

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 60

April 13, 2020

ISSN 1068-2341

**Entangled Educator Evaluation Apparatuses:
Contextual Influences on New Policies**

Jake Malloy

University of Wisconsin - Madison
USA

Citation: Malloy, J. (2020). Entangled educator evaluation apparatuses: Contextual influences on new policies. *Education Policy Analysis Archives*, 28(60). <https://doi.org/10.14507/epaa.28.5016> This article is part of the special issue, *Policies and Practices of Promise in Teacher Evaluation*, guest edited by Audrey Amrein-Beardsley.

Abstract: Drawing on an actor-network articulation of evaluation theory, this article examines the Wisconsin Department of Public Instruction's transition from a punitive teacher evaluation model to a promising new development and support model, which focuses on teacher growth and environmental adjustments. Supported by dozens of interviews and observations of teachers, school and district administrators, support staff, and regional and state education organization employees, the article explains how material, discursive, and affective entanglements within and outside the evaluation apparatus constrained the realization of the new growth model of teacher evaluation. Actor-network evaluation provides a new articulation of evaluation contextualization that provides insight into why some promising changes may run out of steam.

Keywords: Wisconsin; Actor-Network Theory; Educator Effectiveness

Journal website: <http://epaa.asu.edu/ojs/>
Facebook: /EPAAA
Twitter: @epaa_aape

Manuscript received: 8/31/2019
Revisions received: 1/15/2020
Accepted: 2/20/2020

Aparatos de evaluación de educadores enredados: Influencias contextuales en las nuevas políticas

Resumen: Basándose en una articulación actor-red de la teoría de la evaluación examina la transición del Departamento de Instrucción Pública de Wisconsin de un modelo punitivo de evaluación docente a un nuevo modelo prometedor de desarrollo y apoyo, que se centra en el crecimiento docente y los ajustes ambientales. Apoyado por docenas de entrevistas y observaciones de maestros, administradores de escuelas y distritos, personal de apoyo y empleados de organizaciones educativas regionales y estatales, el artículo explica cómo los enredos materiales, discursivos y afectivos dentro y fuera del aparato de evaluación limitaron la realización del nuevo crecimiento modelo de evaluación docente. La evaluación de la red de actores proporciona una nueva articulación de la contextualización de la evaluación que proporciona información sobre por qué algunos cambios prometedores pueden no tener éxito.

Palabras clave: Wisconsin; Teoría de actor-red; Eficacia del educador

Aparelhos de avaliação de educadores emaranhados: Influências contextuais em novas políticas

Resumo: Desenho de uma articulação ator-rede da teoria da avaliação, este artigo examina a transição do Departamento de Instrução Pública de Wisconsin de um modelo punitivo de avaliação de professores para um promissor novo modelo de desenvolvimento e apoio, que se concentra no crescimento do professor e em ajustes ambientais. Apoiado por dezenas de entrevistas e observações de professores, administradores de escolas e distritos, equipe de apoio e funcionários de organizações regionais e estaduais de educação, o artigo explica como os entrelaços materiais, discursivos e afetivos dentro e fora do aparato de avaliação restringiram a realização do novo crescimento. modelo de avaliação de professores. A avaliação da rede de atores fornece uma nova articulação da contextualização da avaliação que fornece informações sobre por que algumas mudanças promissoras podem não ter êxito.

Palavras-chave: Wisconsin; Teoria Ator-Rede; Eficácia do educador

Introduction

But if you may ask it's just the things that you do, that reminds me of all of the mess that I've been through. (Mary J. Blige, *Baggage*, 2005)

Since the passage of the Every Student Succeeds Act (ESSA) in 2015, states have been granted renewed discretion to design their teacher evaluation systems. Some states, such as Wisconsin, are attempting to move beyond high-stakes, value-added teacher evaluation toward a growth and support model. This article illuminates the implementation challenges the new evaluation model has encountered as it deals with the baggage of previous evaluations and contemporary discourses with which it is entangled. Using an Actor-Network (ANT) lens, I examine the material, discursive, and affective entanglements Wisconsin's Educator Development and Support program (EDS) wrestles with while transitioning to teacher evaluation that focuses on formative assessment, training, and structural support. I then explain how ANT can highlight aspects of contextualization important for policymakers when attempting policy change, and end by reiterating the exciting direction Wisconsin is trying to proceed as it disentangles itself. This article seeks to both highlight a promising new approach to evaluation and also explain the challenges encountered in its implementation.

In 2011, Wisconsin Act 166 initiated educator effectiveness legislation, under which all classroom teachers (and later principals) were to be assessed as to their individual effectiveness and placed into a performance category (2011).¹ According to that law, half the evaluation score must be based on student performance measures. Observation of practice compared to interstate teaching and school leadership standards determines the remainder (Wisconsin Law 115.415, 2014). This division of teacher quality into practices and results mirrors a national push generally for teacher evaluations to increasingly be measured by student test score (outcomes) rather than observation of pedagogy (inputs), seen in both ESEA flexibility waivers and Race to the Top guidelines (*RTT-D FAQ*; U.S. Department of Education, 2013).

Recently, Wisconsin's Department of Public Instruction (DPI) has embedded its Educator Effectiveness system (EE),² the apparatus by which educators are evaluated, into the more encompassing EDS in an effort to shift the purpose of educator evaluation from accountability to direct improvement (WI Educator Development & Support, 2017). EDS hopes teachers are "inspired and empowered" to teach well, focusing on professional development through increasing resources and mentoring (Empowered Educators, 2017).

Drawing on Actor-Network Theory perspectives that conceptualize evaluation as an entangled material-discursive apparatus, this article explores why Wisconsin's EDS system has struggled to elicit full engagement from educators despite most educators favoring the switch from punitive accountability to growth-based logic. This article explains how and why a desirable change in policy may not be quickly enacted, explores the challenges of shifting teacher evaluation policy

¹ Though the 2011 Accountability Design Team, made up of "key stakeholders from the business community, parent organizations, philanthropic representatives, elected officials, student advocacy groups, and education leaders, including tribal leaders," recommended three categories, *developing*, *effective*, and *exemplary*, where an "educator will not be allowed to remain at the developing level and continue to practice indefinitely" (U.S. Department of Education, 2013, pp. 19, 130), Wisconsin's Department of Public Instruction (DPI) settled on a four category system: 1 - *Unsatisfactory*, 2 - *Basic*, 3 - *Proficient*, and 4 - *Distinguished* (Wisconsin Department of Public Instruction, 2018).

² Each district uses one of two virtually identical evaluation platforms: the DPI-created Educator Effectiveness program or the CESA 6-created Effectiveness Project. For simplicity without significant loss of insight, I will refer to both as EE.

and practice away from high-stakes approaches, and glimpses promising policies and practices in Wisconsin.

Actor-Network Perspectives on Teacher Evaluation

Actor-Network Theory (ANT) perspectives focus on associations and foreground enrollment and mobilization of allies in order to act at greater distance and “force dissenters into believing new facts and behaving in new ways” (Latour, 1986, p. 6). An actor-network, as the name suggests, considers each object as simultaneously singular, capable of acting, that is, altering its environment, and as plural, comprised of ever-changing associations of other material-semiotic objects. ANT provides the framework for conceptualizing both the entanglements that resist transformation and the mechanisms by which new attachments engender change. By entanglements and attachments, I mean the material and discursive relationships that constraint and afford various actions. Viewing an evaluation policy-practice³ as a network allows me, as the researcher, to trace the connections and interactions among various human and nonhuman actors that negotiate alliances.

I juxtapose a *theory* of improvement (TI) and an *apparatus* of improvement (AI). The theory represents a logic, whether tacit or explicit, explaining how a system progresses toward a goal state. DPI is attempting to transition away from a TI my participants referred to as the *Rate, Rank, Remove* approach (RRR),⁴ in which improvement is assumed to occur through rating the quality of each member of the set, ranking each, and then removing the bottom portion, which is not returning its investment. Such pruning ostensibly allows more resources to be directed to higher performing members (Welch, 2005).

AI represents the mechanisms and processes (the creating, collecting, calculating, and communicating of data and decisions) by which the TI is enacted. For the above TI, this would involve the processes from creating criteria upon which teachers are evaluated to analyzing results informing the rating and ranking to the processes of removing and replacing or reallocating personnel and resources. By considering EDS as the AI, one is able to consider the ways component parts support or contest a particular TI. I use *apparatus* (Anderson, 2017; Foucault et al., 2008) as a special type of actor-network, or a particular assembly of discursive and material entities that inscribes objects in reality. Under the ANT paradigm, the realness of the *facts* produced by an evaluation apparatus is contingent on their acceptance within a network of allies who subsequently encourage/pressure others to accept the facts. This is to say, an apparatus defines an object’s (in our case, a teacher) boundaries and properties in order to make the object knowable and actionable by constructing, collecting, and calculating data and conveying new facts. The realness of the object corresponds to the object’s capacity to alter the actions, decisions, and/or the decision-space of the surrounding objects (Daston, 2000). I use the term decision-space to represent the range and arrangement of considerations leading to a decision. In this article, EDS, the evaluation-as-process, is framed as an apparatus, interacting with other actors to shape new conceptual objects—teacher quality, which are, in large part, experienced through the inscriptions of the evaluation-as-product—reports which give teachers a score on a 4-point scale.

³ Throughout the article I use the Policy as Practice paradigm (Sutton & Levinson, 2001), whereby policy is understood broadly to encapsulate the everyday practices of those who enact and interact with policies in the traditional sense, as it aligns well with actor-network onto-epistemology.

⁴ For a public document explaining that evaluation in Wisconsin should not use RRR (see Milwaukee Public Schools, 2014, p. 2).

Methods

In order to strike a balance between the sprawling network of the Wisconsin school system and the richness of experiences of local participants, this study focuses primarily on five particular and interconnected organizations: the state Department of Public Instruction (DPI), two regional educational organizations focused on opposite sides of the state (CESAs), and two districts, each served by one of the respective regional organizations.

With IRB approval, over the course of a year I attended several meetings related to evaluation and assessment. I attended five open forums; four regarding the Every Student Succeeds Act (ESSA), and one on Educator Effectiveness. I conducted nineteen semi-structured interviews, each approximately an hour long and another fifteen each about half an hour long. These included teachers, school and district administrators, support staff, and CESA and DPI employees. I also observed several training sessions held at the CESA West⁵ building for representatives from its member districts. I observed Waterside District on several occasions, attending an EE training session, several classroom observations, and associated administrator-teacher conferences. At Three Farms District I observed three daylong teacher-staff ‘data meetings’ which discuss ESSA services for students. I also used down time during all observations as unstructured interviews, asking clarifying questions or gaining more contextual understanding of the sites and participants.

Using Nvivo to code typed field notes, transcribed interview recordings, and collected documents, I explored themes germane to intertextual and agential roles of the various actors, including nonhuman ones, such as documents, aligning with the expectations of the theoretical lens (Foucault, 1991; Latour, 2005; Prior, 2003). Salient codes were synthesized into memos, which have been organized into the narrative below.

Wisconsin’s New(ish) Evaluation: The Challenge of Change

In this section I explore more fully first the material-discursive components of the EDS apparatus of improvement that have been altered and then those that have remained unchanged. For each I examine their effects on educators and others within the evaluation system.

Moving Toward Growth

As one DPI employee explained, the department is in the process of rebranding itself as the “kinder, gentler DPI.” Messaging, according to another, is a vital component of their transition. The growth model transition is a continual process. One DPI employee explained in an interview that the initial 2012 Educator Effectiveness design is a chimera born of the competing concerns of the governor and legislature for an accountability model on one side and DPI and teachers for a “professional development model” on the other. According to this participant, the Governor’s office wanted teacher accountability to be 100% value-added, based on student test scores. Through negotiations within the Design Team, the 50% legislation was set. She called this the “Wisconsin Miracle.” As a continued effort to challenge the legislated focus on student scores, DPI, as shown in the Process Manual, liberally interpreted the 50-50 split as two dimensions of teaching (see Figure 1). Doing so ignored an overall single score, leaving the interpretation of quality up to the district (Wisconsin Department of Public Instruction, 2014, p. 29). One study participant from DPI referred to this as “creative compliance.” In the following years DPI has been attempting to subdue EE’s value-added components and focus on a “learning-centered, continuous improvement process”

⁵ All proper names that are not public record are pseudonyms to protect the anonymity of my participants.

(Wisconsin Department of Public Instruction, 2018, p. 1). In the previously quoted document, the student achievement scores are completely ignored.

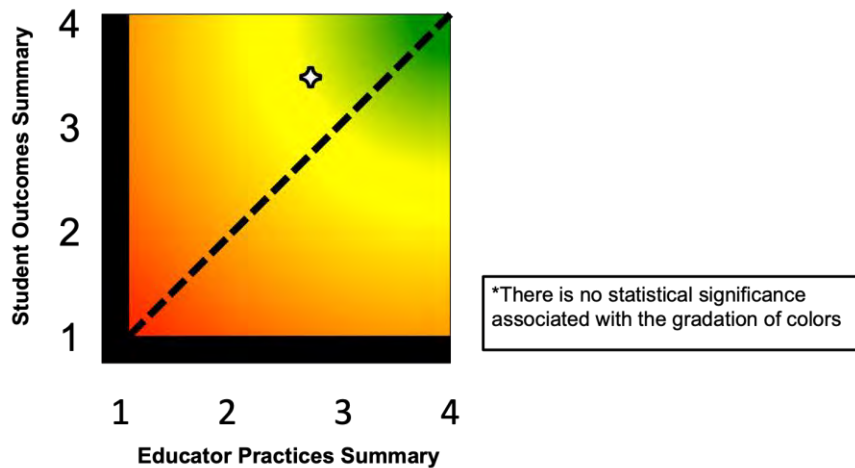


Figure 1. Educator Effectiveness Summary Graph

Though both evaluation approaches were understood as improvement processes, the new approach encourages districts and DPI to consider it their responsibility to create a “system that becomes a learning process” and a culture that supports teachers to “risk, struggle, learn, [and] improve,” as one panelist at an Educator Effectiveness open forum said.

DPI is pushing for a different way of evaluating. One study participant from DPI relayed a scene from a conference of educational leaders from numerous states:

So if you had two bad years of an evaluation, you lost your license [see footnote 1].

So could no longer teach ever. And we [WI DPI] said we would never do that. And basically all of the states laughed at us. And were kind of like why would you even do the [teacher evaluation] system? And we said we were more focused on helping educators improve. At the time, [they] literally laughed at [us].

The 2018 EE User Guide eschews accountability language and strikes a tone of partnership with educators as they set their own goals, acknowledging the importance of trust in the evaluation cycle (Wisconsin Department of Public Instruction, 2018). The importance of this paratext is apparent to DPI employees, as one explained she feels she is always struggling to “get in front of” and “counteract” the associations made between evaluation and accountability in a *Rate, Rank, Remove* (RRR) paradigm.⁶ The RRR paradigm aligns with the national discourse and is encouraged by Wisconsin’s Act 10. Act 10 was a 2011 budget bill also known as the Wisconsin Budget Repair Bill that reduced public funding and constrained teachers’ right to collective bargaining (Wisconsin Act 10, 2011). Proposal of the bill resulted in substantial protests by teachers and other civil servants at the state capitol for several months (Kelleher, 2011). From several teachers and school administrators, I heard, unprompted, the continued weight felt by Act 10. Because of this, much of the change that DPI is making rests on trying to convince educators to accept EE as a tool of professional development within professional learning communities rather than as proof of worth for human resource decisions.

⁶ Several participants used this “Rate, Rank, Remove” shorthand to describe evaluation policies that focused on teachers proving their value and incurring sanctions for failing to do so.

One way DPI attempts to detach the new EDS system from an RRR theory is to challenge its logic and expose the perceived ineffectuality of the process. For example, a DPI employee made the case that, given the interdependent nature of schooling, to “accurately attribute” quality to one teacher is virtually impossible, especially regarding outcome-based measures such as value-added scores on standardized tests. She explained, “Co-teaching, intervention with pull out coaches, even just your basic content areas. Science and social studies are focused on reading just as much as reading [class]. So how are you going to accurately attribute a reading growth score to any teacher that ever supported that student’s reading instruction?”

DPI also has been chipping away at the rating and categorizing components of EE. As one participant stated, “giving a teacher a score of one to four doesn’t tell them how to get better . . . Let’s use [EE] as a way to actually identify ways to improve.” In the spirit of rejecting value-added measures for teacher evaluation, understanding that effectiveness exists at the team level, and recognizing that competition runs counter to the desire of many teachers, DPI has begun dismantling aspects of the teacher effectiveness score that they deemed unhelpful in informing individual teachers in ways to grow. DPI has attempted to work around the constraints of legislation to the betterment of teachers and schools.

Finally, DPI has restructured the legislatively required EE to function within the broader, newly formed EDS program. In this way the AI aligns more closely to the TI through teacher learning rather than teacher replacement. The apparatus includes resources, such as grants, trainings about mentoring and induction, and support for communities of practice. For example, through EDS, DPI has introduced the Every Teacher a Leader program, which seeks to highlight the expertise of teachers and promote teachers’ voices as leaders in the classroom and community (Every Teacher a Leader, 2017). Title II funds focus on improving student achievement through teacher professional development (Title II Programs, 2016). DPI provides guidance on professional learning opportunities in order to help teachers “discern which opportunities are best” for their “continuous improvement process” as teachers “demonstrate a growth mindset” (Resources for Professional Learning, 2017). As attested in my observations and interviews, the catchphrases *growth mindset* and *continuous improvement* have thoroughly infiltrated discussion of teacher evaluation at DPI, CESA, and many districts.

Additionally, I observed how integral CESA employees have been to altering the discourse of educator evaluation from one of accountability, in the sense that teachers and administrators need to prove their worth, to one of growth. Every month one or two leaders from each of the twelve CESAs meet together as School Improvement Specialists. Throughout the day representatives from various offices in DPI explain new policies and practices. Combining this with their experience in schools throughout the month, the CESA leaders shape a “big picture” of Wisconsin schooling, as one study participant explained. Likewise, during an interview another leader conveyed how she views each policy as “a slice of a system of improvement.” Many of the CESA members are deeply committed to the logic of holistic and continuous improvement for schools, which rests on organizing support systems and providing ongoing, meaningful feedback.

The discursive work of the CESA members as they travel among districts involves attaching educator evaluations to the values of “teaching and learning.” For example, during a district principal training session a CESA representative explained to the group that EE, with the blessing of DPI, wanted to “focus on coaching, not judging; on growth, not error; and on helping educators feel supported, not attacked.” She then reiterated to the teacher just observed, “it really is about growth.” Throughout the training session, the representative guided administrators to give focused feedback, with not more than three “opportunities for growth” to increase the likelihood of the recommendations leading to changes in teacher practice. She also encouraged administrators to have

the teachers lead the discussion of their teaching in order to evoke personal reflection, again with the intention of realizing pedagogical changes.

The CESA trainings for educator evaluation shape the discursive elements of the apparatus to align with the new TI but also shape the material components, such as the data gathered by administrators. At the discursive level, DPI and CESA staff have encouraged district administrators to see professional development as their responsibility. The Waterside superintendent discussed the importance of focusing on *growth* not *gotcha* regarding teacher evaluations.⁷ The superintendent explained to me that she also found it important to convince teachers that changes in the district were “continuous and not episodic,” that they represented a shift in the policies and practices of the district holistically rather than a series of unrelated events. Such catchphrases permeate districts throughout the state as CESA leaders who meet together regularly visit districts in their respective regions.

Though the teacher evaluation framework (criteria and platform) has remained largely unchanged, the most significant material change has been enveloping the judgment apparatus, EE, in a system of support, EDS. Included in this concept is a systemic/holistic view of school improvement, where a specific teacher’s quality of instruction is not perceived as solely the responsibility of that teacher but instead as the system’s. While watching principals observe teachers and conduct post-observation meetings, I was able to see how the principal considered various levels of system adjustments in hopes of improving the effectiveness of the teachers. In one instance, the principal noticed significant losses of instructional time and invited the teacher to consider how that could be remedied. The two discussed pedagogical adjustments, such as the addition of student peer feedback, and structural adjustments, such as room arrangements and resources, and adjustments to duties, so as to provide the teacher the time to mentally transition from one class period to the next.

Another teacher had recently transferred to the school from within the district. The principal believed the previous low performance reviews of the teacher were due to a combination of the teacher’s poor rapport with that building principal and an ill-fitting teaching assignment. The teacher, now in a new teaching role, was receiving additional observations and coaching. The principal and teacher both conveyed satisfaction with the new assignment and continued growth of the teacher. Examples like this, where the onus for performance is not placed solely on the teacher but is distributed among the entire system better reflect the EDS approach than the RRR approach. I find this the most important aspect of teacher evaluation as integrated with a system of improvement: the information from observations of teaching does not simply feedback to the teacher but to other parts of the schooling system that collectively share responsibility for improving the schooling experience. However, these examples do not tell the whole story inside that district or the EDS system at large.

Maintaining Punitive Accountability

Though significant discursive work (and some material work) has occurred to disconnect Wisconsin teacher evaluation from the RRR theory and apparatus of improvement, it remains active in several respects. Over the past decade, the state has introduced policies that reflect the discourse about teachers promoted during NCLB and continued through the rhetoric surrounding both Race to the Top and Common Core State Standards, which framed teachers as cantankerous and mediocre yet indispensable, who must be incentivized and placed in systems of constraint lest they

⁷ The opposition of EE to a “gotcha system” is also present in the system guide for largest Wisconsin district, which I note to express that this is not an isolated occurrence (Milwaukee Public Schools, 2014, p. 2).

cease work or stray from the path (Cochran-Smith & Lytle, 2006; Kumashiro et al., 2012). And, as mentioned previously, the impetus for the educator effectiveness program came from the NCLB waiver requirements and its initial shape satisfied the conditions and interests of that policy. As Payne (2008) laments, “part of what is distinctive about the current wave of reform is the degree to which it is founded on (and undermined by) disrespect for educators” (p.91). Though DPI seeks to buck this trend, teachers who worked through the era of NCLB and Act 10 are keenly aware of the broad discourse on teachers, felt the state government has taken an adversarial stance toward teachers, and are wary of evaluative policies. Numerous teachers I interviewed discussed the continued negative effects of ACT 10 and the distrust it engendered. These associations prevent teachers from fully enrolling in the EDS network. Recall the earlier quote about DPI being laughed at. Because of the strength of the general sentiment of policymakers regarding teachers, it did not even make sense to many states’ leaders to have an evaluation system that did not involve substantial sanctions for poor performance.

For educators I observed, *evaluation* was largely synonymous with *accountability*, which in turn was associated with sanctions and rewards. For example, one CESA employee training district leaders in EE, who seemed strongly committed to a growth mindset, stated, “evaluation *is* about evaluation [accountability] but it’s mostly about growth.” Evaluation as (punitive) accountability is set as the default category against which growth evaluation must press, which is precisely the work that EDS is attempting.

Though the 2018 User Guide for EE begins with the statement that “Evaluation systems, implemented in isolation as an accountability or compliance exercise, will not improve educator practice or student outcomes” (Wisconsin Department of Public Instruction, 2018, p. 1), accountability is regularly discussed as a goal of EE. As one panelist, a representative of a Wisconsin teachers’ union, at a public forum hosted by the University of Wisconsin – Madison Educational Leadership and Policy Analysis department discussing Educator Effectiveness claimed to the audience of about twenty members of the public, “accountability is incredibly important.” Though the goal of the forum was to explain the usefulness of EE within a growth paradigm, and though the audience did not press the issue of accountability, because of the policy context, panelists continued to claim accountability could be maintained in the midst of a growth focus. Part of this insistence may have been to proactively rebuff the notion that teachers do not want to be held accountable, but whatever the intention, the panel repeated that accountability would persist even under the new growth model. Though another panelist, from DPI, explained that *rate, rank, remove* was never the intention, she conveyed understanding that trust would need to be built between DPI and teachers for the growth approach to take hold. A third panelist, from UW’s School of Education, undercut this, however, by recommending using EE to find “weak” teachers and “coach them out.” I came away from the forum feeling like EE wanted to have its cake and eat it too, so to speak.

Given the long history of teachers’ struggle for legitimacy, policies that maintain external and punitive accountability are likely to raise teacher defenses (Lortie, 1977; Rousmaniere, 2005). The sentiment of frustration with a system that adds work to busy teachers so they can prove their worth is reflected in one 36-year veteran teacher’s comment at the beginning of the EE panel, “Let’s hope educator effectiveness has a quick death.” She said this because she felt EE was “miserable” and “demoralizing.” Policies that are seen as questioning the professionalism of teachers have negative affective consequences (Achinstein & Ogawa, 2006) and lead to teachers becoming “defensive or protective” (Lasky, 2005, p. 901). One veteran teacher I interviewed assumed the increased attention on educator evaluations indicated that they were doing something wrong, adding, “I wasn’t evaluated for half my career; makes you wonder what’s changed,” and connected the need for

evaluation to the school's recent low report card score.⁸ This was echoed by another teacher who wondered, "If the kids did better, would this evaluation process be happening?"

The rate and rank style of DPI's *school* accountability information affects how teachers read DPI's intentions for *teacher* evaluation. Many educators, though they may appreciate the idea of a development and support-oriented evaluation process are positioned through a long history of delegitimization continued in present discourse to be apprehensive regarding evaluation policies and therefore are more inclined to begrudging compliance than full participation.

Just as the accountability-heavy discourse of teacher evaluation permeates Wisconsin, the materiality of accountability evaluation persists as well. One DPI employee explained,

Our vision was always that [EE] would be a professional development model, not an accountability model, but in the original design, it looked very similar to what other states were doing. So, even though our intention was not to use it like other states (and we were very clear that the results from it would not be used like other states), in the design it looked similar.

The current design is not substantially different from the iteration of the "origin design" that the DPI employee discussed, though to their credit, it is far from the 100% student test-score value-added pitch made by the Governor's office. The current design still maintains formative and summary cycles, asks teachers to collect artifacts on the same categories, and requires evaluators to enter their observations into an electronic database. It also still places teachers into one of four "levels of performance" (Wisconsin Department of Public Instruction, 2018, p. 8).

As one might imagine, attempting to change the TI without changing the material infrastructure or AI poses difficulties. In this section, I discuss three aspects of Wisconsin's current evaluation system that are more suited to an RRR system than a support and development one: total evaluation score, evaluation cycles, and artifact collection.

Though educator effectiveness law assumes the existence of a "total evaluation score" (Wisconsin Law 115.415, 2014), DPI has already shown its willingness to loosely interpret the law, as evidenced by its reimagining of the 50/50 score split as a two-dimensional space (recall Figure 1). Yet the evaluation system sorts teachers into four categories whose titles⁹ correspond roughly with the four categories into which student's standardized test scores can be placed. Given the "complex activity of teaching" organized into 76 elements by the Danielson Group (2013),¹⁰ I find it hard to imagine how the aggregated score adds any information on *how* to improve as a teacher or how an administrator should support a teacher though perhaps it can indicate *who* needs support.

The scores, however, are not overlooked. One third year teacher, in a district training on "growth mindset," explained that she was excited to be in her next reporting year, because though she "thought [she] did pretty well," she "had to rate [her]self a *two*" for her first year. Though she would have scored herself as a *three*, her principal explained that all first year teachers must be *twos*. The teacher went on to explain that a *two* is perceived as a bad thing and indicates the teacher did

⁸ In Wisconsin each school and district receives an annual score and rating on a five-star system (Office of Educational Accountability, 2016). From my observations the school report cards affected teachers and principals more deeply than their individual evaluations.

⁹ The Effectiveness Project labels these *Unacceptable*, *Developing/Needs Improvement*, *Effective*, and *Distinguished*. Educator Effectiveness uses *Unsatisfactory*, *Basic*, *Proficient*, and *Distinguished*. The Report Card uses *Below Basic*, *Basic*, *Proficient*, and *Advanced*.

¹⁰ About two-thirds of districts use Educator Effectiveness based on the Danielson Framework and the remainder use the Effectiveness Project based on the Stronge (Tonneson, 2012) framework. Space does not permit a discussion of the nuances, but for this analysis they are virtually equivalent.

something wrong. She related, “I remember being devastated about that.” On the other hand, though one may sometimes reach a *four*, and here she repeated verbatim a phrase I had on a separate visit heard the superintendent use, “Three is where you want to live.” In this way, the total score, especially with the category names, undermines the sentiment that the evaluations are primarily for growth rather than accountability for performing at a proficient level. Though my participant from DPI claimed the numbers were no longer going to be used, I have found no evidence so far in official documents or observation of practice of such a change.

In a similar vein, the notion of an evaluation cycle with a *Summary Year* (previously named *Rating Year*) and *Supporting Years* makes less sense for a growth model than an accountability model. This is another example of reworking the language of the evaluation while minimally affecting the evaluation process. The teacher excited to become a *three* mentioned collecting artifacts to get ready for her summary year. Educators are supposed to collect “artifacts that will provide adequate evidence for the Summary Year evaluation” (Wisconsin Department of Public Instruction, 2018, p. 31). According to the guide, “artifacts contain evidence of certain aspects of professional practice that may not be readily visible through an observation” (p.32). Though this evidence collection is intended to provide additional data to drive decisions for support and improvement, it relies on the exact same processes and products as it would if teachers were proving their worth. As a teacher in Waterside district stated, “It’s sad to have to prove I’m doing a good job.” Others noted the collection of artifacts is frustrating extra work. By asking teachers to perform the same tasks under the growth model as they did under the accountability model, the evaluation remains tied to all the affective baggage of the former.

The 2019 Policy Guide supports using EE “as one source of evidence ... to inform compensation decisions at the individual level,” though warned some uses of the scores may cause “invalid or unreliable decisions” (Wisconsin Department of Public Instruction, 2019, p. 24). Even when DPI guides evaluators away from making high-stakes decisions on the basis of EE scores, the design of the system and the discursive context in which evaluators find themselves make it easy to use EE for building a case to remove teachers. As one principle expressed, one can use EE to “coach them out, or better yet, have them come to the conclusion themselves that it is ‘time to go’.” One district puts low-performing teachers on a “growth plan,” which includes a pay freeze. A teacher in the district explained how the “growth plan” means “I’m terrible” and “the administrator wants to get rid of me instead of ‘here’s how we’re going to help’.” The district manages to undercut the transition to the new development TI by associating the term *growth* with a punitive system. The principal of the school explained to me the challenge of the growth mindset. Though he reported being very much in favor of it and, as I observed, used teacher evaluation mostly for that purpose, he also believed his role as principal required staffing decisions. As such he used EE to build a case against someone who was “not a very good teacher,” but had been at the school a long time. These practices blur the lines between evaluation for growth and for rating. Because of this blurring, the principal explained the challenge for teachers, who wonder whether the principal, in a given moment, is “coaching or evaluating,” where *evaluating* carries the possibility of punitive action. He proposed the usefulness of a teaching coach that was confidential so teachers could trust that person enough to be vulnerable and improve. Sadly, the district did not have such a system in place.¹¹

Other system constraints also limited the transformation from RRR to the growth model. The overworked¹² teachers and administrators of the average school struggle to accomplish the

¹¹ Reading ANT perspectives across literature on data use (for example, Coburn & Turner, 2011; Marsh, 2012) may be a fruitful way to understand some of the challenges expressed here.

¹² Busyness and exhaustion were nearly ubiquitous self-descriptions among educators I observed.

reflection and feedback necessary to cultivate growth. One principal discussed this with the EE trainer, saying, “Our plates are full and there’s not much we can take off.” Numerous teachers expressed how rushed, late, or missing feedback from observations hindered their acceptance of the evaluation process as a legitimate system of improvement.¹³ Administrators I talked with believe they cannot possibly observe classrooms more than they already do, and teachers believe the small number of observations provides too little feedback for meaningful reflection and improvement, especially the case during *support* years, where the evaluator might only visit part of one class period once or twice that year. Therefore, the system, as understood by educators, does not actually support growth.

Considering the material-discursive collection and calculation components of the educator evaluation apparatus, a great number of components remain tied to an RRR theory and process of improvement rather than a growth theory and process.

Entangled Apparatus of Improvement as Contextualized Evaluation

No evaluation is an island. As discussed above DPI has been struggling to get educators fully enrolled in the new teacher evaluation system in part because the larger material-discursive apparatus on which evaluation relies is not (and cannot be) entirely new. Though they have managed to detach from some RRR-style components, they remain entangled in others. Below I discuss how the ANT framing of evaluation as material-semiotic assemblages can help make sense of implementation challenges. However, after addressing the problems Wisconsin’s EDS faces, I return to the promising avenues of teacher evaluation they are pursuing.

Imagining evaluation as a tool that can be constructed to accomplish a particular purpose, which can simply be set aside in favor of a new tool when the goal changes, like replacing a hammer with a flashlight, the metaphors used by my participants, does not deal with the challenges facing DPI’s EDS. Instead, if schooling is a vast network, a tapestry of connections, then replacing an evaluation policy involves the tedious work of disentangling and disconnecting, and then reconnecting, multitudes of threads, including material, discursive, and affective relationships.

The entanglement of previous and contemporary evaluations with future ones necessitate that no evaluation, in its logic (TI) and mechanisms (AI), is understood and implemented on its own terms. An evaluation is not solely determined by its expressed purpose but by the interactions among its mechanisms of evaluating and the constellation of policies, discourses, and material constraints which contextualize the evaluation that encourage actors to interact with it in particular ways. As sensemaking apparatuses¹⁴ evaluations are entangled with the sense already being made, both extra-evaluationally through the general discourses surrounding the evaluated object (ie, teachers) and inter-evaluationally through the previous evaluations and contemporary competing evaluations, through evaluations measuring the same object, or parallel objects, and those measuring analogous objects in a separate sphere. As partially material actor-networks they are entangled with various human and nonhuman, including systemic and infrastructural, material constraints and affordances. Such a Gordian knot has no Alexander.

As one principal explained, changing evaluation purposes and shifting practice is not straightforward because of “all that baggage” educators bring from their interactions with other evaluations. ANT-influenced evaluation theory can help make sense of materially and discursively

¹³ Administrators made similar comments regarding the testing and reporting cycle of the Report Card Apparatus.

¹⁴ I am drawing on Mark, Henry, and Julnes (2000) who describe all evaluation as assisted sensemaking.

entangled evaluation policies and processes. I follow the concerns of Preskill (2012) and Alkin (2012a), evaluation theorists who draw attention to contexts with which an evaluator should be familiar, including stakeholder values, the program's theory of action¹⁵, social, political, and historical contexts, and evaluation and evaluator context. My approach, using ANT sensibilities, illuminates the historical and contemporary contextualization of the policyscape¹⁶ (evaluation and evaluation-adjacent policies) and helps explain why an evaluation presumably designed aligned to stakeholder desires does not automatically engender complete investment. Much evaluation theory discussing context focuses on designing evaluations stakeholders want (Alkin, 2012b). I use ANT to further explore making evaluations meaningful to stakeholders by analyzing the historical and contemporary relationships influencing the way they make meaning of the evaluation.

Under an ANT paradigm, an evaluation can never be constructed and enacted narrowly, without concern for surrounding evaluations, other discourses regarding the object of evaluation, and the material conditions of the evaluation. These material-discursive actors shape the construction, perception, and use of the new evaluation. Because of this, a new TI with its attendant AI requires considerable work to disassociate from the previous attachments and create new associations, discursively, affectively, and materially.

In order to effectively implement a new policy, evaluation creators should consider the space the evaluation takes up, the reality it imposes, the effect it has on the actions, decisions, and decision-space (the possibility for particular choices) of those that interact with it. They must recognize the ways in which evaluation systems associate themselves with material-discursive-affective networks in order to reify themselves and their constructed objects, and how they are sometimes unable to accomplish that task when competing actor-networks oppose them. To rearticulate into ANT terms the axioms of Realist evaluation (Pawson & Tilley, 1997) regarding context, evaluation should answer two questions: 1. What apparatuses does a program assemble, and how do those apparatuses interact with tangent apparatuses? 2. From what material and discursive actors are the change-apparatuses composed and how are those actors mobilized among interrelated networks?

Considering the material, discursive, and affective associations in which a new evaluation process is necessarily entangled provides a way to understand participants' reservations for and readings of a system they may otherwise favor. Accounting for any previous or current alternative evaluations of the same objects, evaluations of similar or related objects, discourses from which sense about the objects, specifically or as a class, is already being made, and the material conditions which also shape the possibilities of understanding and using the evaluation apparatus will benefit the construction of an evaluation system.

Wisconsin Educator Development and Support: Partially Realized Promise

Given the research that challenges the benefit of high stakes accountability, it seems appropriate that Wisconsin push in a different direction (eg, Amrein-Beardsley & Collins, 2012; Papay, 2011). However, Wisconsin's teacher evaluation remodel suffers from two connected problems. First, its change in theory of improvement (TI) has not been matched by a radical restructuring of the apparatus of improvement (AI). Second, the changes that have been made are

¹⁵ I use theory of improvement rather than theory of action simply to highlight the action of evaluation as improvement specifically.

¹⁶ I borrow this term from Carney (2009).

often not read as authentic because of the broader context in which Wisconsin educators find themselves.

For DPI, improvement officially means more fully realizing the goal of “Every Child a Graduate, College and Career Ready,” a subgoal of which is “Ensuring our educators are both inspired and empowered to teach every student” (*Empowered Educators*, 2017). DPI through EDS has opted for a growth model, which requires teachers and administrators to cooperatively and reflectively identify practices and resources that continuously increase the capacity and improve the ability of teachers to increasingly develop students to be ready for college and careers.

Matching the Apparatus with the Theory: Evaluation Alignment

Ideally, having established a theory of improvement, the evaluative apparatus would be constructed to align with the theory and the purpose(s) it entails.¹⁷ That construction requires the determination of criteria or qualities to be considered, standards or degrees against which the quantity and quality of the criteria can be compared, measures by which the quantity and quality can be determined, and the analysis and synthesis of those determinations. The type of data collected/constructed, the process by which it is, and the analysis and synthesis of that data should reasonably differ according to the TI because of the use to which the results will be put. Crucially, the growth theory promoted by DPI requires a different relationship between educators and the evaluation. It is important that educators go, in a common refrain among participants, “beyond compliance.” As an EE trainer mentioned to a group of administrators, “We want them to *want* it [to use the evaluation to improve].” Therefore, part of the evaluation process is preparing educators to receive the feedback with a growth mindset, making sure “they’re ready to hear,” in the words of one evaluator. Achieving this requires addressing existing remnants of the system connected to the RRR model, such as the four-point rating system and pay freezes for teachers on a *growth plan*, which poorly fit the new evaluation intention. One way to improve enrollment (in the ANT sense) into the new teacher evaluation system is scrub the vestiges of the old paradigm from the apparatus, rather than trying to use an old system in new ways for which it is not well-designed.

Contextualized Interactions: Shifting Expectations

Additionally, even when the theory and apparatus seem to match, due to the historical and contemporary policy context, educators do not always fully enroll. In instances when the evaluative apparatus does align to the TI, that is, it is designed to assist in making sense of particular phenomena as directed by the improvement theory, it may not be *perceived* as aligned. Previous iterations of teacher evaluations in Wisconsin, teacher evaluations in others states, related evaluations such as school evaluations via the state report cards, as well as the infrastructural/material and ideological/discursive phenomena that made those possible can be important associations in determining how actors make sense of and decide how to interact with an evaluation. In this particular case, the generic national discourse, which largely disparages teachers, is present in NCLB, whose waiver initiated educator evaluation in Wisconsin. Additionally, Act 10 continues to provide evidence to teachers that the state is hostile. Therefore, even if EDS were designed solely with growth components, it would be viewed with suspicion. Further, the material constraints, most noticeably *a lack of time*, which hinder the purposeful feedback cycles and reflection necessary for intentional growth, indicate to teachers that the stated purpose could not possibly match the actual purpose.

¹⁷ Though as stated previously, Wisconsin’s evaluation policy still suffers from some mixed motives, hanging onto aspect of an accountability TI in addition to unchanging AI.

For teacher evaluation systems, such as EDS, taking into account the decades-long struggle for legitimacy by teachers and the general deprofessionalization of teachers through other federal and state policies is necessary to understand educators' approach to evaluation. To their credit, DPI seems aware of this and has aggressively conveyed their support of the value of teachers and improvement through growth both to Wisconsin teachers and to education departments in other states.

To progress DPI, the CESAs, and districts will need to care for the evaluation environment. For example, DPI and districts could encourage legislation that counteracts the negative effects of Act 10. Districts, with the support of DPI, could adjust the working conditions of teachers and administrators to provide more opportunity for peer and administrator observation, feedback, and reflection. DPI's educator evaluation office could encourage other offices, especially the Report Card, to adopt school and district evaluations more suited to growth rather than comparison. Rather than primarily attempting to enroll educators directly through the messaging campaign, more attention could be given to creating the conditions that would favor full educator enrollment.

Twofold Potential: Development and Support

Despite focusing much of the article on challenges to implementing new evaluation policy-practices, I believe Wisconsin provides an excellent case of a state committed to an alternative to high-stakes value-added teacher evaluation. I reinterpret the DPI terms *development* and *support* to frame two key aspects of Wisconsin teacher evaluation I find most promising.

By shifting to a development, or growth, theory of improvement Wisconsin has drawn on the teaching and learning ethos of the schooling enterprise. This is a bold deviation from rate, rank, remove/replace strategies. Wisconsin is shifting from using teacher evaluation to prove which teachers are worthy of continued employment to using it for determining specific ways teachers can improve their practice. In this model, teacher evaluation is used to change how a teacher is teaching rather than change who is teaching. Development and growth-based evaluation is a promising first step.

As a second step, administration and staff at various locales I observed are considering holistic systems of support that adjust the way teachers are evaluated and the way administrators approach evaluation. Holistic improvement does not place the onus for performance solely on the teacher but is distributed among various components of schooling. A DPI employee challenged the reasonability that value-added evaluation can "accurately attribute" student improvements to particular teachers within the cooperative, integrated environment of a school. A school principal explored teaching assignment changes and reorganization of facilities to improve teacher quality. The school improvement specialists from the various CESAs are leading the way by viewing the various evaluation processes each as "a slice of a system of improvement." The holistic approach, which does not restrict the results of teacher evaluation to transforming the teacher, informs changes to other human and nonhuman actors within the school system (assignments, facilities, and resources) to make the most of teachers' particular skills sets to improve the teacher's ability to actualize the school's goals.

As I have shown, by formulating evaluation as a material-discursive apparatus composed of a network of actors partially mobilized in support of its theory of improvement, one can consider the entanglements among evaluations and between an evaluation and the broader environment. The positive changes attempted in Wisconsin and those highlighted in the other works in this issue do not occur in a vacuum. Each policy-practice has particular affordances and constraints determined by the policy-practices preceding and the tangent policy-practices and attending discourses. The

entanglements require work to unravel but the unraveling (and subsequent alternative attaching) needs to be accomplished in order to realize the evaluation change we desire.

References

- Achinstein, B., & Ogawa, R. (2006). (In) Fidelity: What the resistance of new teachers reveals about professional principles and prescriptive educational policies. *Harvard Educational Review*, 76(1), 30–63. <https://doi.org/10.17763/haer.76.1.e14543458r811864>
- Alkin, M. C. (2012a). Context sensitive evaluation. In *Evaluation roots: A wider perspective of theorists' views and influences* (pp. 283–292). Sage.
- Alkin, M. C. (2012b). *Evaluation roots: A wider perspective of theorists' views and influences*. Sage.
- Amrein-Beardsley, A., & Collins, C. (2012). The SAS Education Value-Added Assessment System in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives*, 20(12). <https://doi.org/10.14507/epaa.v20n12.2012>
- Anderson, B. (2017). *Encountering affect: Capacities, apparatuses, conditions*. Routledge. <https://doi.org/10.4324/9781315579443>
- Blige, M. J. (2005). *Baggage: Vol. The breakthrough*. Geffen.
- Carney, S. (2009). Negotiating policy in an age of globalization: Exploring educational “policyscapes” in Denmark, Nepal, and China. *Comparative Education Review*, 53(1), 63–88. <https://doi.org/10.1086/593152>
- Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research & Perspective*, 9(4), 173–206. <https://doi.org/10.1080/15366367.2011.626729>
- Cochran-Smith, M., & Lytle, S. (2006). Troubling images of teaching in No Child Left Behind. *Harvard Educational Review*, 76(4), 668–697. <https://doi.org/10.17763/haer.76.4.56v8881368215714>
- Danielson Group. (2013). *Danielson Group » The Framework*. <http://danielsongroup.org/framework/>
- Daston, L. (2000). The coming into being of scientific objects. In *Biographies of scientific objects* (pp. 1–14). University of Chicago Press.
- Foucault, M. (1991). Questions of method. In *The Foucault effect: Studies in governmentality* (pp. 73–86). University of Chicago Press.
- Foucault, M., Davidson, A. I., & Burchell, G. (2008). *The birth of biopolitics: Lectures at the Collège de France, 1978-1979*. Springer.
- Kelleher, J. (2011, February 23). Inside Capitol, Wisconsin protesters create a city. *Reuters*. <https://www.reuters.com/article/us-wisconsin-protests-camping-idUSTRE71M0A220110223>
- Kumashiro, K. K., Ayers, W., & Quinn, T. (2012). *Bad teacher! How blaming teachers distorts the bigger picture*. Teachers College Press.
- Lasky, S. (2005). A sociocultural approach to understanding teacher identity, agency and professional vulnerability in a context of secondary school reform. *Teaching and Teacher Education*, 21(8), 899–916. <https://doi.org/10.1016/j.tate.2005.06.003>
- 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.
- Lortie, D. C. (1977). *Schoolteacher: A sociological study*. University of Chicago Press.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding and improving policies and programs*. Jossey-Bass Pfeiffer.

- Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record*, 114(11), 1–48.
- Milwaukee Public Schools. (2014). *Educator effectiveness system guide for teachers*.
https://www.nctq.org/dmsView/Attachment_Educator_Effectiveness_System_Guide
- Office of Educational Accountability. (2016). *Accountability Report Cards: Technical Guides 2015-16*. Wisconsin Department of Public Instruction.
<https://dpi.wi.gov/sites/default/files/imce/accountability/pdf/Report%20Card%20Technical%20Guide%202015-16.pdf>
- Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163–193.
<https://doi.org/10.3102/0002831210362589>
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.
- Payne, C. M. (2008). *So much reform, so little change: The persistence of failure in urban schools*. Harvard Education Press.
- Preskill, H. (2012). The transformational power of evaluation. In M. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influences* (pp. 323-333). Sage.
- Prior, L. (2003). *Using documents in social research*. Sage. <https://doi.org/10.4135/9780857020222>
- Rousmaniere, K. (2005). In search of a profession: A history of American teachers. *Portraits of a profession: Teaching and teachers in the 21st century* (pp. 1–26). Praeger.
- Stronge, J. H. (2012). *Stronge Teacher Effectiveness Performance Evaluation System*. <http://mnprek-3.wdfiles.com/local--files/teacher-effectiveness/TEPES%20-%20Stronge.pdf>
- Sutton, M., & Levinson, B. A. (2001). *Policy as practice: Toward a comparative sociocultural analysis of educational policy*. Greenwood Publishing Group.
- Tonneson, V. C. (2012). *James H. Stronge, Ph.D. Stronge and Associate Educational Consulting, LLC*. 178.
- U.S. Department of Education. (n.d.). *Race to the Top – District guidance and frequently asked questions*.
<http://www2.ed.gov/programs/racetothetop-district/faqs.pdf>
- U.S. Department of Education. (2013). *Wisconsin ESEA Flexibility Request*.
<http://esea.dpi.wi.gov/sites/default/files/imce/esea/pdf/waiver-final.pdf>
- Welch, J. (2005). The vitality curve. *Leadership Excellence*, 22(9), 4.
- Wisconsin Act 10, Pub. L. No. 2011 Wisconsin Act 10 (2011).
<https://docs.legis.wisconsin.gov/2011/related/acts/10>
- Wisconsin Act 166. (2011b). <https://docs.legis.wisconsin.gov/2011/related/acts/166>
- Wisconsin Department of Public Instruction. (2014). *Teacher Evaluation Process Manual 2014-2015 Version 4*.
<https://wi01919813.schoolwires.net/cms/lib/WI01919813/Centricity/Domain/113/Teacher%20Evaluation%20Process.pdf>
- Wisconsin Department of Public Instruction. (2016, January 22). *Title II Programs*.
<https://dpi.wi.gov/title-ii>
- Wisconsin Department of Public Instruction. (2017, January 3). *Every Teacher a Leader*.
<https://dpi.wi.gov/teacher-leader>
- Wisconsin Department of Public Instruction. (2017, July 3). *Empowered Educators*.
<https://dpi.wi.gov/statesupt/every-child-graduate/empowered-educators>
- Wisconsin Department of Public Instruction. (2017, July 27). *WI Educator Development & Support*.
<https://dpi.wi.gov/educator-development-support>
- Wisconsin Department of Public Instruction. (2017, July 27). *Resources for Professional Learning*.
<https://dpi.wi.gov/educator-development-support/support-development/professional-learning>

Wisconsin Department of Public Instruction. (2018). *Wisconsin Educator Effectiveness System: User Guide for Teachers, Teacher Supervisors, and Coaches*.

<https://dpi.wi.gov/sites/default/files/imce/ee/pdf/teacherprocessmanual.pdf>

Wisconsin Department of Public Instruction. (2019). *Wisconsin Educator Effectiveness System Policy Guide*.

Wisc. Stat. Ch. 118 §42 Low-performing school districts and schools Pub L. No. WI Law 115.415 (2014). <https://docs.legis.wisconsin.gov/statutes/statutes/118/42>

About the Author

Jake Malloy

University of Wisconsin - Madison

jake.malloy@wisc.edu

Jake Malloy, PhD, is a recent graduate of the University of Wisconsin – Madison’s Educational Policy Studies department. His research interests include Evaluation Theory, Narrative Network Analysis, Science and Technology Studies, particularly of the poststructural and Actor-Network Theory variety.

About the Guest Editor

Audrey Amrein-Beardsley

Arizona State University

audrey.beardsley@asu.edu

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.

SPECIAL ISSUE
Policies and Practices of Promise
in Teacher Evaluation

education policy analysis archives

Volume 28 Number 60

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

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: **Melanie Bertrand, David Carlson, Lauren Harris, Eugene Judson, Mirka Koro-Ljungberg, Daniel Liou, 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
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

David E. DeMatthews
University of Texas at Austin

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

A. Chris Torres
Michigan State 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

Miri Yemini
Tel Aviv University, Israel

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, 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