & Jongho, S. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. S*chool Psychology Review, 30* (4), 507-524.

Di Carlo, M. (2012). How to use value-added measures right. *Educational Leadership, 70*(3), 38-42.

Ding, C., & Sherman, H. (2006). Teaching effectiveness and student achievement: Examining the relationship. *Educational Research Quarterly*, *29*(4), 40-51.

Downey, M. (March 23, 2012). Good teachers can be fired for bad math. *ajc.com.* Retrieved from http://blogs.ajc.com/get-schooled-blog/2012/03/23/good-teachers-can-now-be-fired-because-of-bad-math/

Duncan, A. (2012). *Remarks of U. S. Secretary of Education Arne Duncan at the education trust conference.* [Press Release (202) 401-1576]. Retrieved from http://www.ed.gov/news/speeches/remarks-us-secretary-education-arne-duncan-he-education-trust-conference

Education Week (2013). *NCLB waivers: A state-by-state breakdown* [map]. Retrieved from

http: //www. edweek. org/ew/section/infographics/nclbwaivers. html

Harris, D., Sass, T., & Semykina, A. (2010). Value-added models and the measurement of teacher productivity [CALDER Working Paper 54]. *Center for Analysis of Longitudinal Data in Education Research.* Retrieved from http://www. urban.org/UploadedPDF/1001508-Measurement-of-Teacher-Productivity.pdf

Herlihy, C., Karger, E., Pollard, C., Hill, H. C., Kraft, M. A., Williams, M., & Howard, S. (2012). State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teacher College Record*. Retrieved from http://www.gse.harvard.edu/cepr-resources/files/news-events/ncte-valid-reliable-high-stakes-conference-herlihy-hill.pdf

Hoff, D. J. (2009). Schools struggling to meet key goal on accountability: Number failing to make AYP rises 28 percent. *Education Week*, *28*(16), 1-14.

Klein, J. I., & Rice, C. (2012). US education reform and national security (No. 68). *Council on Foreign Relations.* Retrieved from http://www.cfr.org/united-states/us-education-reform-national-security/p27618

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System. *Educational Evaluation and Policy Analysis, 25*(3), 287-298. doi: 10.3102/01623737025003287

Los Angeles Times (2012). *Grading the teachers: Value-added analysis* [database]. Retrieved from http://www.latimes.com/news/local/teachers-investigation/

Martineau, J. A. (2006) Distorting value added: the use of longitudinal, vertically scaled student achievement data for growth-based Value-added accountability. *Journal of Educational and Behavioral Statistics, 31*(1), 35-62.

McCaffrey, D., Lockwood, J. R., Koretz, D., & Hamilton, L. (2004). Evaluating value-added

models for teacher accountability. Santa Monica, CA: RAND Corporation.

McCaffrey, D. F., Lockwood, J. R., Mihaly, K., & Sass, T. R. (2010). *A review of Stata routines for fixed effects estimation in normal linear models.* Retrieved from http://www.researchgate.net/publication/228420110_A_Review_of_Stata_Routines_for_Fixed_Effects_Estimation_in_Normal_Linear_Models/file/60b7d51894cc96aa01.pdf

Murphy, J., Hallinger, P., & Heck, R. H. (2013). Leading via teacher evaluation the case of the missing clothes?.

Educational Researcher, *42*(6), 349-354.

Noell, G., & Burns, J. (2006). Value-added assessment of teacher preparation: An illustration of

emerging technology. *Journal of Teacher Education, 57*(1), 37-50. doi: 10.

1177/0022487105284466

O'Malley, K. J., Murphy, S., McClarty, K. L., Murphy, D., and McBride, Y. (2011). *Overview of student growth models* (Pearson White Paper). Retrieved from http: //researchnetwork. pearson.com/wp-content/uploads/StudentGrowthWP083111.pdf

Papay, J. (2010). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal, 48*(1), 163–193. doi: 10.3102/0002831210362589

Price, T. A. (2013). *Defending public education from corporate takeover*. Lanham, MD: University Press of America.

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12*(3), 247-256.

Sanders, W. L., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future

student academic achievement. Knoxville, TN: University of Tennessee Value-Added

Research and Assessment Center. Retrieved from http://news.heartland.org/sites/all/modules/custom/heartland_migration/files/pdfs/3048.pdf

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 57-67.

Swanson, C. B., & Stevenson, D. L. (2002). Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessments. *Educational evaluation and policy analysis*, *24*(1), 1-27.

The Center for Greater Philadelphia. (2004). *Value-added assessments*. Retrieved from http: //www.cgp.upenn.edu/ope_value.html#9

U. S. Department of Education. National Governors Association and State Education Chiefs Common Core Standards. (June 2, 2010). Retrieved from: http: //www.ed. gov/news/press-releases/statement-national-governors-association-and-state-education-chiefs-common-core-

U. S. Department of Education. (2009). *Race to the Top program executive summary.* Retrieved from http: //www2. ed. gov/policy/elsec/leg/blueprint/blueprint.pdf

U.S. Department of Education. Office of Planning, Evaluation and Policy Development. (2010). *ESEA blueprint for reform.* Retrieved from

http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf

U.S. Department of Education (June 25, 2013). *Duncan pushes back on attacks on common core standards*. Retrieved from http://www.ed.gov/news/speeches/duncan-pushes-back-attacks-common-core-standards