

**SPECIAL ISSUE**  
**Policies and Practices of Promise**  
**in Teacher Evaluation**

education policy analysis  
archives

A peer-reviewed, independent,  
open access, multilingual journal



Arizona State University

Volume 28 Number 55

April 13, 2020

ISSN 1068-2341

**Policies and Practices of Promise in Teacher Evaluation**

*Audrey Amrein-Beardsley*  
Arizona State University  
United States

**Citation:** Amrein-Beardsley, A. (2020). Policies and practices of promise in teacher evaluation: The introduction to the special issue. *Education Policy Analysis Archives*, 28(55).  
<https://doi.org/10.14507/epaa.28.5443> This article is part of the special issue, *Policies and Practices of Promise in Teacher Evaluation*, guest edited by Audrey Amrein-Beardsley.

**Abstract:** This introduction to the special issue on “Policies and Practices of Promise in Teacher Evaluation,” (1) presents the background and policy context surrounding the ongoing changes in U.S. states’ teacher evaluation systems (e.g., the decreased use of value-added models (VAMs) for teacher accountability purposes); (2) summarizes the two commentaries and seven research papers that were peer-reviewed and ultimately selected for inclusion in this special issue; and (3) discuss the relevance of these pieces in terms of each paper’s contribution to the general research on this topic and potential to inform educational policy, for the better, after the federal government’s passage of the Every Student Succeeds Act (ESSA, 2016).

**Keywords:** education policy; teacher evaluation; teacher accountability; ESSA

**Políticas y prácticas de promesa en la evaluación docente**

**Resumen:** Esta introducción a la número especial sobre “Políticas y prácticas de promesa en la evaluación docente” (1) presenta los antecedentes y el contexto político que rodea los cambios en curso en los sistemas de evaluación docente de los estados de EE. UU. (e.g., la disminución del uso de valor agregado modelos [VAMs] para propósitos de rendición de

Journal website: <http://epaa.asu.edu/ojs/>

Facebook: /EPAAA

Twitter: @epaa\_aape

Manuscript received: 3/29/2020

Revisions received: 3/30/2020

Accepted: 3/30/2020

cuentas del maestro); (2) resume los dos comentarios y siete documentos de investigación que fueron revisados por pares y finalmente seleccionados para su inclusión en este número especial; y (3) analizar la relevancia de estas piezas en términos de la contribución de cada artículo a la investigación general sobre este tema y el potencial para informar la política educativa, para mejor, después de la aprobación del gobierno federal de la ley, *Every Student Succeeds Act* (ESSA, 2016).

**Palabras-clave:** política educativa; evaluación docente; rendición de cuentas del maestro; ESSA

### **Políticas e práticas de promessa na avaliação de professores**

**Resumo:** Esta introdução à dossier sobre “Políticas e práticas de promessa na avaliação de professores” (1) apresenta o contexto e o contexto político em torno das mudanças em andamento nos sistemas de prestação de contas de professores dos estados dos EUA (por exemplo, a diminuição do uso de valor agregado modelos [VAMs] para propósitos de responsabilização de professores); (2) resume os dois comentários e sete trabalhos de pesquisa que foram revisados por pares e, finalmente, selecionados para inclusão nesta dossier; e (3) discutir a relevância dessas peças em termos da contribuição de cada artigo para a pesquisa geral sobre esse tópico e o potencial de informar as políticas educacionais, para melhor, após a aprovação pelo governo federal da lei, *Every Student Succeeds Act* (ESSA, 2016).

**Palavras-chave:** políticas educativas; avaliação de professores; prestação de contas do professor; ESSA

## **Introduction<sup>1</sup>**

In January of 2016, former U.S. President Obama signed into law the Every Student Succeeds Act (ESSA). The primary intent of ESSA (2016) was to restore local control to states, reduce the federal government’s regulation over states, and reset the federal government’s relationship with the nation’s 100,000 public schools, its nearly 50 million public school students, and its approximately 3.4 million public school teachers. ESSA (2016) was to also replace the current national accountability policy scheme as primarily based on high-stakes tests, with state-led accountability systems, while returning back to states responsibility for measuring student, teacher, and school performance. While states are still required to test students annually in mathematics and reading in grades three through eight and once in high school, as per the provisions written into No Child Left Behind (NCLB, 2001), and states are to report on these indicators by race, income, ethnicity, disability, etc., specific to this special issue, ESSA (2016) allowed states to decide whether and how to evaluate teachers with or without using or accounting for teachers’ purportedly causal effects on students’ standardized test scores over time, for example, via the use of student growth models (SGMs), more generally, and value-added models (VAMs), more specifically.

While ESSA (2016) is not without controversy (e.g., states’ uses of SGMs and VAMs are still strongly encouraged by the federal government as written into ESSA [2016]), pertinent to this special issue on “Policies and Practices of Promise in Teacher Evaluation,” is that states can now decide how and to what extent states might (or might not) value or explicitly weight students’ test

---

<sup>1</sup> Please also note that a prior version of this piece was published by the National Education Policy Center (NEPC; see Close, Amrein-Beardsley, & Collins, 2008). The introduction to this special issue was edited by the consulting and managing editors.

scores as components of their revised teacher evaluation policies and systems. In addition, states now have more freedom to implement teacher evaluation systems that might involve other evaluation indicators and measures (e.g., student surveys), might serve more formative (i.e., developmental) versus summative (i.e., outcomes-based) ends, and might permit more innovation, for example, when developing new or working to improve longstanding evaluation measures (i.e., observational systems) formerly dismissed as being (too) subjective (see, for example, Weisberg, Sexton, Mulhern, & Keeling, 2009; see also Kraft & Gilmour, 2017).

Hence, and now four years post-ESSA (2016), perhaps unsurprisingly, states' educational policies, systems, and practices surrounding teacher evaluation are changing, or are beginning to change (Close, Amrein-Beardsley, & Collins, 2020; Ross & Walsh, 2019). Again, this is occurring because ESSA (2016) has allowed states to recover power and authorities over these areas. Accordingly, it is the purpose of this special issue to capture how the teacher evaluation situation is, indeed, changing and ideally changing for the better post-ESSA. Changing for the better is defined herein as aligning with the theoretical and empirical research that is currently available in the literature base surrounding contemporary teacher evaluation systems, as well as the theoretical and empirical research that is presented in this special issue.

Correspondingly, via this special issue I have brought together scholars researching, implementing, and assessing such changes and innovations in teacher evaluation policies, systems, and practices, all of whom are doing this important work while drawing upon diverse theoretical, methodological, and conceptual perspectives. Again, contributors to this special issue are beginning to shed light on policies and practices of promise, throughout the US but also with implications for nations beyond in which leaders might also be grappling with similar issues related to evaluating their nations' teachers (see, for example, Araujo, Carneiro, Cruz-Aguayo, & Schady, 2016; Sørensen, 2016).

Accordingly, included in this special issue are a set of two peer-reviewed theoretical commentaries and seven empirical articles, via which authors present or discuss teacher evaluation policies and practices that may help us move (hopefully, well) beyond high-stakes teacher evaluation systems, especially as solely or primarily based on teachers' impacts on growing their students' standardized test scores over time (e.g., via the use of SGMs or VAMs). In this introduction to this special issue, as such, I review each of these pieces, in order to capture the essence of each of these pieces so that readers interested in these issues (and namely, each of the pieces included in this special issue) might better understand what is included and in store.

### **Special Issue Summaries**

First is a commentary authored by Jessica Holloway, Deakin University, Australia, titled "Teacher Accountability, Datafication and Evaluation: A Case for Reimagining Schooling." In this commentary, Holloway discusses how contemporary teacher accountability systems, throughout the US but also globally, have become rooted in testing, evaluation, and dis/incentivization as means for shaping school reform. In the name of equity, she details how global competitiveness and high-stakes accountability practices have steadily weakened teacher expertise, authority, and professionalism by constraining the capacity for teachers to exercise professional discretion. She argues that this continues despite the passage of ESSA (2016). Consequently, she provides a lens for thinking about the role of education and how to radically disrupt the "norms" we have come to accept as necessary features of modern schooling. More specifically, she draws on the growing subfield of "datafication" (e.g., the use of big data, statistical analyses of big data, and advanced technologies to solve highly complex social issues or problems; see, for example, Kitchin, 2014;

Lupton, 2018; Williamson, 2017) to illustrate how evolving and emerging data-related techniques and technologies are dramatically undermining teacher expertise and authority. Subsequently, she makes two major assertions. The first is that the “datafication” turn marks a distinct impact on the professional teacher, as digital data techniques proliferate our reliance on, access to, and ability to capture more data about teachers and their practice than before. As our fundamental understanding of individual people (in this case, teachers) becomes entangled with these data pictures, teachers’ data profiles begin to supersede teachers themselves. Second, while many features of the “datafied” classroom might seem rather innocuous, she argues it is important to consider how such conditions pave the way for new, more insidious, forms of data surveillance and control. In sum, she argues that the prevalence of numbers, metrics, and data within education is consistent with modern, Westernized views of “what counts” more broadly. Thus, there is an epistemological and ontological view that our problems and solutions of the world can be understood through statistical calculations. We must, consequently, question how this limited way of thinking constrains our capacity to imagine alternative versions of schooling—versions that might help confront the global challenges of our time. The present time urgently demands a radical re-thinking of education, not only because of the dangers associated with excessive datafication, but also because of pressing social and political challenges that require collective action. We must engage in thought experiments to provide some space for imagining new possibilities and thinking “outside of” the traditional accountability “box.”

Second is another commentary authored by Kelley King, University of North Texas, and Noelle Paufler, Clemson University. This commentary titled “Excavating Theory in Teacher Evaluation: Evaluation Frameworks as Wengerian Boundary Objects” is about how educational policymakers, also intent on assessing and evaluating teacher quality, have (or have not) focused on ensuring teacher competence and provided experiences for professional learning. Since the passage of ESSA (2016), some state policymakers have continued to prescribe a standard view of teacher quality across their public schools, as well as their educator preparation programs. Such evaluation frameworks are theory laden; however, they vary in terms of how explicitly they denote the theoretical underpinnings of “quality teaching” and professional learning as embedded within their states’ teacher evaluation models. The purpose of this commentary, accordingly, was to excavate said unstated theoretical underpinnings in order to better consider how contemporary teacher evaluation systems might better intersect theoretically with social learning theory (Wenger, 1998, 2000). Why? Social learning theory and the empirical research associated with social learning theory support the idea of professional learning broadly, and as participation within and across the boundaries of social communities (e.g., Communities of Practice [CoPs]). Hence, King and Paufler make the case for research examining potential connections in theory and practice. They argue that research is needed that examines and critiques the ability and desirability of current teacher evaluation systems to function as boundary objects around which CoPs can or might help to build practitioner identities and common understandings of teacher practice and what good teacher practice might mean. Theorizing evaluation through research that maps structures and processes for social learning, in other words, would effectively contribute to efforts to substantively increase teacher quality. Research conceptualizing teacher learning in social and organizational contexts, as well, would help to begin to build better and more nuanced understandings about the role of teacher evaluation in social learning, especially when making recommendations for organizational efforts to develop and support teacher learning that might better and more substantively contribute to the field.

The third contribution is the first of the set of seven empirical pieces included in this special issue, positioned first here given it provides a national view of what states are actually doing with their teacher evaluation systems post-ESSA (2016). Coauthored by myself, one of my current doctoral students, Kevin Close, and one of my former doctoral students, Clarin Collins, all at

Arizona State University, this piece titled “Putting Teacher Evaluation Systems on the Map: An Overview of States’ Teacher Evaluation Systems Post–Every Student Succeeds Act” is about whether the U.S.’s reauthorization of ESSA (2016) categorically marked a “notable inflection point” in education policy (Ross & Walsh, 2019, p. 3). Via this study we collected nationally representative survey and state website data to investigate how and to what extent states actually changed their teacher evaluation systems post-ESSA (2016). We also gathered key state personnel’s insights to capture their perceptions of the strengths and weaknesses of their states’ teacher evaluation systems once changed. While state-by-state results can be found in the full paper, we found that VAM use substantially decreased, the number of states that explicitly *do not* use or encourage VAM use substantially increased, there has been a substantial shift toward more local control, and states have taken more holistic views of and approaches towards their teacher evaluation systems post-ESSA (2016). While state department personnel expressed concerns about how there is now, perhaps, too much variety across districts’ teacher evaluation systems within their states, and there is not enough capacity to support districts’ teacher evaluation needs given this increase in variety, that states’ post-ESSA (2016) teacher evaluation systems are also more focused on formative (i.e., developmental) versus summative (i.e., outcomes-based) functions and needs was also viewed as a positive, post-ESSA (2016) trend. In short, ESSA has impacted the ways that states’ policymakers are thinking about and enacting or endorsing teacher evaluation systems, that do look different now than they did, prior to the passage of ESSA (2016) and especially after *Race to the Top* (2011). This reversal of trends, as we and many others would argue, constitute steps in the right direction.

Fourth is an empirical piece authored by Alisha Braun, University of South Florida, and Peter Youngs, University of Virginia, titled “How Middle School Special and General Educators Make Sense of and Respond to Changes in Teacher Evaluation Policy.” In this piece Braun and Youngs review the contemporary policy landscape of accountability and teacher evaluation reform, as per the use of classroom observation tools and student growth measures (SGMs, akin to VAMs), and as per the perspectives and experiences of special educators. While numerous scholars have written about the strengths and limitations of these measures for special education teacher use (Johnson & Semmelroth, 2014; Jones & Brownell, 2014; Jones, Buzick, & Turkan, 2013), few have compared the experiences of special and general education teachers. To address this gap in the literature, Braun and Youngs compared the perceptions and experiences of middle school special and general educators given a “new” teacher evaluation system in Virginia, even though that system was at the time of their study still relying on these two measures as their primary teacher evaluation indicators. What they found was considerable differences between the perceptions and experiences of special and general educators. In comparison to general educators, more specifically, special educators felt that the use of SGMs to assess teacher performance failed to evaluate a significant component of their jobs, namely their roles as case managers. Special educators also experienced conflict between the main elements of the teacher evaluation policy and their beliefs about effective teaching for students with disabilities. This conflict left the special educators studied very critical of the appropriateness of the state’s evaluation system. Ultimately, findings from this study illustrate the importance of acknowledging differences in special and general educators’ roles and responsibilities and encourage policymakers to seriously reconsider developing and implementing uniform teacher evaluation policies of the past.

Fifth is a research piece authored by Jake Malloy, University of Wisconsin-Madison, titled “Entangled Educator Evaluation Apparatuses: Contextual Influences on New Policies.” Malloy notes that along with the other states in which leaders are embracing, or at least considering, redesigns of their teacher evaluation systems, Wisconsin educational leaders are also attempting to move beyond a high-stakes, VAM-based, teacher evaluation model, so as to “inspire and empower”

(WI DPI, 2017 July). Put differently, the goal is to help Wisconsin teachers teach well and focus, more on professional development; although, as Malloy illuminates, doing this also presents its own set of challenges, given the residual “baggage” with which Wisconsin leaders must grapple as they move away from the state’s former, post-Race to the Top (2011) teacher evaluation system. Likewise, Malloy explains how and why such desirable changes may not quickly be enacted, in Wisconsin, and likely elsewhere. Drawing on Actor-Network Theory (ANT) perspectives (Latour, 1986, 2005) that conceptualize evaluation as an entangled material-discursive apparatus, Malloy more specifically explores why Wisconsin leaders have struggled to elicit full engagement from educators, despite most educators favoring the switch from a punitive accountability-based logic. Moreover, Malloy found that Wisconsin’s change in theory was not matched by a radical restructuring towards improvement, as also constrained by the state’s teacher evaluation apparatuses (see, for example, Anderson, 2017; Foucault, Davidson, & Burchell, 2008), and that the changes that were made were often not read as authentic because of the broader context in which Wisconsin educators continued to find themselves. Taking into account decades-long struggles for legitimacy by teachers and the general deprofessionalization of teachers through these and other federal and state policies, it is necessary, then, to also understand educators’ approaches to evaluation. To their credit, Wisconsin seems aware of this and has aggressively conveyed their support of the value of teachers and their improvement through professional growth and development.

Sixth, Brady Ridge and Alyson Lavigne, both at Utah State University, offer another empirical piece titled, “Improving Instructional Practice through Peer Observation and Feedback.” In this piece they explore one of the unanticipated costs of prior teacher evaluation reforms—increased pressure on school administrators to observe and provide teachers with feedback more often and in more rigorous and systematic ways. They also note that despite these efforts, only half of teachers have apparently found the feedback they have received from their principals useful (Cherasaro, Brodersen, Reale, & Yanoski, 2016). Subsequently, this problem has led many school leaders to look for alternative forms of support for their teachers. One such strategy is utilizing peers to observe and provide teachers feedback, which is a practice utilized more frequently around the globe, underutilized in the US (OECD, 2014a, 2014b), and relatively understood; although, this does show some promise (see, for example, Ackland, 1991, Lu, 2010). Hence, the purpose of this study was to conduct a systematic literature review to determine what the extant literature currently indicates about the efficacy of such an approach, in order to inform further discourse on whether peer observation and feedback might actually be a practice of promise. They evidenced that, indeed, this is an alternative observational practice of promise (alternative to the, perhaps, overreliance on administrators to do this work), but they also evidenced that this practice still lacks sufficient evidence to prevent blanket versus informed and careful adoption. The most salient benefit noted was increased teacher collaboration, whereby teachers purportedly benefited from the opportunity to work more closely with their peers; however, scholars of still very few studies have actually observed meaningful changes in teachers’ instructional practice as a result. Ridge and Lavigne conclude that future research needs to be conducted, especially if states adopt such approaches, in order to truly measure the effect of peer observation and feedback on teachers’ instructional practice, as well as student learning.

Seventh, and related, Sean Kelly, University of Pittsburgh, Robert Bringe, University of North Carolina-Chapel Hill, Esteban Aucejo, Arizona State University, and Jane Fruehwirth, University of North Carolina-Chapel Hill, contributed a research piece titled “Using Global Observation Protocols to Inform Research on Teaching Effectiveness and School Improvement: Strengths and Emerging Limitations.” In this piece they critique the teacher observation protocols often used to evaluate teachers, used perhaps most notably during the well-known Measures of

Effective Teaching (MET) Study (Bill & Melinda Gates Foundation, 2013; see also Kane & Staiger, 2012), and used to inform instructional improvement, many of which take a “global” approach to observing and measuring teacher pedagogy and instruction in practice. Indeed, this set of scholars interrogate the set of limitations of said global protocols via this study, which may represent the most comprehensive, multi-faceted critique of such protocols to date. In contrast, they argue for the use of more newly developed, fine-grained, teacher observational systems that can be used to record and more carefully analyze the individual particulars related to effective teacher practice (e.g., utterances, questions, turns at talk, etc.). These systems, Kelly, Bringe, Aucejo, and Fruehwirth argue, seem to offer states’ teacher evaluation systems, and the policies and policy-based consequences surrounding such systems, more promise and potential. Ultimately, they argue, using global observation protocols in some cases can be interpreted as positive; for example, when principals report relying on such data when making hiring decisions. Yet, and especially from a purely measurement standpoint, the limitations surrounding these global protocols that they outline in this study are severe and multifaceted. Henceforth, genuine alternatives to global protocols, including methods also relying on the latest technology in automated methods of observation, should be pursued.

Eighth, Timothy Ford, University of Oklahoma, and Kimberly Kappler Hewitt, University of North Carolina-Greensboro, offer a piece, “Better Integrating Summative and Formative Goals in the Design of Next Generation Teacher Evaluation Systems.” In this article, they explore how the two main purposes of teacher evaluation—professional growth/improvement (formative) and accountability/goal accomplishment (summative)—are often at odds with one another. Hence, they argue that the challenge of the next generation of teacher evaluation systems will be to better integrate these two purposes in policy and practice. Correspondingly, they integrate frameworks of self-determination theory (SDT; see, for example, Ford, 2018, Ryan & Brown, 2005) and Stronge’s Improvement-Oriented Model for Performance Evaluation (Stronge, 1995) to critically examine teacher evaluation policy in Hawaii and Washington, DC, two distinctly different approaches to teacher evaluation, to identify a set of policy recommendations for improving the design and implementation of teacher evaluation policies moving forward. What they found were, among multiple other findings, inequitable power relationships at levels of both policy and practice, that influence how evaluation feedback is received and used. What they called “lop-sided power dynamics” seem to stifle two-way, meaningful communication and change; hence, one primary goal surrounding both policy and practice should be to work to reduce power inequities and re-center teachers as key actors in any teacher evaluation system. This, Ford and Kappler Hewitt argue, will help to ensure that feedback gets used, not just for effectiveness judgments, but also for actually improving teaching, especially if coupled with peer support, intensive coaching, and successful modeling. Structured autonomy, clarity of expectations, and self-determined action within evaluation systems, they also argue, promote use of feedback for growth (see also a set of six, more specific recommendations for policy and practice in the full paper). While Ford and Kappler Hewitt recognize and make explicit that there is considerable tension between making evaluation personally meaningful while maintaining systems that also allow for at least some inter- or intra-teacher comparisons, they acknowledge that such comparisons should not be normative, but rather criterion-based as per sets of high professional standards.

Ninth, and finally, Mark Paige, University of Massachusetts-Dartmouth, offers his legal perspective in “Moving Forward While Looking Back: How Can VAM Lawsuits Guide Teacher Evaluation Policy in the Age of ESSA?” He notes that immediately following *Race to the Top* (2011), many states and districts rushed to adopt VAMs for purposes of teacher evaluation and high-stakes employment decisions, which subsequently landed a good number of states and districts (e.g.,

$n \cong 15$ ; see, for example, *Education Week*, 2015) in court. Drawing upon what we as a nation might learn as a result of these lawsuits, Paige provides the most important lessons for states and school districts that continue to use VAMs, or are contemplating their use, so that they might use them in much wiser, more informed, and more defensible ways. Likewise, the significance of understanding these lessons, for states and districts no longer under federal mandates post-ESSA (2016) is even more important, so that states and districts might avoid lawsuits themselves, especially if policy prone to the attachment of high-stakes consequences to VAM-based teacher evaluation output. In addition, even though evidence suggests that the use (and abuse) of VAMs is declining across states (see, for example, Close, Amrein-Beardsley, & Collins, 2020; Ross & Walsh, 2019), several states do still require or permit them, making their continued assessments, especially in terms of the law, relevant. For example, across the cases reviewed in this piece, Paige notes that plaintiffs were generally unsuccessful on theories arising under the substantive due process clause of the Fourteenth Amendment of the U.S. Constitution. However, in at least one federal case, *Houston Federation of Teachers v. Houston Independent School District* (2017), plaintiffs succeeded in their challenges to VAMs as based on the procedural due process clause of the Fourteenth Amendment. Likewise, a court upheld a challenge to the use of VAMs based on state law. Notwithstanding, Paige concludes this paper with several recommendations that caution against the use of VAMs, again, especially for high-stakes decision-making purposes. While other factors must enter such deliberations, including potential vulnerability to claims under procedural due process, state law, or collective bargaining, quite apart from assessing the legal liability associated with using VAMs, districts must consider the “costs” of continued use of VAMs that include some of the following: the acrimony created by the use of VAMs, a district’s capacity to effectively implement and provide actionable feedback based on VAMs, and the costs of defending (in court) their continued use, if needed.

## Conclusion

A close read of these nine articles reveals the tensions still ongoing, really regardless of the passage of ESSA (2016), primarily between policy and research communities, surrounding the evaluation of teacher effectiveness and quality. This is notably evidenced in the commentary authored by Holloway, who notes that these tensions are fundamentally and firmly rooted in epistemological and ontological views that our nation’s (and other nations’) problems and solutions can be understood through “datafication.” For starters, given the freedom we have been afforded by ESSA (2016), Holloway argues, we must consistently question how such limited ways of thinking actually constrain our capacity to imagine new and more innovative solutions to said problems. King and Paufler offer in their commentary one such solution; that is, to excavate the theoretical underpinnings surrounding current teacher evaluation systems in order to better consider how they might better intersect with social learning theory, so as to better support the ideas of professional learning more broadly, also via stakeholder perspectives and participation within and across social and professional boundaries (e.g., via CoPs).

Similar, albeit more pragmatic tensions are evidenced in the pieces by Braun and Youngs and Ford and Kappler Hewitt. Braun and Youngs evidenced how special educators experienced conflict and dissonance, in comparison to their general education peers, when being evaluated using a teacher evaluation system initially developed to be uniform across teachers. Policy implications here include but are not limited to the development and implementation of teacher evaluation policies that not only acknowledge how teacher roles differ by subject area, but also by and as situated within various classroom-, school-, district-, and community-based contexts. The purported need for uniformity may not, in fact, be all that necessary, especially when the goals of a teacher evaluation



system might be to support teachers, in order to better support all types of students in all types of learning. Related, Ford and Kappler Hewitt make explicit that there is considerable tension between making evaluation personally meaningful, especially as situated within the two main purposes of teacher evaluation—professional growth/improvement (formative) and accountability/goal accomplishment (summative)—both of which are often at odds with each other. Hence, they argue that the challenge of the next generation of teacher evaluation systems will be to better integrate these two purposes into both policy and practice, with emphases on offering solutions to ensure that teachers are not pitted against one another, and to also receive better feedback that can be more easily accessed, understood, internalized, and then used in order to actually improve teaching.

At a larger scale, tensions are noted in the empirical piece authored by Close, Collins, and myself, in terms of how the state-level changes observed post-ESSA (2016) might be interpreted as progressive; although, some states are still very much grappling with adopting and applying such changes. This is true, we argue, likely given the substantial financial and human resources invested in states' post-Race to the Top (2011) teacher evaluation systems and the residual effects of these systems. Put differently, even though the US is four years past the passage of ESSA (2016), the sweeping reforms called for and incentivized via Race to the Top (2011) are not shifting as rapidly as one might have thought, especially given the enthusiasm that followed after ESSA (2016) was passed (see, for example, Strauss, 2016). Notwithstanding, change is obvious, as is another set of tensions arising as changes take place (e.g., state leaders facing difficulties when trying to support states' districts' now more varied teacher evaluation systems). Associated, in his contribution, Malloy explains how and why such changes may not quickly (or as quickly as possibly anticipated) be enacted, with a case in point coming from Wisconsin. Malloy, more specifically, explores why Wisconsin leaders have struggled to elicit full engagement from educators, despite most educators favoring the switch from their state's former and relatively punitive accountability-based logic. He ultimately argues that delays can be attributed to the fact that Wisconsin's change in theory was not matched by the radical restructuring for which said theory called. Again, and to their credit, however, Wisconsin seems aware of this and is continuing to move forward with a new and improved teacher evaluation system.

One practice of promise that states like Wisconsin might consider is presented in the systematic literature review offered by Ridge and Lavigne. In sum, they evidenced that developing and implementing, all the while studying peer observation and feedback systems, may offer a sound alternative to the more traditional teacher observational practices of the past, whereby administrators do this work and, apparently and in general, do not do it very well. Inversely, it is becoming increasingly apparent that teachers engaged with peer observation and feedback systems purportedly benefit from the increased opportunities to work more closely with on another such observational approaches offer; although, scholars of few studies have thus far documented significant changes in teachers' actual instructional practices as a result. Hence, while Ridge and Lavigne do offer a practice of promise in this piece, they note that states might move forward with care and concern about the intended (and unintended) effects that might result if peer observation and feedback systems are developed and implemented. Related, Kelly, Bringe, Aucejo, and Fruehwirth, after offering another thorough and thoughtful critique of traditional observation systems (or "global" observational protocols), argue for the use of more newly developed and fine-grained teacher observational systems that can be used to record and more carefully analyze more nuanced and individual particulars related to effective teacher practice, as well as rely on the latest technologies in automated methods of observation. These systems, they posit, will also offer states' teacher evaluation systems more promise and potential in terms of actually supporting teachers with better feedback, which would likely lead to more internalization and effective use.

While not necessarily a practice of promise, is a set of policy recommendations that come from the final piece in this special issue. This piece, authored by Paige, I would interpret, perhaps, most important for states still using or contemplating using VAMs in their post-ESSA (2016) teacher evaluation policies, systems, and plans. Drawing from the approximately 15 lawsuits that came about a result of states' adoptions and implementations of high-stakes teacher evaluation policies, as primarily (or solely) based on VAM-based teacher evaluation output (*Education Week*, 2015), Paige provides us with the most important lessons for states and school districts to use, to not only move their teacher evaluation systems forward in wise and informed ways, but also in more legally defensible ways, especially so as to keep them out of court. In terms of policy implications, in other words, this set of law-based recommendations I would interpret as critical.

Otherwise, it is in this context that these theoretical and empirical papers are presented to readers, individually and collectively, as these papers stand to “add” much “value” to our current thought, with implications for both practice and policy, in and of themselves. While these pieces not only contribute to the literature regarding teacher evaluation systems and the federal and state educational policies that surround them, they also contribute to our collective thought about how policymakers, their affiliates, and others might think in more forward-thinking and innovative ways when moving (or attempting to move) their teacher evaluation systems and measures forward so as to, ultimately, help teachers improve upon their practice and help students learn and achieve more, and more in terms of what actually matters.

## References

- Ackland, R. (1991). A review of the peer coaching literature. *Journal of Staff Development*, 12(1), 22–26.
- Anderson, B. (2017). *Encountering affect: Capacities, apparatuses, conditions*. Routledge.
- Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in Kindergarten. *The Quarterly Journal of Economics*, 1415–1453. <https://doi.org/10.1093/qje/qjw016>
- Bill & Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study*. Seattle, WA. Retrieved from <http://www.gatesfoundation.org/press-releases/Pages/MET-Announcement.aspx>
- Cherasaro, T. L., Brodersen, R. M., Reale, M. L., & Yanoski, D. C. (2016). *Teachers' responses to feedback from evaluators: What feedback characteristics matter?* (REL 2017-190). Regional Educational Laboratory Central. Retrieved from [https://ies.ed.gov/ncee/edlabs/regions/central/pdf/REL\\_2017190.pdf](https://ies.ed.gov/ncee/edlabs/regions/central/pdf/REL_2017190.pdf)
- Close, K., Amrein-Beardsley, A., & Collins, C. (2020). Putting teacher evaluation systems on the map: An overview of states' teacher evaluation systems post-Every Student Succeeds Act. *Education Policy Analysis Archives*, 28(58). <https://doi.org/10.14507/epaa.28.5252>
- Close, K., Amrein-Beardsley, A., & Collins, C. (2018). *State-level assessments and teacher evaluation systems after the passage of the Every Student Succeeds Act: Some steps in the right direction*. Nation Education Policy Center (NEPC). Retrieved from <http://nepc.colorado.edu/publication/state-assessment>
- Education Week. (2015). *Teacher evaluation heads to the courts*. Retrieved from <http://www.edweek.org/ew/section/multimedia/teacher-evaluation-heads-to-the-courts.html>
- Ford, T. G. (2018). Pointing teachers in the wrong direction: Understanding Louisiana elementary teachers' use of *Compass* high stakes teacher evaluation data. *Educational Assessment, Evaluation, and Accountability*, 30(3), 251-283. <https://doi.org/10.1007/s11092-018-9280-x>

- Foucault, M., Davidson, A. I., & Burchell, G. (2008). *The birth of biopolitics: Lectures at the Collège de France, 1978-1979*. United Kingdom: Palgrave Macmillan.
- Houston Federation of Teachers v. Houston Independent School District, 251 F. Supp. 3d 1168 (2017).
- Johnson, E., & Semmelroth, C.L. (2014). Special education teacher evaluation: Why it matters, what makes it challenging, and how to address these challenges. *Assessment for Effective Intervention*, 39, 71-82. <https://doi.org/10.1177/1534508413513315>
- Jones, N. D., & Brownell, M. T. (2014). Examining the use of classroom observations in the evaluation of special education teachers. *Assessment for Effective Intervention*, 39(2), 112-24. <https://doi.org/10.1177/1534508413514103>
- Jones, N., Buzick, H., & Turkan, S. (2013). Including students with disabilities and English language learners in measures of educator effectiveness. *Educational Researcher*, 42(4), 234-241. <https://doi.org/10.3102/0013189X12468211>
- Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Bill & Melinda Gates Foundation. Retrieved from <http://files.eric.ed.gov/fulltext/ED540960.pdf>
- Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). <https://doi.org/2053951714528481>
- Kraft, M. A., & Gilmour, A. F. (2017). Revisiting the Widget Effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, 46(5) 234-249. <https://doi.org/10.3102/0013189X17718797>
- Latour, B. (1986). Visualization and cognition. *Knowledge and Society*, 6, 1–40.
- Latour, B. (2005). *Reassembling the social: An introduction to Actor-Network-Theory*. Oxford University Press.
- Lu, H. L. (2010). Research on peer coaching in preservice teacher education – A review of literature. *Teaching and Teacher Education*, 26(4), 748–753. <https://doi.org/10.1016/j.tate.2009.10.015>
- Lupton, D. (2018). How do data come to matter? Living and becoming with personal data. *Big Data & Society*, 5(2). <https://doi.org/2053951718786314>
- No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115 Stat. 1425. (2002). Retrieved from <http://www.ed.gov/legislation/ESEA02/>
- Organisation for Economic Co-operation and Development (OECD). (2014a). *TALIS 2013 results: An international perspective on teaching and learning*. Retrieved from [http://www.keepeek.com/Digital-Asset-Management/oecd/education/talis-2013-results\\_9789264196261-en](http://www.keepeek.com/Digital-Asset-Management/oecd/education/talis-2013-results_9789264196261-en)
- Organisation for Economic Co-operation and Development (OECD). (2014b). *Results from TALIS 2013: United States of America*. Retrieved from <http://www.oecd.org/unitedstates/TALIS2013-country-note-US.pdf>.
- Race to the Top Act of 2011, S. 844—112th Congress. (2011). Retrieved from <http://www.govtrack.us/congress/bills/112/s844>
- Ross, E. & Walsh, K. (2019). *State of the states 2019: Teacher and principal evaluation policy*. National Council on Teacher Quality (NCTQ). Retrieved from <https://www.nctq.org/pages/State-of-the-States-2019:-Teacher-and-Principal-Evaluation-Policy>
- Ryan, R. M., & Brown, K. W. (2005). Legislating competence: The motivational impact of high-stakes testing as an educational reform. In C. Dweck & A. Elliot (Eds.), *Handbook of competence and motivation* (pp. 354-372). Guilford Press.

- Sørensen, T. B. (2016). *Value-added measurement or modelling (VAM)*. Education International. Retrieved from [http://download.ei-ie.org/Docs/WebDepot/2016\\_EI\\_VAM\\_EN\\_final\\_Web.pdf](http://download.ei-ie.org/Docs/WebDepot/2016_EI_VAM_EN_final_Web.pdf)
- Strauss, V. (2016). Explaining key points of the new K-12 education law. *The Washington Post*. Retrieved from <https://www.washingtonpost.com/news/answer-sheet/wp/2016/01/21/explaining-key-points-of-the-new-k-12-education-law/>
- Stronge, J. H. (1995). Balancing individual and institutional goals in educational personnel evaluation: A conceptual framework. *Studies in Educational Evaluation*, 21, 131-151. [https://doi.org/10.1016/0191-491X\(95\)00010-R](https://doi.org/10.1016/0191-491X(95)00010-R)
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.
- Wenger, E. (2000). Communities of practice and social learning systems. *Organization*, 7, 225-246. <https://doi.org/10.1177/02F135050840072002>
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The Widget Effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. New Teacher Project (TNTTP). Retrieved from [http://tntp.org/assets/documents/TheWidgetEffect\\_2nd\\_ed.pdf](http://tntp.org/assets/documents/TheWidgetEffect_2nd_ed.pdf)
- Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. Sage.
- Wisconsin Department of Public Instruction. (2017, July 3). *Empowered Educators*. <https://dpi.wi.gov/statesupt/every-child-graduate/empowered-educators>

## About the Author/Guest Editor

### **Audrey Amrein-Beardsley**

Arizona State University

Email: [audrey.beardsley@asu.edu](mailto:audrey.beardsley@asu.edu)

ORCID: <https://orcid.org/0000-0002-1250-2281>

Audrey Amrein-Beardsley, PhD., is a Professor in the Mary Lou Fulton Teachers College at Arizona State University. Her research focuses on the use of value-added models (VAMs) in and across states before and since the passage of the Every Student Succeeds Act (ESSA). More specifically, she is conducting validation studies on multiple system components, as well as serving as an expert witness in many legal cases surrounding the (mis)use of VAM-based output. Audrey Amrein-Beardsley is also Lead Editor of EPAA; however, it should be noted that this piece was editorially reviewed and the piece she authored with two others (summarized as the third piece above) went through a double-blind, peer-review process of which Amrein-Beardsley had no part and over which Amrein-Beardsley has no influence.

## SPECIAL ISSUE Policies and Practices of Promise in Teacher Evaluation

## education policy analysis archives

Volume 28 Number 55

April 13, 2020

ISSN 1068-2341



Readers are free to copy, display, distribute, and adapt this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, the changes are identified, and the same license applies to the derivative work. More details of this Creative Commons license are available at <https://creativecommons.org/licenses/by-sa/2.0/>. EPAA is published by the Mary Lou Fulton Institute and Graduate School of Education at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A1 (Brazil), SCImago Journal Rank, SCOPUS, SOCOLAR (China).

Please send errata notes to Audrey Amrein-Beardsley at [audrey.beardsley@asu.edu](mailto:audrey.beardsley@asu.edu)

Join EPAA's Facebook community at <https://www.facebook.com/EPAAAPE> and Twitter feed @epaa\_aape.

education policy analysis archives  
editorial board

Lead Editor: **Audrey Amrein-Beardsley** (Arizona State University)

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Associate Editors: **David Carlson, Lauren Harris, Eugene Judson, Mirka Koro-Ljungberg, Scott Marley, Molly Ott, Iveta Silova**, (Arizona State University)

**Cristina Alfaro** San Diego State University

**Gary Anderson** New York University

**Michael W. Apple** University of Wisconsin, Madison

**Jeff Bale** OISE, University of Toronto, Canada

**Aaron Bevenot** SUNY Albany

**David C. Berliner** Arizona State University

**Henry Braun** Boston College

**Casey Cobb** University of Connecticut

**Arnold Danzig** San Jose State University

**Linda Darling-Hammond** Stanford University

**Elizabeth H. DeBray** University of Georgia

**Chad d'Entremont** Rennie Center for Education Research & Policy

**John Diamond** University of Wisconsin, Madison

**Matthew Di Carlo** Albert Shanker Institute

**Sherman Dorn** Arizona State University

**Michael J. Dumas** University of California, Berkeley

**Kathy Escamilla** University of Colorado, Boulder

**Yariv Feniger** Ben-Gurion University of the Negev

**Melissa Lynn Freeman** Adams State College

**Rachael Gabriel** University of Connecticut

**Amy Garrett Dikkers** University of North Carolina, Wilmington

**Gene V Glass** Arizona State University

**Ronald Glass** University of California, Santa Cruz

**Jacob P. K. Gross** University of Louisville

**Eric M. Haas** WestEd

**Julian Vasquez Heilig** California State University, Sacramento

**Kimberly Kappler Hewitt** University of North Carolina Greensboro

**Aimee Howley** Ohio University

**Steve Klees** University of Maryland  
**Jaekyung Lee** SUNY Buffalo

**Jessica Nina Lester** Indiana University

**Amanda E. Lewis** University of Illinois, Chicago

**Chad R. Lochmiller** Indiana University

**Christopher Lubienski** Indiana University

**Sarah Lubienski** Indiana University

**William J. Mathis** University of Colorado, Boulder

**Michele S. Moses** University of Colorado, Boulder

**Julianne Moss** Deakin University, Australia

**Sharon Nichols** University of Texas, San Antonio

**Eric Parsons** University of Missouri-Columbia

**Amanda U. Potterton** University of Kentucky

**Susan L. Robertson** Bristol University

**Gloria M. Rodriguez** University of California, Davis

**R. Anthony Rolle** University of Houston

**A. G. Rud** Washington State University

**Patricia Sánchez** University of University of Texas, San Antonio

**Janelle Scott** University of California, Berkeley

**Jack Schneider** University of Massachusetts Lowell

**Noah Sobe** Loyola University

**Nelly P. Stromquist** University of Maryland

**Benjamin Superfine** University of Illinois, Chicago

**Adai Tefera** Virginia Commonwealth University

**Tina Trujillo** University of California, Berkeley

**Federico R. Waitoller** University of Illinois, Chicago

**Larisa Warhol** University of Connecticut

**John Weathers** University of Colorado, Colorado Springs

**Kevin Welner** University of Colorado, Boulder

**Terrence G. Wiley** Center for Applied Linguistics

**John Willinsky** Stanford University

**Jennifer R. Wolgemuth** University of South Florida

**Kyo Yamashiro** Claremont Graduate University

archivos analíticos de políticas educativas  
consejo editorial

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Editores Asociados: **Felicitas Acosta** (Universidad Nacional de General Sarmiento), **Armando Alcántara Santuario** (Universidad Nacional Autónoma de México), **Ignacio Barrenechea**, **Jason Beech** (Universidad de San Andrés), **Angelica Buendía**, (Metropolitan Autonomous University), **Alejandra Falabella** (Universidad Alberto Hurtado, Chile), **Carmuca Gómez-Bueno** (Universidad de Granada), **Veronica Gottau** (Universidad Torcuato Di Tella), **Carolina Guzmán-Valenzuela** (Universidade de Chile), **Antonia Lozano-Díaz** (University of Almería), **Antonio Luzon**, (Universidad de Granada), **María Teresa Martín Palomo** (University of Almería), **María Fernández Mellizo-Soto** (Universidad Complutense de Madrid), **Tiburcio Moreno** (Autonomous Metropolitan University-Cuajimalpa Unit), **José Luis Ramírez**, (Universidad de Sonora), **Axel Rivas** (Universidad de San Andrés), **César Lorenzo Rodríguez Uribe** (Universidad Marista de Guadalajara), **María Veronica Santelices** (Pontificia Universidad Católica de Chile)

**Claudio Almonacid**

Universidad Metropolitana de  
Ciencias de la Educación, Chile

**Miguel Ángel Arias Ortega**

Universidad Autónoma de la  
Ciudad de México

**Xavier Besalú Costa**

Universitat de Girona, España

**Xavier Bonal Sarro**

Universidad  
Autónoma de Barcelona, España

**Antonio Bolívar Boitia**

Universidad de Granada, España

**José Joaquín Brunner**

Universidad  
Diego Portales, Chile

**Damián Canales Sánchez**

Instituto Nacional para la  
Evaluación de la Educación,  
México

**Gabriela de la Cruz Flores**

Universidad Nacional Autónoma de  
México

**Marco Antonio Delgado Fuentes**

Universidad Iberoamericana,  
México

**Inés Dussel**, DIE-CINVESTAV,

México

**Pedro Flores Crespo**

Universidad  
Iberoamericana, México

**Ana María García de Fanelli**

Centro de Estudios de Estado y  
Sociedad (CEDES) CONICET,  
Argentina

**Juan Carlos González Faraco**

Universidad de Huelva, España

**María Clemente Linuesa**

Universidad de Salamanca, España

**Jaume Martínez Bonafé**

Universitat de València, España

**Alejandro Márquez Jiménez**

Instituto de Investigaciones sobre la  
Universidad y la Educación,  
UNAM, México

**María Guadalupe Olivier Tellez**,

Universidad Pedagógica Nacional,  
México

**Miguel Pereyra**

Universidad de  
Granada, España

**Mónica Pini**

Universidad Nacional  
de San Martín, Argentina

**Omar Orlando Pulido Chaves**

Instituto para la Investigación  
Educativa y el Desarrollo  
Pedagógico (IDEP)

**José Ignacio Rivas Flores**

Universidad de Málaga, España

**Miriam Rodríguez Vargas**

Universidad Autónoma de  
Tamaulipas, México

**José Gregorio Rodríguez**

Universidad Nacional de Colombia,  
Colombia

**Mario Rueda Beltrán**

Instituto de  
Investigaciones sobre la Universidad  
y la Educación, UNAM, México

**José Luis San Fabián Maroto**

Universidad de Oviedo,  
España

**Jurjo Torres Santomé**,

Universidad  
de la Coruña, España

**Yengny Marisol Silva Laya**

Universidad Iberoamericana,  
México

**Ernesto Treviño Ronzón**

Universidad Veracruzana, México

**Ernesto Treviño Villarreal**

Universidad Diego Portales  
Santiago, Chile

**Antoni Verger Planells**

Universidad Autónoma de  
Barcelona, España

**Catalina Wainerman**

Universidad de San Andrés,  
Argentina

**Juan Carlos Yáñez Velazco**

Universidad de Colima, México

arquivos analíticos de políticas educativas  
conselho editorial

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Editoras Associadas: **Andréa Barbosa Gouveia** (Universidade Federal do Paraná), **Kaizo Iwakami Beltrao**, (Brazilian School of Public and Private Management - EBAPE/FGV), **Sheizi Calheira de Freitas** (Federal University of Bahia), **Maria Margarida Machado**, (Federal University of Goiás / Universidade Federal de Goiás), **Gilberto José Miranda**, (Universidade Federal de Uberlândia, Brazil), **Marcia Pletsch** (Universidade Federal Rural do Rio de Janeiro), **Maria Lúcia Rodrigues Muller** (Universidade Federal de Mato Grosso e Science), **Sandra Regina Sales** (Universidade Federal Rural do Rio de Janeiro)

**Almerindo Afonso**

Universidade do Minho  
Portugal

**Alexandre Fernandez Vaz**

Universidade Federal de Santa  
Catarina, Brasil

**José Augusto Pacheco**

Universidade do Minho, Portugal

**Rosanna Maria Barros Sá**

Universidade do Algarve  
Portugal

**Regina Célia Linhares Hostins**

Universidade do Vale do Itajaí,  
Brasil

**Jane Paiva**

Universidade do Estado do Rio de  
Janeiro, Brasil

**Maria Helena Bonilla**

Universidade Federal da Bahia  
Brasil

**Alfredo Macedo Gomes**

Universidade Federal de Pernambuco  
Brasil

**Paulo Alberto Santos Vieira**

Universidade do Estado de Mato  
Grosso, Brasil

**Rosa Maria Bueno Fischer**

Universidade Federal do Rio Grande  
do Sul, Brasil

**Jefferson Mainardes**

Universidade Estadual de Ponta  
Grossa, Brasil

**Fabiany de Cássia Tavares Silva**

Universidade Federal do Mato  
Grosso do Sul, Brasil

**Alice Casimiro Lopes**

Universidade do Estado do Rio de  
Janeiro, Brasil

**Jader Janer Moreira Lopes**

Universidade Federal Fluminense e  
Universidade Federal de Juiz de Fora,  
Brasil

**António Teodoro**

Universidade Lusófona  
Portugal

**Suzana Feldens Schwertner**

Centro Universitário Univates  
Brasil

**Debora Nunes**

Universidade Federal do Rio Grande  
do Norte, Brasil

**Lílian do Valle**

Universidade do Estado do Rio de  
Janeiro, Brasil

**Geovana Mendonça Lunardi**

**Mendes** Universidade do Estado de  
Santa Catarina

**Alda Junqueira Marin**

Pontifícia Universidade Católica de  
São Paulo, Brasil

**Alfredo Veiga-Neto**

Universidade Federal do Rio Grande  
do Sul, Brasil

**Flávia Miller Naethe Motta**

Universidade Federal Rural do Rio de  
Janeiro, Brasil

**Dalila Andrade Oliveira**

Universidade Federal de Minas  
Gerais, Brasil