



REVIEW OF *FOR GOOD MEASURE? TEACHER EVALUATION POLICY IN THE ESSA ERA*

Reviewed By

Amy N. Farley
University of Cincinnati

February 2017

Summary of Review

Although teacher evaluation reform has rapidly expanded across the U.S. in recent years, the newly enacted Every Student Succeeds Act (ESSA) seems to mark a federal deprioritization of teacher evaluation reform. This may in turn prompt states to consider changes to their evaluation systems. A new report from Bellwether Education Partners, *For Good Measure? Teacher Evaluation Policy in the ESSA Era*, argues for the maintenance of key elements of high-stakes teacher evaluation, including the use of student outcomes to evaluate teachers and a heavy focus on accountability. It urges policymakers to move slowly in revising their systems and to invest in management, capacity, and strategies to capture lessons learned. While the report raises several good questions with regard to the future of teacher evaluation, it overstates the likelihood that ESSA will result in widespread changes to evaluation systems, it ignores the literature regarding substantial technical challenges and unintended consequences of growth measures, and it dismisses the ideological and political debates surrounding teacher accountability. The unsubstantiated claims and dogged defense of student growth metrics provides little fresh or worthwhile new directions to policymakers seeking a nuanced and research-based discussion of teacher evaluation reform in the ESSA era.



Kevin Welner
NEPC Director

William Mathis
Managing Director

Alex Molnar
Publishing Director

National Education Policy Center

School of Education, University of Colorado
Boulder, CO 80309-0249
Telephone: (802) 383-0058

Email: NEPC@colorado.edu
<http://nepc.colorado.edu>

This material is provided free of cost to NEPC's readers, who may make non-commercial use of the material as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.

This review is one of a series made possible in part by funding from the Great Lakes Center for Education Research and Practice.



<http://www.greatlakescenter.org>
GreatLakesCenter@greatlakescenter.org

REVIEW OF *FOR GOOD MEASURE? TEACHER EVALUATION POLICY IN THE ESSA ERA*

Amy N. Farley, University of Cincinnati

I. Introduction

In the last several years, local schools, districts, and states have seen radical changes in K-12 teacher evaluation practices, particularly as it relates to the use of standardized observation protocols and student growth measures.¹ This focus on educator effectiveness can be traced in part to the 2001 authorization of the Elementary and Secondary Education Act (ESEA), known as No Child Left Behind (NCLB), which included a requirement to ensure that all teachers be “highly-qualified”.² This intensified in 2009 as a result of the \$4.35 billion federal Race to the Top (RTTT) grant competition, designed to “trail-blaze effective reforms” in U.S. states.³ In order to be competitive for RTTT funds, states made assurances regarding four policy reform agendas, including (a) standards and assessments, (b) data systems, (c) school-turnaround efforts, and (d) teacher and principal evaluation linked to student performance.⁴

Spurred by NCLB and RTTT, teacher evaluation reform is now widespread: 40 states submitted RTTT applications with proposed strategies for performance-based teacher evaluation⁵ and 28 states enacted legislation *requiring* that teacher evaluation use student achievement measures between 2009 and 2015.⁶ There is also a significant and growing body of research and scholarship related to teacher evaluation systems in general and the use of student growth linked to teachers in particular.

The 2016 reauthorization of ESEA, known as the Every Student Succeeds Act (ESSA), seems to mark a federal deprioritization of teacher evaluation reform. Compared to NCLB, ESSA grants states more flexibility with regard to state accountability plans in general and also removes requirements from RTTT that states evaluate teachers on the basis of student outcomes.⁷ Given this flexibility, states and districts are left to ponder the best path forward and whether that still includes high-stakes teacher evaluation systems linked to student achievement. A new report from Bellwether Education Partners – *For Good Measure? Teacher Evaluation Policy in the ESSA Era*,⁸ written by Kaitlin Pennington and Sara Mead – takes up these very questions. The report seeks to inform revisions to teacher evaluation policies by “reviewing the evolution of the teacher evaluation policy movement over the last several years, identifying positive outcomes of new systems and negative consequences, and describing risks that should be considered in a post-ESSA world.”⁹ The report concludes with policy recommendations for states and districts.

II. Findings and Conclusions of the Report

The report has three primary substantive sections. The first provides an introduction, historical summary of teacher evaluation reform, and a review of the primary arguments used in its favor and opposition. The second section identifies six potential “risks” associated with the flexibility provided under ESSA regarding teacher evaluation. Finally, the report concludes by offering five policy recommendations.

The introductory section begins with a brief historical overview of teacher evaluation reform in the United States, referencing the oft-cited literature that teachers are the “greatest in-school factor affecting student achievement”¹⁰ and that traditional teacher evaluation systems do not meaningfully differentiate between teachers. Embedded within the introductory section, the report also highlights two case studies of teacher evaluation reform. The first is a review of IMPACT, one of the earliest teacher evaluation systems adopted in the District of Columbia Public Schools’ (DCPS) in 2009. The second summarizes the teacher evaluation system developed by Achievement First, a charter management organization with a network of 32 charter schools.

The report then seeks to review the major lessons learned from teacher evaluation reform, summarizing both the so-called good and bad outcomes. The report identifies several consequences of teacher evaluation reform it considers to be positive, including (a) more teachers engaging in *meaningful conversations about their practice*; (b) *increased collaboration and communication* between teachers and school leaders; (c) preliminary evidence of *improved teacher quality*, based on preliminary research conducted in D.C. Public schools and Chicago; (d) *increased differentiation* between teachers’ performance, particularly the identification of highly effective teachers; and (e) *expanded data systems* that allow schools, districts, and states to link teacher and student data. The report also acknowledges several areas of improvement with regard to teacher evaluation reform: (a) the *speed and timeline* of reform adoption, which has left “stakeholders unhappy and important tenets of the reforms vulnerable to attack;”¹¹ (b) the *persistent perception among teachers* that reformed evaluation systems are “a mechanism to harm them,”¹² although the report suggests these fears are overstated; (c) *teachers’ resistance to the use of student achievement* in evaluation systems generally, which the report attributes to mistrust of state achievement tests and an ideological disagreement about whether results are a meaningful reflection of teacher quality; and (d) an *inadvertent reduction in opportunities for educational innovation* as a result of rigid teacher evaluation policies, particularly with regard to blended and personalized learning models.

Following the introductory section of the report, the second section summarizes the possible risks inherent in the “post-ESSA world.”¹³ The report identifies six potential risks for states and policymakers to consider when revising evaluation systems. First, it warns not to shift the focus of teacher evaluation systems away from human resource decisions – that is, employment decisions like hiring and firing of teachers – to professional growth and development.¹⁴ Second, the report urges against the elimination of student achievement measures, claiming they are better measures of teacher quality than other available measures and that

their loss could impact the ability to “identify effective teaching or ensure equitable access to quality teaching for underserved students.”¹⁵ Third, the report claims that it is possible under ESSA that neither schools nor teachers will be held accountable for student outcomes. This could “significantly reduce urgency to improve outcomes for chronically underserved students and subgroups—ultimately leading to increased educational inequity and backsliding on recent progress narrowing achievement gaps.”¹⁶ Fourth, the report cites concerns regarding the capacity of local agencies, including districts and states, to implement evaluation systems without sufficient guidance. They report that although the reformed systems did include flexibility and encourage locally designed systems, many districts have limited capacity, not to mention insufficient expertise. They argue that while these capacity issues are pre-existing, they are likely to be exacerbated by ESSA’s flexibility regarding evaluation systems. Fifth, the report highlights equity risks and challenges in sharing lessons across states, argues that because ESSA may increase variation, lower capacity districts may be unable to implement so-called “strong reforms.”¹⁷ Finally, the report reminds policymakers of the risks inherent in ignoring the larger human capital ecosystem, claiming that teacher evaluation reform must be linked to a “more comprehensive rethinking of human capital systems.”¹⁸ Specifically, they cite concerns that evaluation be connected to reforms related to teacher preparation and pipeline issues, teacher retention and leadership, and school leadership capacity particularly as it relates to the implementation of evaluation systems for teachers.

The report concludes with post-ESSA recommendations for state policymakers. This section begins by acknowledging the wide variability in state contexts, which makes the authors “reluctant to offer prescriptive recommendations.”¹⁹ Nevertheless they go on to advance the following five policy recommendations to help policymakers avoid mistakes, and ostensibly to minimize the risks outlined above:

1. Don’t rush to action.

The report argues that the rapid adoption of teacher evaluation reform may have, in fact, contributed to implementation problems, and that policymakers should avoid compounding this problem by quickly revising their evaluation systems again. They also point out that under ESSA states are likely making large policy changes to their school accountability systems and that significant reform to teacher evaluation systems could create confusing interactions between two major program overhauls. Instead, they recommend states “wait until the dust is settled on accountability and other issues.”²⁰

2. Preserve a role for student achievement in teacher evaluation systems.

To some extent, the report takes its strongest stance here by clearly calling it a mistake to eliminate student achievement measures from teacher accountability systems. The report cites many of the concerns outlined above, and also argues that student

achievement measures are critical in sending a message that student learning is fundamental to quality teaching.

3. Consider the relationship between teacher evaluation and accountability systems.

The report once again points out the considerable flexibility regarding school accountability under ESSA and encourages state policymakers to consider the dependencies between school accountability and teacher evaluation systems as they make changes to both. They recommend aligning the metrics used in school and teacher accountability systems to avoid “creating conflicting or misaligned incentives for different actors in the system,”²¹ which essentially reiterates the argument for the inclusion of student achievement in teacher evaluation. Finally, the report urges states to ensure the availability of high quality, valid, and reliable assessments.

4. Invest in management and capacity to develop teachers.

The report argues that developing school leader capacity is critical to improving teacher quality, regardless of status of teacher evaluation reform. They further argue that increased flexibility from the state and federal systems regarding accountability and interventions for low-performing schools increase the need for strong leadership.

5. Identify strategies to capture and learn from variation.

Finally, based on their assumption that teacher evaluation policies will become more diverse and heterogeneous under ESSA, the report reminds policymakers that this will create learning opportunities that must be captured, suggesting the need for more consistent evaluation and evidence regarding teacher evaluation systems and their consequences on teachers, students, and the ecosystems of schools and districts at large.

III. The Report’s Rationale for Its Findings and Conclusions

At its core, the report seeks to mitigate alleged risks associated with the increased flexibility in ESSA – flexibility the report claims will lead to vast changes in teacher evaluation policies across the U.S. and will roll back some of the most controversial components of recently legislated teacher evaluation reform (e.g., the use of student outcomes as a measure of teacher effectiveness). The report cautions that “state policymakers may retreat from these policies without learning from the successes and shortcomings,”²² a decision they suggest may be premature and even dangerous.

While the report acknowledges that the speed of adoption of teacher evaluation reforms has far outpaced readiness, including technical expertise and district/state capacity, it also expresses concerns that “just as the rapid passage of teacher evaluation policies created risks five years ago, the movement away from them creates risks today.”²³ However, the risks identified are not clearly drawn from the research literature or even a review of the theoretical foundations of teacher evaluation policy; instead, they appear to represent the worst fears of teacher evaluation advocates.

IV. The Report’s Use of Research Literature

The report does not rely on a robust review of academic or peer-reviewed literature. Within the 64 endnotes included in the report, which include 77 total citations,²⁴ only five come from peer-reviewed journals. Furthermore, the report almost entirely omits the body of literature regarding teacher evaluation reform or value-added models in general and the particular challenges of attributing student growth to educators. This is particularly unfortunate because the literature is quite robust: In the 2015 edition of the *Handbook for Research on Teaching* published by the American Educational Research Association (AERA), the chapter on “Teacher Evaluation in American Schools”²⁵ includes over 200 unique references, derived almost entirely from peer-reviewed research.

Table 1: Citations and sources included in *For Good Measure?*

Citation type	Count	Selection of most-frequently cited examples (<i>n</i>)
Popular press	26	Chalkbeat (7); Education Week (7); NY Times (3); NJ Spotlight (2)
Advocacy organization/think tank	20	National Council on Teacher Quality (9); The New Teacher Project (3); Data Quality Campaign (2)
Bellwether Education Partners	8	Bellwether Education Partners (8)
Other classifications	8	Southern Regional Education Board (2), District Reform Support Network, Council of Chief State School Officers
Peer-reviewed journal	5	Journal of Policy Analysis and Management (2), Journal of Human Resources, Econometrica, Education Next
Research center or organization	5	American Institutes for Research, Gallup, RAND corporation
Government agency	4	U.S. Department of Education (3), Indiana Department of Education
Foundation	1	Bill & Melinda Gates Foundation

Instead of summarizing the research literature, 46 of the report’s 77 references hail from the popular press or advocacy-oriented organizations. The popular press is referenced more than any other source (26 times), with the majority of references coming from *Chalk-*

beat and *Education Week* (seven times apiece). Advocacy organizations and/or think tanks are referenced an additional 20 times; among these, the report cites the National Council on Teacher Quality (NCTQ) nine times,²⁶ more than any other single source. This reliance on NCTQ is noteworthy given that it advocates aggressively for teacher evaluation reform and the inclusion of student growth measures in the evaluation of teachers.²⁷ The report also cites a handful of other advocacy organizations, including The New Teacher Project and the Data Quality Campaign.

The underutilization of peer-reviewed literature is particularly problematic in this report given its supposed reliance on research. Throughout the report, the narrative misleadingly claims that the “research shows” various trends, and then the text either provides no citations at all²⁸ or links to only one or two sources, the quality of which is highly variable. Sometimes these assertions do reflect the tenor of the research and other times they do not, but by including only limited citations, the report authors ensure that policymakers or others reviewing the report are unable to independently validate their claims.

V. Review of the Report’s Methods

As is described in detail above, the report provides little explanation regarding the methods used to arrive at its conclusions. While the report raises several good questions with regard to the future of teacher evaluation – particularly as it relates to district and state capacity and the critical importance of maintaining a research and learning agenda related to the systems that evolve – the majority of the conclusions drawn and policy recommendations that follow remain relatively underdeveloped or unsubstantiated.

The inclusion of two case studies regarding real-life teacher evaluation systems also raises methodological questions. Although the report never explicitly says so, both appear to be elevated as exemplary programs. Nevertheless, the contexts and program designs differ in important ways from typical evaluation systems in public schools across the country, making the replicability of their successes questionable: Achievement First is a charter management organization, meaning that it operates without many of the regulations that exist for traditional public schools, and DC Public Schools’ program is a hybrid that blends elements of performance pay with traditional teacher evaluation).²⁹

Furthermore, the report misuses the research literature in at least one instance, raising serious questions about the veracity and reliability of its conclusions. In the section outlining post-ESSA risks, the report warns against eliminating student achievement measures, arguing it could lead to a more expensive evaluation system. The report cites research from three districts regarding evaluation-focused expenditures, noting that in Memphis City Schools “the activities to produce student growth measures connected to teacher performance accounted for just one percent.”³⁰ While this is technically an accurate description

of the study's findings for Memphis City Schools, it dramatically misrepresents the results across all three districts from the original research.³¹ In fact, the percentage of the evaluation system expenditures related to value-added models in the other two districts, Hillsborough County Public Schools and Pittsburgh Public Schools, were 13% and 45% respectively (see Figure 1).³²

Exhibit B. Overview of Expenditures on the Evaluation Systems in HCPS, MCS and PPS between November 2009 and June 2012

	Hillsborough County Public Schools (HCPS)	Memphis City Schools (MCS)	Pittsburgh Public Schools (PPS)
Total evaluation system expenditures	\$24.8 million	\$8.5 million	\$6.4 million
Percentage of evaluation system expenditure by component			
Teacher observations	87%	82%	47%
Value-added model (VAM)	13%	1%	45%
Student surveys	Not Applicable	17%	8%
Funding sources			
Philanthropic funds	62%	94%	58%
Federal funding	19%	6%	27%
District funding	19%		8%
Mixed funding			7%

Source: Author calculations based on effective teaching initiative expenditure figures from IPS grantee financial reports.

Figure 1: Original AIR results regarding expenditures related to evaluation systems.
Source: Chambers, J., Brodziak de los Reyes, I. & O'Neil, C. (2013). RAND/AIR

This kind of selective reporting of data is especially concerning given that the original report explicitly acknowledges that the expenditures were lower in Memphis than the other two sites because Memphis relied on an existing state model (i.e., the Tennessee Value-Added Assessment System) that had been in place for several years.

VI. Review of the Validity of the Findings and Conclusions

The report's conclusions are weakened by three significant limitations: First, the report overstates the likelihood that the relaxed ESSA requirements will quickly result in a complete overhaul of teacher evaluation systems; second, the argument in favor of a preserved role for student growth in teacher evaluation completely ignores the robust literature regarding technical challenges and potential unintended consequences; and, third, the report brazenly dismisses the ideological and political debates surrounding teacher accountability when those debates are not yet settled. Below, each of these limitations is discussed at greater length.

Overstated concern regarding post-ESSA evaluation reform.

As described above, the report seeks to identify the inherent risks for teacher evaluation reform in the post-ESSA policy landscape. However, it exaggerates the likelihood that states will radically alter their existing evaluation systems. Although the report acknowledges that “ESSA requires states to have a definition of teacher ineffectiveness,” it also states several times that states are no longer “required to implement teacher and principal evaluation systems.”³³ While the new ESSA regulations may not specifically require evaluation systems, it is a bit of a misrepresentation to claim that states are no longer required to implement systems of evaluation at all. A large number of states are still beholden to state and local legislation that does, in fact, require performance-based teacher evaluation, often linked to student achievement; while the passage of ESSA may prompt states to revisit that legislation, it does not immediately nullify those legal requirements. Furthermore, the report leads the reader to believe that ESSA has almost completely abandoned teacher accountability, when in fact significant oversight remains regarding the equitable distribution of effective teachers to low-income and minority students, a requirement that inherently requires states to determine educator effectiveness in some manner.

Here in Ohio, the draft ESSA plan, released on January 19, 2017, acknowledges the ubiquitous concerns from educators and other stakeholders regarding the use of student growth measures in teacher evaluation.³⁴ However, Ohio’s Department of Education appears poised to stay the course, at least in the short term, citing current state law and the existing equity plan as justification for those decisions. And while Ohio’s draft plan does contend that the “freedom from prescriptive mandates gives the state the opportunity to reflect on our current system, have a larger conversation with stakeholders, and consider how we want to define effectiveness moving forward,” Ohio is still a considerable time away from abandoning the current system wholesale.³⁵

Underdeveloped discussion of student growth measures.

At its outset, the report dismisses universally some of the most important lingering questions about teacher evaluation, the most significant of which is whether teachers should be held singularly accountable for student outcomes. In fact, educational researchers and scholars largely agree that the technical requirements for the value-added models in teacher evaluation are substantial and the unintended consequences of using them inappropriately are vast: In 2014 and 2015 respectively, the American Statistical Association (ASA)³⁶ and the American Educational Research Association (AERA)³⁷ issued formal position statements cautioning states and policymakers about the potential negative consequences from the high-stakes application of VAMs. AERA’s statement expresses both technical concerns and overarching concerns about their use:

Although VAM may be superior to status models, it does not mean that they are ready for use in educator or program evaluation. There are potentially serious negative consequences in the context of evaluation that can result from

the use of VAM based on incomplete or flawed data, as well as from the misinterpretation or misuse of the VAM results. Teachers and leaders, for example, with low VAM scores can experience loss of advancement, lost compensation, and even termination. Also, when large numbers of teachers and leaders are misidentified, then resources may be misdirected, and the educational system as a whole can be degraded.³⁸

As highlighted in the excerpt above, VAMs are imperfect empirical estimates that are vulnerable to distortion and corruption.³⁹ Furthermore, questions remain about the reliability and stability of these measures over time and the most appropriate methodological approach to consider outside factors (e.g., student demographic characteristics, classroom/school characteristics).⁴⁰

Despite these well-documented questions in the literature, the Bellwether report almost entirely eschews a robust discussion of the technical limitations of using student outcomes in teacher evaluation. Instead, throughout the report, the reader is led to believe that any policy changes that reduce the role of student outcomes is politically and/or ideologically motivated, rather than rooted in the real technical challenges. Consider, for example, the case of New York state. According to the report, “union leaders’ efforts to undermine evaluations and tie them to an anti-testing backlash” devastated the new evaluation system in New York, resulting in “successive years of legislation undoing previous reforms.”⁴¹ One example cited in the report was legislation from 2013-14 that prohibited the use of new, Common Core-aligned assessments in the nascent teacher evaluation system. While the report interprets this change as an attempt to weaken New York’s system, this action was just as likely grounded in best practice regarding assessment and growth models: In fact, AERA’s formal position explicitly states that growth estimates “should generally not be employed across [assessment system] transitions.”⁴² Nevertheless, the report concludes that “national teachers’ unions have played a role in fostering anti-testing sentiment, which fueled an opt-out movement and created both technical and political challenges for performance-based teacher evaluations.”⁴³ While the union may have highlighted technical challenges in support of their agenda, it is unfair to characterize these well-documented challenges as a union creation.

Dismissive toward political and ideological debates.

Lastly, even setting the technical concerns aside, the report argues that “the debate and conflict over teacher evaluation continues to focus on issues of politics and ideology,”⁴⁴ leading the reader to believe that any ideological or political discussion is misplaced. In fact, the political and ideological debates about high-stakes teacher evaluation remain largely unresolved. There is little evidence to suggest that high-stakes accountability has increased educational equity or narrowed achievement gaps.⁴⁵

VII. Usefulness of the Report for Guidance of Policy and Practice

Although the report seeks to provide useful guidance and recommendations for policymakers, its reliance on faulty logic and incomplete analyses of the existing literature prevent it from accomplishing those goals. In particular, the selective omission of critical content – including cautions regarding the use of student growth measures in teacher evaluation put forth by ASA and AERA – presents an incomplete picture of the issues at hand. While there may be a continued role for student outcomes in teacher evaluation, best practice demands that states and districts attend to the technical requirements necessary to produce valid estimates.

For example, in their statement, AERA specifies eight technical requirements that must be met to consider the use of VAM scores.⁴⁶ They explicitly argue that any use of VAM that fails to meet all eight of their requirements should *preclude its use entirely*. Here they are quite clear: “When there is credible evidence that there are negative consequences, every effort should be made to mitigate them.”⁴⁷ Had the report addressed these concerns and provided practical solutions for districts and states to move forward in this context, their recommendations would be much more defensible.

However, as the report currently stands, it is imbalanced, particularly in its dogged defense of the student growth metrics in teacher evaluation systems. While the report criticizes the teacher evaluation debates as being overly ideological, its own recommendations appear to rely almost exclusively on ideological commitments rather than on research. The report reveals these commitments when it acknowledges that “[w]hile it’s unlikely that all states will eliminate the role of student growth in their evaluation policies, even a few states reversing course will make it difficult for advocates to protect measures of student achievement in teacher evaluation policy.”⁴⁸ It seems that this drive to protect reformed evaluation systems is really at the heart of the report’s analyses and subsequent recommendations.

Notes and References

- 1 Rowan, B. & Raudenbush, S.W. (2016) Teacher evaluation in American schools. In D.H. Gitomer & C.A. Bell (Eds.), *Handbook of research on teaching* (fifth ed.), (pp. 1159-1216). Washington, D.C., American Education Research Association.
- 2 No Child Left Behind (NCLB) Act of 2001, Public Law No. 107-110 (2002). Washington, D.C.: United States Congress. Retrieved January 20, 2017, from <https://www.congress.gov/bill/107th-congress/house-bill/1/text?overview=closed&r=31>
- 3 U.S. Department of Education. (n.d.). *Race to the Top fund, program description*. Washington, D.C.: Author. Retrieved January 20, 2017, from <http://www2.ed.gov/programs/racetothetop/index.html>
- 4 U.S. Department of Education. (2009, November). *Race to the Top program executive summary*. Washington, D.C.: Author. Retrieved January 20, 2017, from <https://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- 5 Learning Point Associates. (2010, May). *Evaluating teacher effectiveness: Emerging trends reflected in the state phase 1 Race to the Top applications*. Naperville, IL: Author. Retrieved January 20, 2017, from <http://files.eric.ed.gov/fulltext/ED509942.pdf>
- 6 Dougherty, K. M. & Jacobs, S. (2015, November). *State of the states 2015: Evaluating, teaching, leading, and learning*. Washington, D.C.: National Council on Teacher Equality, as cited in Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 7 Every Student Succeeds Act. (Public Law No: 114-95). (2015). Washington, D.C.: United States Congress. Retrieved January 2, 2017, from <https://www.congress.gov/bill/114th-congress/senate-bill/1177/text>
- 8 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 9 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 3. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 10 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 8. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 11 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 16. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 12 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 16. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>

- 13 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 3. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 14 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 20. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 15 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 21. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 16 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 23. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 17 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 24. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 18 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 24. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 19 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 28. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 20 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 29. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 21 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 29. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 22 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 7. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 23 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 7. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 24 Each of these 77 citations is not unique. Quite the contrary, several resources are referenced repeatedly throughout the document.
- 25 Rowan, B. & Raudenbush, S.W. (2016) Teacher evaluation in American schools. In D.H. Gitomer & C.A. Bell (Eds.), *Handbook of research on teaching* (fifth ed.), (pp. 1159-1216). Washington, D.C., American Education Research Association.
- 26 It is worth noting that this includes seven unique references to NCTQ's 2015 *State of the States* publication:

Dougherty, K.M. & Jacobs, S. (2015, November). *State of the states 2015: Evaluating, teaching, leading, and learning*. Washington, D.C.: National Council on Teacher Quality. Retrieved January 25, 2017, from <http://www.nctq.org/dmsView/StateofStates2015>

- 27 In fact, NCTQ's formal state policy position regarding the evaluation of teacher effectiveness asserts that "the state should either require a common evaluation instrument in which evidence of student learning is the most significant criterion or should specifically require that student learning be the preponderant criterion in local evaluation processes" (<http://www.nctq.org/statePolicy/2015/approach/goals/>).
- 28 The two references to the research literature without any accompanying citations include: Recommendations to keep student achievement in evaluation systems but not use them mathematically toward a final rating "... fall short, however, in grounding the teaching profession in outcomes-based measures that research shows are the best predictors of teacher effectiveness" (p. 22) and "Research shows that school leaders have tremendous influence on teacher quality through the hiring decisions they make, expectations they set, the feedback they provide to teachers, and the school culture they create" (p. 30).
- 29 IMPACT provides teachers with both positive and negative "high-powered incentives," including "the threat of dismissal for low-performing teachers as well as substantially larger financial incentives for high-performing teachers", a fact acknowledged only in passing in the report. The addition of significant financial incentives (i.e., bonuses upwards of \$25,000 for teachers of tested grades and subjects in high-poverty schools) to the DCPS IMPACT program make it in some ways distinct from many teacher evaluation reforms across the country, a point that was not explicit in the report. Citation: Dee, T. S. & Wyckoff, J. (2015, Spring). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of policy analysis and management*, 34(2), 267-297.
- 30 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 21. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 31 Chambers, J., Brodziak de los Reyes, I. & O'Neil, C. (2013, May). *How much are districts spending to implement teacher evaluation systems? Case studies of Hillsborough County Public Schools, Memphis City Schools, and Pittsburgh Public Schools*. Washington, D.C.: RAND Education; Washington, D.C.: American Institutes for Research. Retrieved January 25, 2017, from http://www.air.org/sites/default/files/downloads/report/What_Districts_Are_Spending_to_Implement_Teacher_Evaluation_Systems_Final_o.pdf
- 32 It is also worth noting here that these figures do not represent the total cost of calculating growth measures because they exclude entirely the existing costs of administering high-stakes assessment systems to students.
- 33 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 10. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 34 Ohio Department of Education. (2017, January). *Driving education excellence: Securing the future for all Ohio students: A draft overview of Ohio's state plan for The Every Student Succeeds (ESSA) Act*. Columbus, Ohio: Author. Retrieved January 25, 2017 from http://education.ohio.gov/getattachment/Topics/Every-Student-Succeeds-Act-ESSA/ESSA_Framework_Report.pdf.aspx
- 35 Ohio Department of Education. (2017, January). *Driving education excellence: Securing the future for all Ohio students: A draft overview of Ohio's state plan for The Every Student Succeeds (ESSA) Act*. Columbus,

Ohio: Author. Retrieved January 25, 2017, from http://education.ohio.gov/getattachment/Topics/Every-Student-Succeeds-Act-ESSA/ESSA_Framework_Report.pdf.aspx

- 36 ASA's position outlines the many technical challenges inherent in creating robust VAM estimates, and claims that the full impact of VAMs on education remains relatively unknown, particularly with regard to their impact on teachers, principals, and other educators. While ASA doesn't advocate abandoning the use of VAMs entirely, it does urge caution in their interpretation and application to high-stakes uses. For the full statement, see: American Statistical Association. (2014, April). *ASA statement on using value-added models for educational assessment*. Alexandria, VA: Author. Retrieved January 25, 2017, from <http://www.amstat.org/asa/files/pdfs/POL-ASAVAM-Statement.pdf>
- 37 American Educational Research Association Council. (2015, November). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448-452. Retrieved January 2, 2017, from <http://journals.sagepub.com/doi/full/10.3102/0013189X15618385>
- 38 American Educational Research Association Council. (2015, November). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448-452, p. 449. Retrieved January 2, 2017, from <http://journals.sagepub.com/doi/full/10.3102/0013189X15618385>
- 39 Rowan, B. & Raudenbush, S.W. (2016) Teacher evaluation in American schools. In D.H. Gitomer & C.A. Bell (Eds.), *Handbook of research on teaching* (fifth ed.), (pp. 1159-1216). Washington, D.C., American Education Research Association.
- 40 Rowan, B. & Raudenbush, S.W. (2016) Teacher evaluation in American schools. In D.H. Gitomer & C.A. Bell (Eds.), *Handbook of research on teaching* (fifth ed.), (pp. 1159-1216). Washington, D.C., American Education Research Association.
- 41 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 10. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 42 American Educational Research Association Council. (2015, November). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448-452, p. 450. Retrieved January 2, 2017, from <http://journals.sagepub.com/doi/full/10.3102/0013189X15618385>
- 43 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 10. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 44 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 6. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 45 For a more detailed discussion of these issues, see Berry, B. (2013). Good schools and teachers for all students: Dispelling myths, facing evidence, and pursuing the right strategies. In P.L. Carter & K.G. Welner (Eds.), *Closing the opportunity gap: What America must do to give every child an even chance*. New York: Oxford University Press;

Zhao, Y. & Tienken, C. (2013). How common standards and standardized testing widen the opportunity gap. In P.L. Carter & K.G. Welner (Eds.), *Closing the opportunity gap: What America must do to give every child an even chance*. New York: Oxford University Press.

- 46 Those requirements range from ensuring that the assessments used to create VAM estimates meet professional standards of reliability and validity, to requiring that all VAM models utilize multiple years of data from tests that are comparable over time (e.g., not disrupted by major shifts in assessment systems), to requiring that all VAM models include statistical estimates of error. They also caution against singularly using VAM estimates to make decisions and argue for ongoing monitoring of both intended and unintended consequences. American Educational Research Association Council. (2015, November). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448-452. Retrieved January 2, 2017, from <http://journals.sagepub.com/doi/full/10.3102/0013189X15618385>
- 47 American Educational Research Association Council. (2015, November). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448-452, p. 450. Retrieved January 2, 2017, from <http://journals.sagepub.com/doi/full/10.3102/0013189X15618385>
- 48 Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners, p. 10. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>

DOCUMENT REVIEWED: For Good Measure? Teacher Evaluation Policy in the ESSA Era

AUTHORS: Kaitlin Pennington & Sara Mead

PUBLISHER/THINK TANK: Bellwether Education Partners

DOCUMENT RELEASE DATE: December 5, 2016

REVIEW DATE: February 21, 2017

REVIEWER: Amy N. Farley, University of Cincinnati

E-MAIL ADDRESS: amy.farley@uc.edu

PHONE NUMBER: (513) 556-5111

SUGGESTED CITATION:

Malen, B. & Rice, J.K. (2017). *Review of “For Good Measure? Teacher Evaluation Policy in the ESSA Era.”* Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-teacher-evaluation>