

Eastern Illinois University The Keep

Faculty Research and Creative Activity

Educational Leadership


November 2012

Growth Model Evaluations: Possibilities and Pitfalls

John A. Dively Jr

Eastern Illinois University, jdively@eiu.edu

Follow this and additional works at: http://thekeep.eiu.edu/edadmin_fac

 Part of the [Education Law Commons](#), and the [Elementary and Middle and Secondary Education Administration Commons](#)

Recommended Citation

Dively, John A. Jr, "Growth Model Evaluations: Possibilities and Pitfalls" (2012). *Faculty Research and Creative Activity*. 1.
http://thekeep.eiu.edu/edadmin_fac/1

This is brought to you for free and open access by the Educational Leadership at The Keep. It has been accepted for inclusion in Faculty Research and Creative Activity by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.

Growth Model Evaluations: Possibilities and Pitfalls

A Paper Presented at the Education Law Association National Conference

Sonesta Resort, Hilton Head Island, South Carolina

Saturday, November 10, 2012

John Dively J.D., Ed.D.

Associate Professor

Educational Leadership Department

Eastern Illinois University

Abstract: In response to Race to the Top mandates, student academic growth models are being incorporated into teacher evaluation processes across the country. Illinois' version of the reform is the Performance Evaluation Reform Act. This paper briefly summarizes the new law and its impact to date. Further, the paper provides reflection upon the current research related to VAMs, and the possible legal consequences of relying on student growth models as a significant component of teacher personnel decisions.

The Performance Evaluation Reform Act

The Illinois General Assembly passed the Performance Evaluation Reform Act (PERA) on January 13, 2010, and Public Act 096-0861 was signed into law by Governor Quinn on January 15, 2010.¹ PERA was enacted in part to establish the prerequisites necessary for Illinois' successful application to the federal Race to the Top (RTTT) grant program. The state legislation, coupled with the accompanying rules and regulations² established by the Illinois State Board of Education in May 2012, mandates that districts and teachers' associations establish new teacher evaluation systems consistent with its prescribed criteria. Of particular interest here is the requirement that future evaluations for teachers incorporate student growth as a significant factor:

b) By no later than the applicable implementation date, each school district shall, in good faith cooperation with its teachers or, where applicable, the exclusive bargaining representatives of its teachers incorporate the use of data and indicators on student growth as a significant factor in rating teaching performance into its evaluation plan for all teachers....³

Under the legislation, districts are required to create a joint committee composed of equal representation from the district, teachers and/or the collective bargaining unit. Once this committee has convened, it has 180 days to complete its work, or the default state evaluation plan (to be devised) will be mandated.⁴ The state plan requires 50% of the teacher evaluations to incorporate student achievement measures.⁵ District devised plans must incorporate student achievement measures as at least 25% of the evaluation for the first two years and at least 30% thereafter.⁶

On June 13, 2011, P.A. 97-0008 (SB7) became law and thereby incorporated PERA further into future school personnel decision-making. In particular, its provisions require the application of PERA to decisions such as remediation,⁷ reductions-in-force and recall rights,⁸ tenured teacher dismissals,⁹ certificate revocation for incompetency,¹⁰ and tenure acquisition.¹¹

One-hundred and one days after the state's rules and regulations became effective on May 21, 2012,¹² three hundred Chicago Public Schools (CPS) were required to implement the evaluation components of the new legislation effective September 1, 2012.¹³ Issues relating to how the performance evaluation requirements would be negotiated into the collective bargaining agreement were among the key issues that led to a well-publicized seven day work stoppage by the Chicago Teachers Union (CTU) which ended on September 18, 2012.¹⁴ The rest of the CPS schools are required to implement PERA by September 1, 2013. The deadline for "down-state" schools' (all districts except Chicago Public Schools) implementation of the student achievement evaluation components of PERA is set for September 1, 2016, except for the lowest performing 20% which were given a deadline of September 1, 2015.¹⁵

PERA requires that evaluation plans incorporate multiple measures of student growth.¹⁶ Three types of assessments are identified. Type I assessments are standardized tests that are administered statewide or outside Illinois. Type II assessments are used district-wide by all teachers in a given grade or subject area. Type III assessments are assessments that are approved by the evaluator and the teacher. Under the rules, every teacher's evaluation growth factor will be judged by at least one Type I or Type II and one Type III assessment. If no Type I or II assessment is appropriate, then two Type III assessments will be used.¹⁷

Value-Added and Growth Models

On the one hand, many policy-makers believe that data collected through VAMs is a key component for improving teacher quality and for informing broader educational reform efforts.

Andrew J. Rotherham summarized this sentiment in his 2010 article:

Because value-added models can control for other factors impacting student test scores, the most important being whether a student arrived in a teacher's classroom several grade levels behind, this method of analysis can offer a more accurate estimate of how well a particular teacher is teaching than simply looking at the latest set of student test scores. High-flying teachers can be recognized and low performers can be identified before they spend years doing a disservice to kids. Science and technology to the rescue!¹⁸

Despite important technical differences, the titles VAM and Growth Model are often used interchangeably to describe systems that measure student learning gains over a period of time. In sum, value-added models attempt to identify how much of a student's achievement is due to the influence of a particular teacher while controlling for other variables. Growth models employ less sophisticated statistical methodologies to identify simple measures of student achievement (e.g., pre and post-test models) without controlling for other variables.¹⁹ Joint committees throughout Illinois must determine what type of growth model to implement as part of the new teacher evaluation processes. The type and sophistication of the measurement models and the type of assessments adopted in Illinois have been left up to individual districts to negotiate with their teachers through the joint committees.²⁰

PERA required districts to design and implement performance evaluation systems that assess teachers' professional practice and incorporate measures of student growth. Through the joint committee, district administrators must work

with teachers/union representatives to develop evaluation systems that incorporate student growth. School districts and the state must ensure that these performance evaluation systems are valid and reliable and help teachers to improve student outcomes.²¹

Critiques of Value-Added and Growth Models

However, there is no shortage of concern expressed regarding recent efforts to reform teacher evaluation.²² As Helen M. Hazi and Daisy Arredondo Rucinskcki so aptly put it:

Examining changes in statutes and policy on teacher evaluation may shed light on the assumptions underlying such policies, and illustrate that “theories of action” connecting increased controls of teacher performance may rest on tenuous and uncertain linkages.²³

Shortly after RTTT was adopted, the Board on Testing and Assessment (BOTA) of the National Research Council (which is a branch of the National Academies) sent a letter to the Secretary of Education, Arne Duncan, on October 5, 2009 addressing the use of student performance measures to inform the process of educational reform.²⁴ The bulk of the letter outlined the limitations of RTTT student achievement measures for the purposes of evaluation.

In part, the Board stated:

However, BOTA has significant concerns that the Department’s proposal places too much emphasis on measures of growth in student achievement (1) that have not yet been adequately studied for the purposes of evaluating teachers and principals and (2) that face substantial practical barriers to being successfully deployed in an operational personnel system that is fair, reliable, and valid.²⁵

The Board characterized “value-added models” as a method to measure student achievement attributable to a teacher or a school in a given year, but cautioned the Secretary about using such models:

The term “value-added model” (VAM) has been applied to a range of approaches, varying in their data requirements and statistical complexity. Although the idea has intuitive appeal, a great deal is unknown about the potential and the limitations of alternative statistical models for evaluating teachers’ value-added contributions to student learning. BOTA agrees with other experts who have urged the need for caution and for further research prior to any large-scale, high-stakes reliance on these approaches....²⁶

Similarly, a coalition of eighty-eight professors known as the Chicagoland Researchers and Advocates for Transformative Education (CReATE) provided an open letter to Mayor Rahm Emanuel, CPS CEO Jean-Claude Brizard and the Chicago School Board on May 26, 2012. In it, CReATE recommended that CPS “pilot and adjust the evaluation system before implementing it on a large scale,” and “minimize the percentage that student growth counts in teacher or principal evaluation.”²⁷ These recommendations were based on three stated concerns:

1. CPS is not ready to implement a teacher-evaluation system that is based on significant use of “student growth.”
2. Educational research and researchers strongly caution against teacher-evaluation approaches that use Value-Added Models (VAMs).
3. Students will be adversely affected by the implementation of this new teacher-evaluation system.²⁸

Researcher Richard Rothstein noted the intuitive appeal of using growth models to gauge teacher effectiveness. However seductive the models may be, Rothstein cautioned that their use would lead the education community to over-attribute student learning to school factors. As he stated, “...differences in quality of schools can explain about one-third of the variation in student achievement. But the other two-thirds is attributable to non-school factors.”²⁹

Rothstein identified some of the non-school factors that have an effect on student learning as follows:

1. Parental factors: Education levels, low vocabulary level, poor health
2. Economic stress
3. Lack of a stable and secure home environment
4. Lack of preventative health care
5. Lack of travel and cultural experiences
6. Residence in a zip code without educated adult role models³⁰

In another article that same year, Rothstein and his colleagues provided an even fuller critique.³¹ In sum, the authors contended that "...there is broad agreement among statisticians, psychometricians, and economists that student test scores alone are not sufficiently reliable and valid indicators of teacher effectiveness to be used in high-stakes personnel decisions, even when the most sophisticated statistical application, such as value-added modeling is employed. For a variety of reasons, analyses of VAM results have led researchers to doubt whether the methodology can accurately identify more and less effective teachers."³²

Linda Darling-Hammond, et. al., wrote recently that attributing student achievement gains (or lack thereof) to an individual teacher's effectiveness assumes too much. She and her co-authors wrote that there are many other factors that impact student learning beyond the control of the teacher. Some of the factors included academic growth of classmates, classroom context, school resources, home and community supports or challenges, individual student characteristics, peer culture, previous teachers, other current teachers, summer learning loss differentials, and the tests used.³³ These researchers summarized three main problems for using VAM models for teacher evaluation:

1. Value-added models of teacher effectiveness are inconsistent.
2. Teachers' value-added performance is affected by the students assigned to them.
3. Value-added ratings cannot disentangle the many influences on student progress.³⁴

Potential Causes of Action to Adverse Employment Decisions

Under Illinois' PERA, teachers will have up to 50% of their evaluations based on fundamentally flawed methods of measuring student achievement. Based on the psychometric problems with VAMs and Growth Models that have been documented in the research, it seems certain that dismissed teachers' attorneys will seek redress in part based on a lack of due process based on the Fourteenth Amendment. As in the past, local administrative procedural errors in conducting the evaluation process, related collective bargaining issues, and state contract law in general would all be reviewable. In 2010, on his blog, School Finance 101, Bruce D. Baker wrote the following:

This new crop of state statutes and regulations which include arbitrary use of questionable data, applied in a questionably appropriate way will most likely lead to a flood of litigation like none that has ever been witnessed.³⁵

Specifically, Baker suggests that the clustering of poor and minority students, will lead to adverse employment decisions against a higher ratio of minority teachers that serve those communities. As a result, "...black teachers of low-income black students will be several times more likely to be dismissed on the basis of poor value-added test scores."³⁶ Further, Baker foresees that this lack of non-random student assignment will create numerous court cases involving racially disparate impact teacher dismissals under Title VII of the Civil Rights Act.³⁷ Unless the state and schools can demonstrate that the score differences were not the result of race (i.e., not attributable to the result of non-random assignment of students), a court could find that PERA was not a neutral policy and therefore violates Title VII.

In Illinois, a recent decision related to educational malpractice may be instructive as to how future courts might view adverse teacher employment actions based on VAM or Growth

Models. The authors (one is a judge) of a recent Illinois State Bar Association article analyzing this 2012 appellate court case concluded that the tort of educational malpractice "...is not cognizable in Illinois."³⁸ The court in *Waugh*, noted the following:

Those courts that have refused to recognize claims of educational malpractice have done so based on various public policy grounds, including: 1) the lack of a satisfactory standard of care by which to evaluate an educator; 2) the inherent uncertainties about causation and the nature of damages in light of such intervening factors as students' attitude, motivation, temperament, past experience and home environment; 3) the potential for a flood of litigation against schools; and 4) the possibility that such claims will "embroil the courts into overseeing the day-to-day operations of the schools. *Alsides*, 592 N.W.2d at 472."³⁹

In particular, the appellate court's recognition of "inherent uncertainties about causation" would appear to be a theory plaintiffs' counsel should consider in an adverse employment decision action under PERA. Imagine the psychometric uncertainties an educational research expert could provide to a jury related to the use of a VAM as a "significant factor."

On the other hand, a recent law review article has suggested that the use of VAM may have the unintended consequence of building a pathway to educational malpractice as a viable tort.⁴⁰ As the authors Todd DeMitchell, Terri DeMitchell and Douglas Gagnon concluded,

...a pathway to educational malpractice may be being built through articulated standards, increased accountability, and now value-added measures of teacher effectiveness.... The courts already hold other professions responsible for the breach of their duties, which causes an injury. Why not education?⁴¹

This is an interesting theory. However, after considering the growing body of research regarding the unreliability for current VAMs, and the decision in *Waugh*, educational malpractice in Illinois (without advances in VAM psychometrics) appears to be a long-shot.

Conclusion

In conclusion, Illinois PERA contains a mixture of potential benefits and pitfalls for teachers, administrators, students, parents and the larger school community. On the one hand, Illinois' PERA presents a number of opportunities. The enhanced evaluator training (once it is fully de-bugged) presents an opportunity to inform evaluators and improve evaluations state-wide. Second, PERA establishes a state-wide framework for evaluation, and a common vocabulary for discussing and identifying best practices. Third, the flexibility afforded to local districts to consider and select the student assessments used for teacher evaluation will allow schools to customize their evaluation programs consistent with the law.

On the other hand, the use of student achievement growth data as a significant component of personnel decision-making demands further reflection and considerable refinement. At best, this will take time and further state investment to establish reliable methods and systems to support such an initiative. In the meantime, local districts will need to establish short-term strategies and reasonable evaluation solutions until more sophisticated data collection methods are available. Perhaps current and future district-university partnerships can develop more reliable methods for collecting and analyzing student achievement data related to teacher evaluation.

From a legal perspective, school districts proceed at their own risk by including student growth data as part of a negative employment decision. The law says districts are to consider student achievement data as at least 30% of the overall evaluation. And, districts should always follow the law. However, given the proven unreliability of using student growth data to determine teacher effectiveness, districts should be very cautious in making personnel rating decisions based on that particular input. In sum, growth models don't meet fundamental principles of scientifically based inquiry. Despite being promoted as central components of RTTT by USDE, it has been left to local districts to make the best of a flawed model. In the meantime, what happens when a teacher's position on the district Reduction in Force list is lowered (due to the requirements of SB7 and PERA), and that leads to her being honorably discharged in favor of a younger, less experienced and lower salaried teacher? See you at the courthouse!

It is difficult to imagine how USDE could adopt a reform that doesn't meet the "scientifically-based" standard set for instructional interventions under NCLB. Yes, it sounds good in theory—teachers should be held accountable for student learning. But if I may borrow Bill Clinton's line from the Democratic National Convention regarding the proposed Republican budget plan, the problem is arithmetic. Statistically, these growth models are not reliable or valid measures of teacher effectiveness. Yet, the state of Illinois, in adherence to the RTTT initiative now requires school districts to base personnel decision upon this unproven methodology.

As long as local administrators fully recognize the statistical and practical weaknesses associated with the student growth portion of the system, they can work collaboratively with their teachers and collective bargaining units to make the best of the situation. Perhaps VAMs can be refined in the years to come to solve the practical and statistical problems. Efforts to assess teacher effectiveness are difficult, but that shouldn't stop the education community from continuing to create methods that are reliable. In the meantime, the reaction as evidenced by the CTU-CPS strike indicates that future conflict on this issue is only in its infancy.

Promisingly, however, the conflict is also an opportunity for PK-12 educators and sympathetic legal counsel to illustrate through the courts and the court of public opinion the variables that impact student learning. A larger discussion needs to be held about the effects of poverty as well as the many school factors beyond the teacher's control that impact student learning. In fact by doing so, this may be the impetus by which meaningful discussions regarding adequate funding and social supports can be achieved. If policy-makers truly want to close the achievement gap, these factors must be eventually addressed. Perhaps RTTT's insistence on the use of VAMs has the potential to expand the discussion about what is needed to improve educational outcomes beyond unreliable measures of teacher effectiveness.

References

- ¹ The School Code (105 ILCS) was amended by changing Sections 2-3.25g, 24A-3, 24A-4, 24A-5, 24A-7, 24A-8, 24A-15, 34-8, and 34-85c and by adding Sections 24A-2.5, 24A-7.1, and 24A-20 (Jan. 15, 2010), <http://www.ilga.gov/legislation/publicacts/96/096-0861.htm>.
- ² 23 Ill. Admin. Code pt. 50, (2012), <http://www.isbe.state.il.us/rules/archive/pdfs/50ARK.pdf>.
- ³ 105 ILCS 5/24A-4b.
- ⁴ *Id.*
- ⁵ 105 ILCS 5/24A-7.
- ⁶ 23 Ill. Admin. Code pt. 110 a, (2012).
- ⁷ 105 ILCS 5/24-16.5.
- ⁸ 105 ILCS 5/24-12-b.
- ⁹ 105 ILCS 5/10-22.4.
- ¹⁰ 105 ILCS 5/21-23.
- ¹¹ 105 ILCS 5/24-11.
- ¹² 23 Ill. Admin. Code pt. 50, (2012).
- ¹³ 105 ILCS 5/24A-2.5.
- ¹⁴ *Comparison of CPS Proposals*, Chicago Tribune (Sept. 18, 2012), <http://www.chicagotribune.com/news/nationworld/chi-comparison-of-cps-proposals-20120917,0,892291.htmlpage>
- ¹⁵ 105 ILCS 5/24A-2.5.
- ¹⁶ Peter S. Goodman, *Chicago Teachers Strike a Push-Back to Education Reform*, Huffington Post (Sept 17, 2012), http://www.huffingtonpost.com/2012/09/15/chicago-teachers-strike_n_1886142.html.
- ¹⁷ 23 Ill. Admin. Code pt. 110 b, (2012).
- ¹⁸ *Rating Teachers: The Trouble with Value-Added Data*, Time, Andrew J. Rotherham, September 23, 2010, [www.time.com/time/nation/article/0.8599,2020867,00.html](http://www.time.com/time/nation/article/0,8599,2020867,00.html).
- ¹⁹ Glynn D. Ligon, *The Optimal Reference Guide: Comparison of Growth and Value-Add Models Growth Model Series – Part II*, Sept. 18, 2012, http://www.espsolutionsgroup.com/espweb/assets/files/ESP_Comparison_of_Growth_and_Value_Add_Models_ORG.pdf.
- ²⁰ 23 Ill Admin. Code pt. 50.30
- ²¹ Performance Evaluation Advisory Council, General FAQs about Teacher & Principal Evaluation (Sept. 27, 2012) , <http://www.isbe.state.il.us/peac/html/faqs.htm>.
- ²² Sean P. Corcoran, *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*. Education Policy of Action Series, (2010).
- ²³ Helen M. Hazi & Daisy Arredondo Rucinshki, *Teacher Evaluation as a Policy Target for Improved Student Learning: A Fifty-State Review of Statute and Regulatory Action since NCLB*. Education Policy Analysis Archives, Vol. 17 No. 5, 15, (March 10, 2009).
- ²⁴ Edward Haertel, et. al, *Comments on the Department of Education's Proposal on the Race to the Top Fund*, Board on Testing and Assessment Letter, (October 5, 2009).
- ²⁵ *Id.* at 8.
- ²⁶ *Id.*
- ²⁷ Chicagoland Researchers and Advocates for Transformative Education, *Misconceptions and Realities about Teacher and Principal Evaluation: Open Letter of Concern*, (March 26, 2012). www.ctunet.com/blog/text/OpenLetterCPSTeacherEvalSignersFinal.doc
- ²⁸ *Id.* at 2-3.
- ²⁹ Richard Rothstein, *How to Fix Our Schools: Its More Complicated, and More Work, Than the Klein-Rhee 'Manifesto' Wants You to Believe*, Economic Policy Institute Policy Brief 286, (October 14, 2010).
- ³⁰ *Id.*

³¹ Richard Rothstein, Helen F. Ladd, Diane Ravitch, Eva L. Baker, Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Robert L. Linn, Richard J. Shavelson, & Lorrie A. Shepard (2010). *Problems with the Use of Test Scores to Evaluate Teachers*, Economic Policy Institute, Briefing Paper #278, (August 27, 2010),

<http://www.epi.org/publication/bp278/>

³² *Id.* at 2.

³³ Linda Darling-Hammond, Audrey Amrein-Beadsley, Edward Haertel & Jesse Rothstein, *Evaluating Teacher Evaluation*, Kappan, 8, (March 2012).

³⁴ *Id.*

³⁵ *Id.* at 2.

³⁶ Bruce D. Baker, *Pondering Legal Implications of Value-Added Teacher Evaluation*, 1, (June 2, 2010),

<http://schoolfinance101.wordpress.com/2010/06/02/pondering-legal-implications-of-value-added-teacher-evaluation/>

³⁷ *Id.*

³⁸ Hon. James Fitzgerald Smith & Julia Illman Maness, *Does Your Claim Allege Educational Malpractice?*, 59 Trial Briefs 2, Aug. 2012,

<http://www.isba.org/sections/civilpractice/newsletter/2012/08/doesyourclaimallegeeducationalmalpr>

³⁹ *Waugh v. Morgan Stanley and Co., Inc.*, 2012 Ill App (1st) 102653, 359 Ill.Dec. 219, 966 N.E.2d, 540, (1st Dist. 2012).

⁴⁰ Todd A. DeMitchell, Terri A. DeMitchell & Douglas Gagnon, *Teacher Effectiveness and Value-Added Modeling: Building a Pathway to Educational Malpractice?*, 2012 BYU Educ. & L. J. 257 (2012).

⁴¹ *Id.* at 297.