

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 63

April 13, 2020

ISSN 1068-2341

**Better Integrating Summative and Formative Goals in the
Design of Next Generation Teacher Evaluation Systems**

Timothy G. Ford

University of Oklahoma



Kim Kappler Hewitt

University of North Carolina at Greensboro
United States

Citation: Ford, T. G., & Hewitt, K.¹ (2020). Better integrating summative and formative goals in the design of next generation teacher evaluation systems. *Education Policy Analysis Archives*, 28(63). <https://doi.org/10.14507/epaa.28.5024> This article is part of the special issue, *Policies and Practices of Promise in Teacher Evaluation*, guest edited by Audrey Amrein-Beardsley.

Abstract: In current teacher evaluation systems, the two main purposes of evaluation—accountability/goal accomplishment (summative) and professional growth/improvement (formative)—are often at odds with one another. However, they are not only compatible, but linking them within a unified teacher evaluation system may, in fact, be desirable. The challenge of the next generation of teacher evaluation systems will be to better integrate these two purposes in policy and practice. In this paper, we integrate the frameworks of Self-determination theory and Stronge’s Improvement-Oriented Model for Performance

¹ The authors contributed equally to the work.

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

Facebook: /EPAAA

Twitter: @epaa_aape

Manuscript received: 9/3/2019

Revisions received: 2/3/2020

Accepted: 2/20/2020

Evaluation. We use this integrated framework to critically examine teacher evaluation policy in Hawaii and Washington, D.C.—two distinctly different approaches to teacher evaluation—for the purposes of identifying a set of clear recommendations for improving the design and implementation of teacher evaluation policy moving forward.

Keywords: teacher evaluation; accountability; educational improvement; formative and summative assessment; self-determination theory; qualitative methods

Mejor integración de objetivos sumativos y formativos en el diseño de los sistemas de evaluación docente de próxima generación.

Resumen: En los sistemas de evaluación docentes actuales, los dos propósitos principales de la evaluación: accountability / logro de objetivos (sumativo) y crecimiento / mejora profesional (formativo), a menudo están en desacuerdo entre sí. Sin embargo, no solo son compatibles, sino que vincularlos dentro de un sistema unificado de evaluación docente puede, de hecho, ser deseable. El desafío de la próxima generación de sistemas de evaluación docente será integrar mejor estos dos propósitos en la política y la práctica. En este documento, integramos los marcos de la teoría de la autodeterminación y el modelo orientado a la mejora de Stronge para la evaluación del desempeño. Utilizamos este marco integrado para examinar críticamente la política de evaluación docente en Hawai y Washington D.C., dos enfoques claramente diferentes para la evaluación docente, con el propósito de identificar un conjunto de recomendaciones claras para mejorar el diseño y la implementación de la política de evaluación docente en el futuro.

Palabras clave: evaluación docente; accountability; mejora educativa; evaluación formativa y sumativa; teoría de la autodeterminación; métodos cualitativos

Melhor integração de objetivos somativos e formativos no projeto de sistemas de avaliação de professores da próxima geração

Resumo: Nos atuais sistemas de avaliação de professores, os dois principais objetivos da avaliação - accountability / realização de objetivos (sumativo) e crescimento / aprimoramento profissional (formativo) - geralmente estão em desacordo. No entanto, eles não são apenas compatíveis, mas vinculá-los a um sistema unificado de avaliação de professores pode, de fato, ser desejável. O desafio da próxima geração de sistemas de avaliação de professores será integrar melhor esses dois propósitos nas políticas e práticas. Neste artigo, integramos as estruturas da teoria da autodeterminação e o modelo orientado para a melhoria de Stronge para avaliação de desempenho. Utilizamos essa estrutura integrada para examinar criticamente a política de avaliação de professores no Havaí e Washington, DC - duas abordagens distintas para a avaliação de professores - com o objetivo de identificar um conjunto de recomendações claras para melhorar o design e a implementação da política de avaliação de professores.

Keywords: avaliação de professores; accountability; melhoria educacional; avaliação formativa e sumativa; teoria da auto-determinação; métodos qualitativos

Introduction

Race to the Top (RttT) spurred states to rethink the ways in which they evaluate teacher performance. However, by many accounts, these efforts have largely failed to produce marked improvements in teaching and learning (Firestone & Donaldson, 2019; Hallinger et al., 2014; Lavigne & Good, 2019, 2020). While the Every Student Succeeds Act (ESSA) has provided some much-needed flexibility, changes in teacher evaluation systems have remained meager at best, though there has been a marked trend away from the accountability-driven rhetoric that typified RttT systems toward language which affirms the importance of providing teachers with meaningful feedback to improve their practice (Close, Amrein-Beardsley, & Collins, 2018).

Similarly, there has been some shift away from reliance on value-added measures (VAMs) as the preponderant measure of teacher performance in state systems. A number of states, including Florida and Wyoming—where the practice was first used—have eliminated the requirement that teacher evaluations rely on student test scores, and this “about-face is picking up speed” (NEPC, 2019, p. 1). Some of this shift can be attributed to abundant evidence of the substantial validity and reliability issues associated with these measures (Amrein-Beardsley, 2014; Bitler, Corcoran, Domina, & Penner, 2019; Close et al., 2018; Lavigne & Good, 2020; Sloat et al., 2018). Furthermore, scholarship exploring the use of multiple measures in teacher evaluation has provided guidance that states and local policy actors can use in selecting alternative measures of teacher performance (Grissom & Youngs, 2016), as well as in reexamining the weighting or use of various measures to provide a more balanced assessment of teacher performance (Kane et al., 2014). Despite substantial monetary investments by states and districts, one area in which research and guidance remains sparse is with respect to encouraging and supporting teachers’ use of performance feedback to improve their practice (Firestone & Donaldson, 2019). If teacher evaluation is truly to function as a formative—not just a summative—tool for improvement, we must more carefully consider how teachers view the information generated from evaluation and the conditions under which they are most likely to use it (Firestone & Donaldson, 2019; Ford, 2018; Pauffer & Sloat, 2020).

Social scientists have long noted the inherently psychological dimensions of feedback and the conditions under which it succeeds in changing (or fails to change) mindsets and behavior (Hattie & Timperley, 2007; Ryan & Deci, 2017; Semke, 1984). Recent innovations in evaluation and feedback in the business sector in particular reflect thoughtful consideration of the psychological dimensions of evaluation to ensure that these efforts result in meaningful changes to practice (e.g., Buckingham & Goodall, 2019). These insights center on the interpersonal nature of feedback, the limitations of praise, and the science of how the brain receives information and how it is integrated into current cognitive patterns. For example, understanding the critical role of timing in the feedback process suggests that feedback should be more frequent and embedded in the performance process, so that leaders and managers can point out successes in real time (Buckingham & Goodall, 2019; Lavigne & Good, 2020). Furthermore, subtle changes in language use can cue better receptivity of feedback on the part of the evaluatee, eschewing judgment while pointing out critical areas for improvement.

To some degree, awareness of the inherent challenges in providing meaningful feedback has permeated pre-service teacher mentoring practice (e.g., Akcan & Tatar, 2010; Copeland, 2010; Lu, 2010; Vasquez, 2004; Waite, 1993), though much of this work remains theoretical and requires stronger empirical support. Good pre-service teacher mentorship often entails working with student teachers in a critical examination of their classroom practice, but this is an inherently threatening prospect for the teacher being evaluated (Vasquez, 2004). Here again, the teacher mentor/supervisor’s ability to productively negotiate these interactions is critical, and the

supervisor's use of language and their ability to address inherent power dynamics in these exchanges is particularly so (Copland, 2010, 2012; Waite, 1993). Insights into how evaluatees receive and respond to feedback have yet, however, to be effectively integrated into and/or used to inform innovations in teacher evaluation that move beyond high-stakes systems based largely on student test scores.

Some barriers to effective teacher evaluation practice lie beyond the micropolitics of providing and using feedback. Scholars have argued that some barriers to quality, equitable, and beneficial teacher evaluation practice are structural in nature and no change to practice will effectively remove them. For example, critical teacher evaluation scholars have noted the neoliberal nature of "new" teacher evaluation policy (Holloway & Brass, 2017; Holloway, Sørensen, & Verger, 2017). Neoliberalism emphasizes the efficiencies brought about by market competition, deregulation, and a more explicit focus on the measurement and tracking of performance outcomes for the purposes of incentivizing improvement (i.e., performance management). Critical teacher evaluation scholars argue that high-stakes, top-down, teacher evaluation reflects an increasing prioritization of the needs of the educational organization for control and certainty over the needs of teachers to feel supported in their learning and development as practitioner-professionals (Ford, Urick, & Wilson, 2018; Holloway & Brass, 2017; Holloway, Sørensen, & Verger, 2017).

Although many aspects of teacher evaluation systems have ostensibly been improved over recent years, numerous studies reveal that some old, persistent challenges remain: lack of support and/or guidance in the use of teacher evaluation results (Amrein-Beardsley & Collins, 2012; Ford, Van Sickle, Clark, Fazio-Brunson, & Schween, 2017); lack of validity and/or reliability (either real or perceived) of evaluation results (Darling-Hammond et al., 2012; Ford, Van Sickle, & Fazio-Brunson, 2016; Kappler-Hewitt, 2015; Jiang et al., 2015; Longo-Schmid, 2016; Reddy et al., 2017); and evidence of increase in stress and anxiety as a result of engaging in teacher evaluation (Ford et al., 2017; Hewitt, 2015; Holloway & Brass, 2017; Ingersoll et al., 2016; Jiang et al., 2015). While disparate, these recent findings of teacher evaluation in practice have one thing in common: They reflect a series of consequences that might plausibly emerge from evaluation systems which do not, to a large enough extent, prioritize the needs of teachers or, at the very least, fail to maximize the benefits they receive from participation.

Current Study

In current teacher evaluation systems, as with other policies, tensions between the two main purposes of evaluation—accountability/goal accomplishment (summative) and professional growth/improvement (formative)—are evident (Firestone, 2014; Firestone & Donaldson, 2019; Schildkamp et al., 2017). In practice, this tension often results in one of these goals being favored, to the detriment of the other. However, these two purposes are not only compatible, but linking them within a unified teacher evaluation system may, in fact, be desirable (Firestone, 2014). The challenge of the next generation of teacher evaluation systems will be to better integrate and balance these two purposes in policy and practice. This project is important, because we know that teacher evaluation policies that rely primarily on test-based measures trigger teacher stress, which decreases teacher satisfaction and increases teacher turnover intention (Ford et al., 2018; Ryan et al., 2017). Churn created by high teacher turnover "undermines student achievement and consumes valuable staff time and resources. It also contributes to teacher shortages throughout the country" (Learning Policy Institute, 2017, para. 1). Moreover, teachers who do not perceive feedback to be useful are generally less satisfied with their work (Smith & Kubacka, 2017).

To develop teacher evaluation systems and processes which are more balanced in terms of their consideration of both individual and collective needs, we need a framework which attempts to reconcile these two approaches. This means we need: a) A theory which can help to explain how performance feedback can be used to motivate individuals to improve and under what conditions

such use is optimized (Ford, 2018); and b) A conceptual framework that addresses the purposes of evaluation and ways in which multiple purposes can be met through a unified evaluation system.

To these ends, we integrate the frameworks of Self-determination theory (Ryan & Deci, 2000, 2017) and Stronge's (1995) Improvement-Oriented Model for Performance Evaluation. We use this framework to critically examine teacher evaluation policy in Hawaii and Washington, D.C.—two distinctly different approaches to teacher evaluation. The research questions which undergirded our examination of these two different evaluation approaches were: 1) What are the discursive characteristics of DC IMPACT and Hawaii evaluation policy and how do they illustrate the various dimensions/tensions present in our integrated framework? 2) What can we learn from studying the similarities/differences in these two systems to inform the development of more promising, balanced, and equitable teacher evaluation policy and practice moving forward?

Theoretical Framework: Self-Determination Theory

In this educational climate, teachers and leaders are bombarded with various sources of information, data, and/or knowledge that they could potentially use for improvement (Danna, 2004; Hill & Rapp, 2012; Mokhtari, Thoma, & Edwards, 2009). The information generated as a result of teacher evaluation is but one of these sources. Thus, if you want teachers to use their evaluation feedback for improvement, they have to *value* the feedback and see its potential for improvement. This entails not trying to control individual behavior but instead creating the conditions that activate intrinsic interest in the feedback, as well as removing some important barriers to feedback receptivity and use (Ford, 2018; Schildkamp et al., 2017). Self-determination theory (SDT) is a broad social-cognitive theory of human motivation and personality development. A foundational assumption of SDT is that human beings are innately driven to seek out new information and experiences, use these experiences to learn and grow, and, doing so, strive to better integrate themselves into the larger social structures in which they are embedded. This assumption often stands in direct opposition to the implicit assumptions of high-stakes evaluation and assessment policies, whose consequences for poor or good performance—such as termination, punishment, or increased teacher pay—assume that what best spurs human improvement derives from external, not internal, sources (Ford, 2018, Ryan & Brown, 2005).

Thus, the question for SDT theorists is not, “Can the individual be motivated?” or “Will an individual use the feedback they are given?”; rather it is: “Under what conditions will their intrinsic interest in an event or goal be activated?” The SDT concept of *functional significance* states that the effects of external events on human motivation hinge on the psychological meaning they have for the recipient, and this concept provides a framework for predicting the extent to which individuals will integrate new events, experiences, or information to which they are exposed (Ryan & Weinstein, 2009). For example, events have a positive effect on motivation when they have *informational* significance—that is, when they provide feedback that helps learners become more effective but without interfering with autonomous action or decision making (Deci & Ryan, 2000). Informationally significant feedback, for example, would be feedback generated from assessments teachers themselves chose or find beneficial in their daily practice (Farrell & Marsh, 2016), or lesson observations from a colleague whom they see as expert in the areas in which they would like to improve. When events or feedback are *controlling*, on the other hand, individuals often respond by exerting the least amount of effort needed to gain reward or avoid punishment (Ryan & Weinstein, 2009).

Finally, events have *amotivating* significance when the recipient feels overwhelmed by feedback, for instance when it is highly negative (Ryan & Brown, 2005). Similarly, individuals will

tend to exhibit amotivational responses to events that they perceive to be out of their control, irrelevant to their immediate work, or that contain no clear action steps (Deci & Ryan, 2000; Ryan & Weinstein, 2009). As Ryan and Deci (2017) point out, even subtle differences in the ways in which an event is introduced can have profound implications for how salient the information will be to the recipient. For example, even if feedback measures and systems are well-designed and intentioned with recipients in mind, something as simple as adding consequences or rewards to the feedback, or having a principal (who is in a position of authority) versus a fellow teacher deliver the feedback, can lead to vastly different teacher responses (Ford et al., 2018).

Conceptual Framework: Stronge's Improvement-Oriented Model for Performance Evaluation

Citing the two main purposes of evaluation—accountability/goal accomplishment (summative) and professional growth/improvement (formative)—Stronge (1995) argued that these purposes are not only compatible but that linking them within a unified teacher evaluation system is desirable. Stronge contended that teacher evaluation should fulfill the needs of both the organization and the individual. Institutional needs include accountability/goal attainment and organizational improvement, which is dependent upon individual improvement, and individual needs include personal goal attainment and improvement.

In addition to institutional goal accomplishment, an evaluation system should facilitate compatibility with and support for individual goals. Goals that are mutually beneficial (i.e., compatible) to the individual as well as the institution are essential. Indeed, if goal accomplishment (both institution and individual) is fundamental to success, then the evaluation system should reflect this balanced perspective (p. 134). Further, because goals “reflect a desired state of being, not an existing state...an emphasis on improvement and monitoring of progress toward goal accomplishment are inherent in a sound evaluation system” (p. 134).

According to Stronge, an evaluation system that moves beyond a focus on minimal competence to emphasize growth “offers greater potential for systematically improving the organization and moving it toward the accomplishment of its stated goals” (p. 134). In order for this to happen, “organizational barriers (i.e., incompatibility of individual and institutional needs) and personal barriers (i.e., disillusionment, distrust, stress, fear of failure) must be removed” (p. 134). Such an evaluation system is marked by five characteristics: 1) mutuality of goals between the organization and individual; 2) emphasis on two-way, systematic communication (that attends to context and clarity); 3) climate for quality evaluation (environment of mutual trust and cooperation, fairness, and humane treatment); 4) technical rationality (conceptually sound, accurate, and ethnically and legally defensible measures); and 5) use of multiple data sources.

Integrating SDT and the Improvement-Oriented Model for Performance Evaluation

What seems evident after reviewing each of these standalone frameworks is that SDT and Stronge's Improvement-Oriented model seem complementary in terms of thinking about what constitutes mutually-beneficial teacher evaluation and why. By integrating the frameworks of SDT and Stronge's Improvement-Oriented Model, we believe we can develop a meaningful lens through which to analyze existing and emerging teacher evaluation systems that provides a clear way forward for teacher evaluation policy.

Figure 1 identifies the ways in which we see the SDT and Stronge frameworks intersecting. The figure illustrates that the focus of evaluation should be the area of overlap between organizational and individual interests, which are (1) growth/improvement—which aligns to basic, human needs for learning and development; and (2) work that supports mutual/compatible goals, which considers the individual but aligns an individual’s need for integration into the larger social structures (organization) in which one is embedded with the concomitant needs of the organization for uniformity and structure.

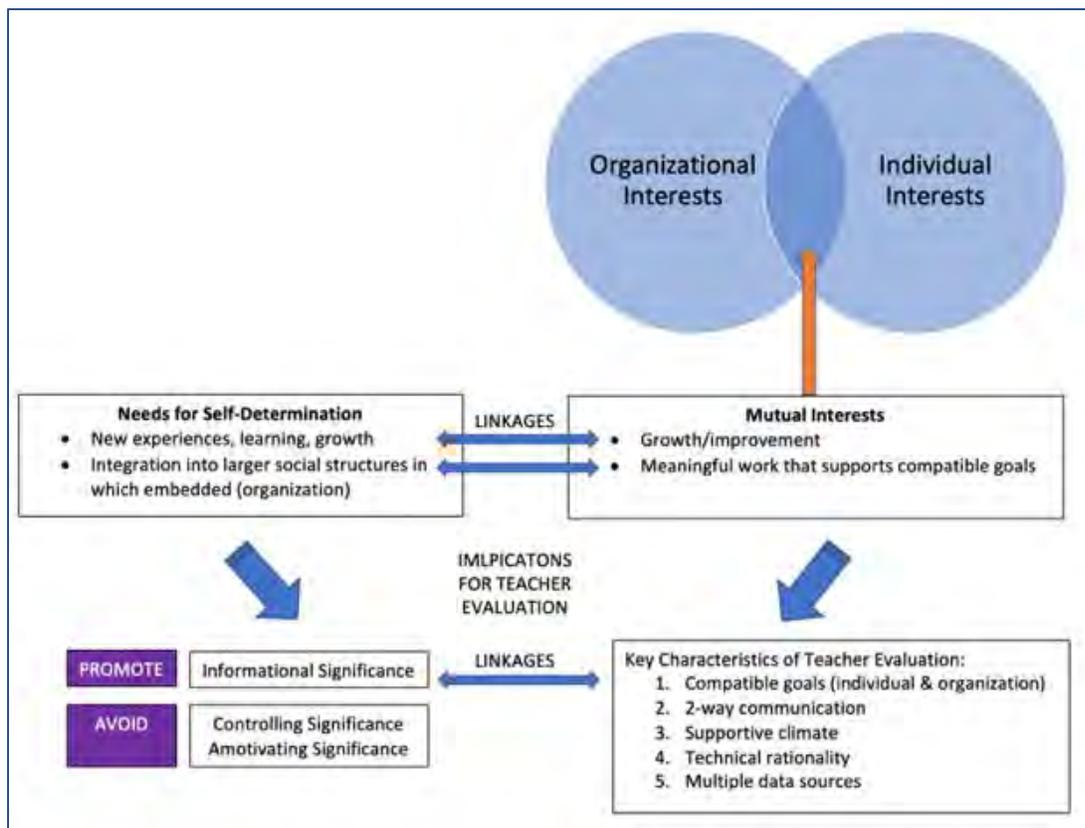


Figure 1. Integrated framework: Intersection of SDT and Stronge’s Improvement-Oriented Model for Performance Evaluation

It is through the lens of SDT that we believe many of the key characteristics of Stronge’s mutually-beneficial model of performance evaluation find their rationale and empirical support. For example, current accountability policies were designed as authority/incentive policy tools (Schneider & Ingram, 1990). The rationale of authority/incentive tools is that using rewards or punishment to change behavior is an effective way to motivate individuals (Ryan & Brown, 2005), and they certainly can be, given the proper conditions. However, the use of punishment and/or rewards frames teacher evaluation as a summative tool, and the pursuit of this goal will produce information more likely to be of interest to stakeholders other than teachers, often at the expense of information teachers might find helpful or useful for improvement. Framing performance information as summative also increases the likelihood of these data being perceived by teachers as controlling or amotivating (Curry et al., 2016; Ryan & Deci, 2017; Ryan & Weinstein, 2009).

A supportive climate for teacher evaluation has to do primarily, in the case of SDT, with the support of key psychological needs on the part of teachers. For example, teachers’ need for

competence for can be supported by providing professional development opportunities aligned with their evaluation results that help teachers improve areas where they are seen as needing improvement. Furthermore, schools can better assist teachers in making meaning out of the evaluation information they are provided, instead of leaving them to their own devices (Lavigne & Good, 2020; Mandinach, 2012). Several studies of effective use of performance information have highlighted autonomy as a key condition which improved teachers' perceptions of the usefulness of performance information for their practice (Farrell & Marsh, 2016; Huguet et al., 2017). According to SDT, people are more inclined to pay attention to performance information from assessments they specifically requested or otherwise had a voice in choosing (Ryan & Deci, 2017). Finally, results from numerous studies demonstrate that, in particular, the use of high-stakes evaluation scores discourages collegial exchange and can engender competition among teachers (Booher-Jennings, 2005; Lane, 2020; Lavigne & Good, 2020). Empirical evidence also suggests that teachers' need for relatedness is both beneficial for the individual as well as the organization, by encouraging more collaborative use of teacher evaluation results (Cosner 2011; Marsh & Farrell, 2015). Collaboration around teacher evaluation results can assist in establishing common language, shared knowledge, understandings and routines, as well as the sense of community needed to better leverage their use (Cosner, 2011; Curry et al., 2016; Datnow & Park 2014).

Finally, a growing body of evidence suggests that teacher evaluation systems should be based on a thorough assessment of teaching practice—not simply rely on only a few measures of performance (whether student test scores or otherwise; Grissom & Youngs, 2016; Lavigne & Good, 2014, 2020). The prioritization (whether intentional or unintentional) of any one measure of teacher performance can not only undermine credibility and perceived fairness of the evaluation system (Ford, 2018; Lavigne & Good, 2020; Rice & Malen, 2016), but it also ignores the many other ways that teachers contribute to student learning (Grissom & Youngs, 2016). Widening the number and type of criteria with which teachers are evaluated provides a way of balancing teachers' needs to be recognized for their many contributions to the work of the school and to have multifaceted data to use for improvement, with the needs of the organization to have a way of summatively evaluating their performance.

Method

For this project, we qualitatively examined current teacher evaluation policy in Hawaii and Washington, D.C. Our analysis focused specifically on the public discourses of teacher evaluation enacted by the Hawaii State Department of Education (HIDOE) and the District of Columbia Public Schools (DCPS) through policy documents. The selection of the DCPS and Hawaii systems was intentional and served as the initial step of our study. First, the DCPS IMPACT (not an acronym) system is one of the most studied teacher evaluation systems in the U.S (see, for example, Dee & Wyckoff, 2015; Gitomer et al., 2015). Hailed by the National Council on Teacher Quality (2018) as one of six teacher evaluation systems that is “getting results,” narrowly defined in terms of meeting accountability goals, IMPACT has also been lionized by the Center for American Progress (2015) as one of 10 “first-mover districts” for reform of teacher compensation, recognizing the role of bonuses in IMPACT as well as the withholding of annual raises (flat salary) for teachers who receive a rating below effective. One quasi-experimental study of the DC IMPACT system found that dismissal threats under the IMPACT system increased the number of low-performing teachers who quit over time as well as improved the performance of teachers that remained (Dee & Wyckoff, 2015).

Table 1
Publicly Available Policy Documents Analyzed

State Evaluation System	Documents
Hawaii Educator Effectiveness System (EES)	<ul style="list-style-type: none"> • EES website • <i>EES Overview 2018-2019</i> video • EES Manual
Washington D.C. IMPACT	<ul style="list-style-type: none"> • IMPACT: The DCPS Evaluation and Feedback System for School-Based Personnel website • The DCPS Essential Practices website • IMPACT Guidebooks

In contrast, Hawaii's system, while also a RttT system, is one of the lesser-known evaluation systems in the U.S., with no known published studies of its design, implementation, or effectiveness. Further, Hawaii and DC are a study in contrast on many levels of consequence for our analysis. Hawaii is, largely rural, sparsely populated, and isolated; DC, in contrast, is a hub of urban activity and federal influence and governance. In our preliminary analysis, we found the language/focus of Hawaii's system to be quite different when compared to DC's IMPACT. Hawaii's system seemed to emphasize to a greater degree the importance of teacher growth and development informed by student growth data—this type of contrast was critical in the exploration of our integrated framework.

The second step we took involved gathering and selecting policy documents that codified teacher evaluation discourse in Hawaii and DC. Specifically, we included only formal documents created by the HDOE and DCPS available in public space (e.g., official websites). We argue that these documents substantially shape what constitutes an effective teacher within each system and, consequently, influence the narrative of teacher evaluation within and beyond the state/district. No formal policy documents produced by HDOE about the Hawaii Educator Evaluation System and by DCPS about the IMPACT system were excluded from analysis (see Table 1 for the list of included documents).

In the third step of our study, we engaged in line-by-line coding (micro-analysis; Stringer, 2009) of each policy document. Specifically, we used provisional coding (Saldaña, 2013) to analyze discourse in the selected policy documents for constructs from our integrated framework in Figure 1 (e.g., compatible goals; supportive culture; two-way communication; mutual interest, and informational, controlling, and amotivating significance). Provisional coding involves establishing *a priori* codes, in our case based on the integrated framework. Provisional codes are modified, expanded, or jettisoned as analysis continues (Saldaña, 2013). Additionally, we also engaged in open coding of the data to capture discursive constructs in the policy documents not encompassed by the integrated framework. The combination of provisional and open coding provided a robust analysis of the data (Marshall & Rossman, 2016).

For the fourth step of our study, we created an operational model diagram (Saldaña, 2013)—a visual map of codes, concepts, and categories from the provisional and open coding process. Through multiple iterations, the operational model diagram informed the construction of the theories of action implicit in the Hawaii Educator Evaluation System and DC IMPACT (see Figure 2, described below).

Limitations

Analysis of formal policy documents has inherent limitations for drawing conclusions about the appropriateness and usefulness of evaluation policy. As people enact policy, they invariably change it (Cohen, 1990; Tyack & Cuban, 1995). By the same token, policy actors' sensemaking is shaped by the framing of policy so that policy itself, aside from enactment, can be consequential (Coburn, 2005; Coburn & Woulfin, 2012; Cohen & Hill, 2001; Rigby, 2015). Additionally, how policy actors—specifically teachers—perceive the fairness and accuracy of their evaluations informs their reactions to them (Ford, 2018; Reddy et al., 2018; Rice & Malen, 2016). As such, interpretations from this study of the productiveness and respectfulness of evaluation policy may be inconsistent with policy actors' interpretations of and responses to these policies.

Additionally, both Hawaii and DC represent single entity systems, such that each is comprised of only one fairly homogenous district, unlike other states that are comprised of myriad districts that vary in size and demographics. As such, the findings from this study likely do not reflect the diversity of other state evaluation systems. Further, the labor market for each system is a key consideration: DC has a robust and stable labor market (Dee & Wyckoff, 2015) that can support an evaluation system with harsher consequences for poor performance. In contrast, Hawaii, which is geographically isolated, has a finite labor market marked by current teacher shortages (Peterkin, 2019), which may require an emphasis on teacher development over teacher sorting. In short, size and stability of labor markets may also inform and influence teacher evaluation systems, but these considerations were largely outside of our study purview.

Findings and Interpretation

According to the integrated framework, evaluation systems cultivate improvement best when they reflect mutual interests (growth/improvement) and when they contain five key characteristics: compatible goals (individual and organization); two-way communication; supportive climate; technical rationality; and use of multiple data sources. The findings of our analysis are organized, respectively, according to these concepts and each of the focal evaluation systems (Hawaii and DC IMPACT), comparing and contrasting, where appropriate, the discursive elements they contain as well as their implications for teacher evaluation policy and practice.

Mutual Interests

The concept of mutual interests refers to policies that emphasize both teacher growth as well as accountability goals. Both Hawaii's Educator Effectiveness System (EES) and IMPACT systems purport to focus on growth. The EES website states:

To help students succeed in college and careers, it is imperative that the Hawaii State Department of Education (HIDOE) support our educators to become highly effective in their schools and classrooms. This means that administrators and teachers need feedback, coaching and data that inform them about how to improve their practice and make an impact. We are holding ourselves accountable at all levels of the organization for providing support and getting results for students. (HIDOE, n.d.-a, para. 1)

This statement communicates a number of things: 1) The EES privileges teacher growth/improvement over teacher sorting, reward, and punishment; 2) the HIDOE recognizes its reciprocal accountability to teachers to provide resources (“feedback, coaching and data”) to support their growth; 3) the EES is built on an implicit theory of action that positions teacher growth as the

means by which the HDOE goals—college and career success—are met, as reflected in Figure 2 below. The discourse of teacher growth and reciprocal accountability is reflected throughout the EES policy artifacts, including seven times in the evaluation manual for teachers.

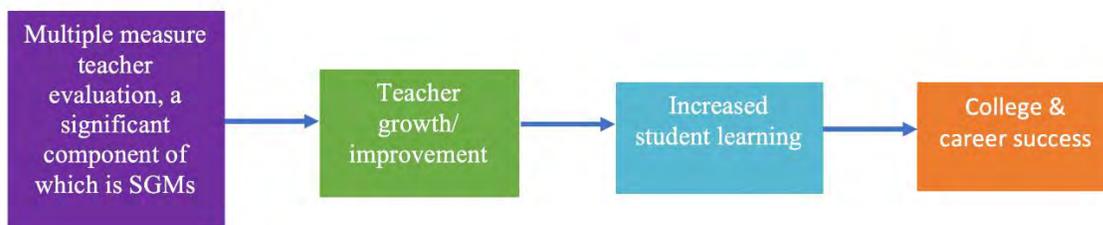


Figure 2. Hawaii EES implicit theory of action.

Note. SGMs = Student growth measures.

In contrast, in DC IMPACT policy documents, growth typically refers to increases in student achievement. In the evaluation guidebooks (discussed below), immediately under the page two title “putting growth first,” the text reads:

DCPS has seen continuous improvement in student achievement because of the extraordinary passion, skill, joy, and talent teachers, school leaders, and staff bring to work each day. . .IMPACT reflects our belief that everyone in our system plays a critical role in improving student outcomes. (DCPS, n.d.-a, p. 2)

Here, growth is referring to improvement in student achievement/outcomes—and that growth is attributed to faculty, staff, and administration. In two places, further down on the page—and also on the IMPACT website—there is also a recognition of the need to support teacher growth through clear expectations and feedback: “With an outstanding teacher in every classroom. . .our students will graduate prepared for success. IMPACT supports professional growth by 1) clarifying expectations. . .; 2) providing frequent and meaningful feedback” (DCPS, n.d.-a, p. 2).

However, in DCPS, rewarding effective teachers is also strongly emphasized: Teachers with Highly Effective ratings earn substantial bonuses and base salary increases, which are outlined more thoroughly in the IMPACT evaluation guidebooks:

Great teachers are essential to student success. That is why DCPS teachers who earn Highly Effective ratings are rewarded with bonuses up to \$25,000 and can earn up to \$3.7 million over the course of their careers through IMPACT *plus* – DCPS’s performance-based compensation system. (DCPS, n.d.-e, para. 4)

Also outlined in guidebooks are the consequences for not scoring effective or higher on the evaluation: one year of an “ineffective” rating results in “separation” (dismissal), two consecutive years of a “minimally effective” rating result in dismissal, and three consecutive years of a “developing” rating result in dismissal. Thus, while there is acknowledgement of the importance of feedback for teacher improvement, the implicit mechanism driving growth in student achievement through evaluation is the use of reward (bonuses and salary increases) and punishment (dismissal) to change the composition of the teacher ranks, as reflected in Figure 3.

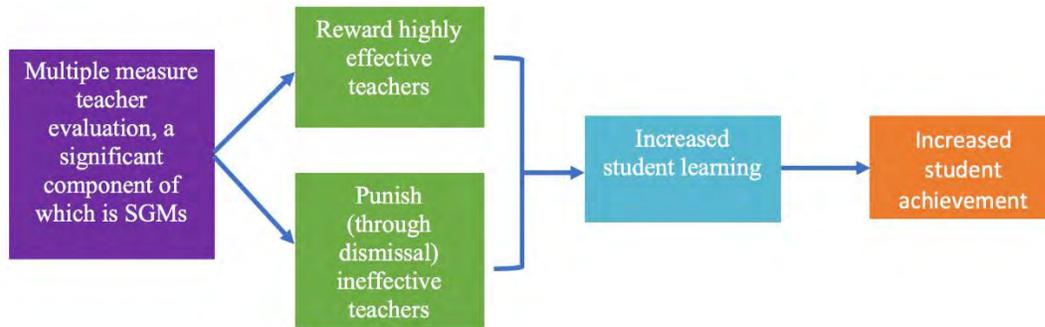


Figure 3. DC IMPACT implicit theory of action.

Note. SGMs = student growth measures.

In summary, a major distinguishing feature between EES and IMPACT is the degree of emphasis each policy has placed on growth/improvement (in the former) and reward and punishment (in the latter). In the integrated framework, *mutual interests* reflect a simultaneous focus on both growth/improvement, which is well-reflected in EES and, while articulated in IMPACT, does not appear to be the central mechanism driving its theory of change.

Multiple Data Sources

Both the EES and IMPACT incorporate multiple measures of educator effectiveness. Within the EES (see Figure 4), those elements are: a) teacher practice, as assessed by observations or, in the case of educators who are not classroom teachers, working portfolios (30% of composite evaluation); b) core professionalism (20% of composite), which involves reflection on student growth percentiles (SGPs) and student survey data (at the individual teacher or school level), as well as other aspects of professionalism; and c) student growth and learning (50%), which is assessed through a form of student learning objectives (SLOs) known as Student Success Plan (SSP), or, in the case of educators who are not classroom teachers, School System Improvement Objectives (SSIOs). Thus the reliance on student growth measures—a hallmark of RttT—is twofold in ESS; however, student growth percentiles, which are calculated using standardized tests, are not directly used to measure teacher effectiveness in EES. Rather, they are a focus of teacher reflection, which is included in core professionalism (20%). SLOs, which comprise the student growth and learning component and account for 50% of a teacher's rating, are selected by teachers based on classroom assessments. SLOs are potentially more relevant to classroom instruction and more meaningful to teachers; however, they are also potentially more susceptible to gaming and bias (Crouse, Gitomer, & Joyce, 2016; Ford et al., 2017). Additionally, SLOs, because they are individual to a specific teacher, make it more difficult to compare teachers' performance across classrooms and schools. This current structure is quite different than Hawaii's original RttT proposal, which based 50% of a teacher's rating on student growth measures from standardized tests (HIDOE, 2010). Indeed, lack of progress in development of the proposed teacher evaluation system was one reason that the Hawaii RttT program was put in "high risk status" by the U.S. Department of Education in December, 2011; Hawaii came fully out of "high risk status" in July, 2013 (HIDOE, 2013; U.S. DoE, 2013).

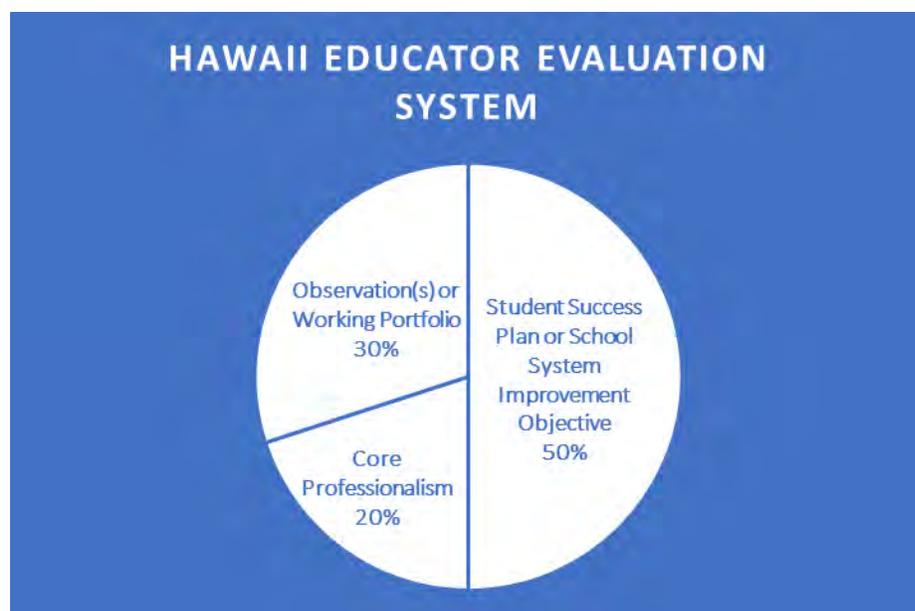


Figure 4. Hawaii Educator Evaluation System multiple measures of teacher effectiveness.

Note. Observations(s) or Working Portfolio (30%) is based on observations for classroom teachers and a portfolio of evidence for non-classroom teachers. Core Professionalism (20%) is assessed holistically using a rubric that includes ethics, professionalism, collaboration, record-keeping, communication, and reflection, including reflection on student growth measure data and student survey data. Student Success Plan or School System Improvement Objective (50%) involves the development of one student learning objective specific to the teacher's course/subject/grade, including baseline data, instructional strategies, growth data, and reflection or – for non-classroom teachers – involves the development of a goal and improvement objective school or system data, strategies for achieving the goal, and outcome evidence.

In the EES model, educators are also required to *reflect* on their student growth data and student survey data, but these data do not receive weight in teachers' evaluation scores. By emphasizing reflection and growth over using the data to calculate "effectiveness," this element of the EES prioritizes teachers' needs for self-determined learning by privileging their understanding and use of the data over the quantitation of teacher effectiveness. Furthermore, the EES process reflects a nuanced understanding of the role of autonomy in shaping teachers understanding of and interest in the data generated from teacher evaluation. To complete the SSP, teachers must collect baseline data on the "most important desired learning" (HIDOE, n.d.-b, p. 27), identify instructional strategies to be used, provide assessment data that demonstrates student growth, and reflect on their practice. SSP's must be approved by a teacher's supervisor. While 50% of the teacher's evaluation is dependent upon student learning data through the SSP/SSIO, the fact that teachers have a great deal of input into the nature and focus of their SSP likely increases teachers' perceptions of the informational significance of the results.

Because it has these characteristics, the information generated from SSPs is much more likely to be relevant to a teacher's classroom practice (in the sense that teachers select the standard/s, instructional practices, and means of assessment) than measuring growth by using standardized tests and situates more choice and control in the hands of the teacher (Farrell & Marsh, 2016; Huguet et al., 2017). Additionally, numerous studies (e.g., Eckert, 2016; Ford, 2018; Hewitt, 2015; Rice & Malen, 2016) indicate that educators generally struggle to understand how to interpret teacher effectiveness scores that are produced through student growth measures and value-added

models (VAM), while the implications of the results of student learning objectives, if they are developed by teachers themselves, are more likely to be self-evident.

Further, the HIDEOE encourages teachers to integrate their SSP efforts into their larger data-driven decision-making efforts. For example, if a group of teachers in the same department, course, or grade level can agree on a common SSP, or if the school develops a school-side SSP, data team meetings could then become a useful forum for analyzing progress towards the SSP and sharing teaching strategies that are successful in helping students demonstrate growth (HIDEOE, n.d.-b, p. 27). This approach ensures that individual teacher evaluation practice is rendered more meaningful through its integration into the larger social structures within a department or school. Under the ESS model, teachers' development of their own SLOs, which prioritizes teacher choice and role relevance, emphasizes teacher meaning-making, and engenders greater opportunity to engage collaboratively with other teachers, maximizes the potential of the evaluation feedback to be viewed as informational and not controlling to teachers.

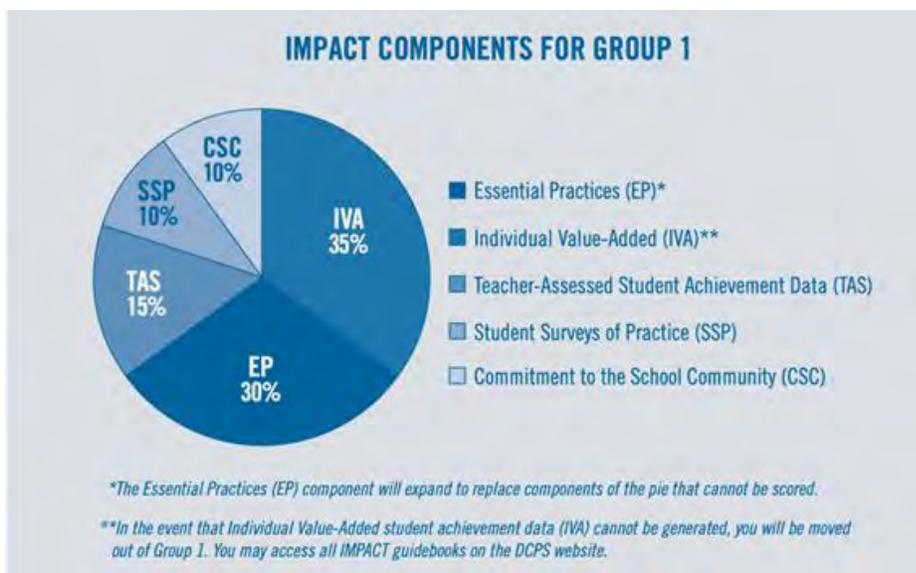


Figure 5. From DCSP (n.d.-a) group 1 guidebook.

Note. IMPACT components for Group 1 (teachers in grades 4+ with individual value-added data and student survey data). Essential Practices are evaluated based on observation data. Teachers are observed one to three times, based on their stage on the DCPS career ladder. TAS data come from student learning objectives that are based on a “measure of your students’ learning over the course of the year, as evidenced by rigorous assessments other than PARCC” (DCPS, n.d.-a, p. 30). Commitment to the School Community (CSC) is evaluated using a rubric that assesses the employee’s support of local school initiatives, support of special education and English language learner initiatives, demonstrating high expectations, partnering with families, and engaging in instructional collaboration. On each employee’s evaluation, across groups, CSC comprises 10% of the final rating. Core professionalism (CP) is a component of each employee’s evaluation, across groups. CP, which reflects attendance, punctuality, following policies and procedures, and showing respect, can result in points being deducted from an employee’s overall evaluation score but cannot result in points being added.

In the IMPACT multiple-measures model (see Figure 5), student growth measures can represent up to 50% of a teacher’s evaluation and include the more opaque measures of teacher effectiveness mentioned above (e.g., individual value-added scores (35%), as well as teacher-assessed student achievement data (TAS), which is a form of SLO (15%). Unlike SLOs, the standardized-test-based student growth measures, such as SGPs and value-added scores, are generated based on algorithms (to which teachers are not privy), and annual standardized tests (that teachers have no

input into), tend to be more difficult to understand, and are often perceived by teachers to be a poor measure of their effectiveness (Amrein-Beardsley, 2014; Amrein-Beardsley & Collins, 2012; Eckert, 2016; Ford et al., 2017). While the limitations of these measures are widely acknowledged by researchers as well as educators, the IMPACT Group 1 Guidebook allocates only one page to the topic of individual value-added (IVA) scores and explains the process of calculating them in one sentence:

Step 2: Statisticians determine the predicted PARCC scale score for each student;
 Step 3, Statisticians determine the difference between the predicted PARCC scale score and the actual PARCC scale score; and Step 4, The difference in all students' predicted and actual PARCC scale scores is combined for each teacher to create a raw IVA score. (DCSP, n.d., p. 26)

Discursively, this language positions the anonymous, remote statistician as the determiner of effectiveness, and the teacher's only contribution to the IVA process is in Step 1: "Teachers confirm their student rosters" (DCSP, n.d., p. 26). In this sense, IVA can be perceived as something done to teachers and a process that is largely out of their control. Performance pressure coupled with a lack of perceived control over the process makes this approach to evaluating effectiveness more likely to engender amotivational orientations to the feedback on the part of teachers.

However, the IMPACT model does take into account an important limitation of past evaluation systems, by acknowledging the need to differentiate evaluation approaches to a greater degree for different types of certified educators within a school building, understanding that different teachers and staff within a school may need different types of feedback. The aforementioned use of IVA is only for teachers in Group 1, which is comprised of teachers in grades 4 and above. Other educators fall into other groups, which are evaluated according to a different mix of elements. As a point of comparison, see Figures 6 and 7 which reflect the respective differences between weighting of multiple measures for Group 2a (early childhood teachers) and Group 10 (school counselors).

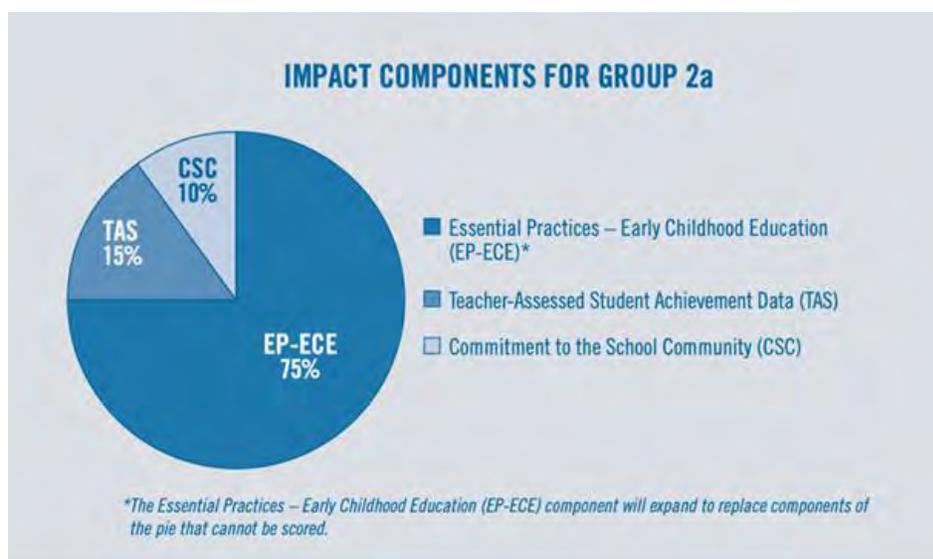


Figure 6. From DCSP Group 2a (n.d.-a). IMPACT components for early childhood teachers, including pre-kindergarten and kindergarten teachers).

In fact, there are 20 groups and multiple subgroups in the IMPACT model, totaling 33 distinct groups/subgroups, each with its own evaluation composite and guidebook. On one hand, the fact that evaluation components are different—and differently weighted—by role group maximizes the potential informational significance of the feedback, ostensibly because the differentiation takes into account elements of work that are more relevant to the educators within each group. Viewed another way, however, the myriad (33) groupings, each with its own “guidebook” and rubric that outlines multiple evaluation elements, makes educator evaluation in the IMPACT system somewhat unwieldy, to some degree overwhelming, and potentially confusing to educators trying to understand how they will be evaluated and what they can do with this information. Indeed, the 33 guidebooks range in length from 30 pages (Group 10 - school counselors) to 80 pages (Group 3a – Special Education Teachers – Communication and Education Support Program).

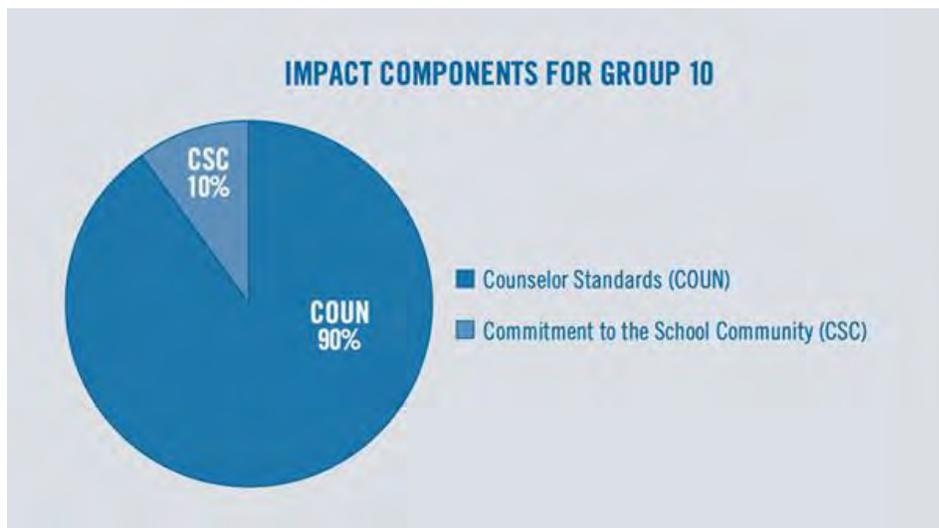


Figure 7. From DCSP Group 10 (n.d.-a), IMPACT components for school counselors.

In summary, Hawaii’s EES and DC’s IMPACT both contain similar deployment of multiple measures that include teacher observation, student growth data, and measures of teacher professionalism. However, there are important differences. The Hawaii model does not directly measure teacher effectiveness through standardized test-based student growth measures, but instead requires teachers to reflect on the data, while DC’s IMPACT model does directly calculate a teacher effectiveness score from standardized test data. Both systems require SLOs. In the Hawaii system, teachers have the option to develop common SLOs across multiple teachers to promote collaboration. Of the two systems, the multiple measures of the Hawaii EES better reflect self-determination and an improvement-orientation, as reflected in the integrated framework.

Two-way Communication

Within the context of evaluation systems, two-way communication can be demonstrated most clearly through the opportunity for teachers to provide input into the system and feedback on it, as well as to seek assistance with it when they have questions or challenges. Two-way communication systems are unique and desirable because most evaluation policy, by its very nature, is top-down in orientation. The HIDOE highlights in its policy documents the multiple ways in which educator input and feedback have been sought and utilized:

Since the beginning of the EES pilot in 2011-12, Hawaii educators have had a significant voice in improving their evaluation system. The feedback has come in a variety of forms including survey responses and in-person conversations with teachers, administrators, and union officials. Continuous improvement has been based on feedback received from various stakeholder groups (see below), Complex Area Superintendents and their EES support staff, and the HSTA-HIDOE Joint Survey (see [April 2015](#) and [April 2014](#) results). (HIDOE, n.d.-a, para. 3)

Additionally, the following stakeholder groups have provided input and/or feedback for the EES: (a) the Teacher Leader Workgroup, which is comprised of over 100 educators from across the state; (b) the Hawaii State Teacher Association (HSTA)/Hawaii Department of Education (HIDOE) Joint Committee, (c) the EES Technical Advisory Group, which ensures a fair evaluation process through a review of data, policies, and practices, and (d) the EES Help Desk which not only answers teachers' questions about the EES but also "documents caller feedback to improve the overall EES training and implementation planning" (HIDOE, n.d.-a, para. 4). Furthermore, the HIDOE uses multiple modalities to communicate about the EES, including a website dedicated to the EES, a 12-minute introductory and overview video about the EES, and the 56-page *Manual for Evaluators and Participants*.

Throughout the EES policy documents, communication appears to be respectful and culturally responsive. Both the video and manual open using second person perspective ("you," "your"), such that it seems as though teachers are directly being addressed, and both begin with a message from the superintendent that greets teachers with "aloha!" and thanks teachers for their commitment and efforts on behalf of students. The message from the superintendent in the manual reflects culturally sustaining language through the use of the Hawaiian word "haumana," meaning students: "Our haumana deserve the best educators to prepare them . . ." (p. i). The superintendent's message in the video ends with, "Thanks for all you do . . . have a great year," and the video contains images of classrooms and diverse teachers and collaborative teacher teams. The overall effect appears respectful and teacher-centered, promoting the concept of teacher as professional. DCPS communicates about IMPACT through the IMPACT website and the 33 Guidebooks mentioned previously. The website states that IMPACT was "designed with input from teachers and administrators" (DCPS, n.d.-e, para. 1) although no further details are provided. While the input/feedback sought by DCPS and provided by teachers may have matched or exceeded that for the EES, discursively two-way communication is neither touted nor evident within IMPACT policy documents themselves.

Interestingly, in contrast to parts of the Hawaii documents that use pronouns "we" and "you," the IMPACT documents generally speak in the third person, but there are also modest deviations from this general style, which reflect a more collaborative discourse. The introductory page of the guidebooks ends with a singular testimonial by a DCPS teacher that is also featured prominently on the IMPACT website:

I'm so lucky to have started my teaching career in DCPS and to have developed my craft under IMPACT. The feedback I have received from observers over the years has made me a better teacher and in turn has helped my students learn more in a joyful classroom. Teaching in DCPS isn't like teaching anywhere else. Having a framework like IMPACT...shows how much DCPS values its teachers and values learning. I'm very thankful to be a part of it. — Teacher, Eastern Senior High School (DCPS, n.d.-a, p. 2)

Notable in this included quote is its effusiveness, that it is the only teacher quote in the IMPACT materials and featured on the IMPACT website, and that its use of words like “joyful,” “inspiring,” “continual focus on growth and collaboration with leadership,” and “values [teacher] learning” seem out of sync with the rest of the Guidebook discourse, which seems detached, officious, impersonal, and filled with rubrics, scoring matrices and charts, framework standards, and impact scale. For all their length, however, the Guidebooks are far from comprehensive. For example, the Teacher-Assessed Student Achievement Data (TAS), which is a form of SLO’s, requires that:

Assessments must be rigorous, aligned to the Common Core State Standards or other appropriate content standards, and approved by your school administration. Please see the TAS guidance document for resources on commonly used assessments, and assessments that cannot be used for TAS. (DCPS, n.d.-c, p. 36, in Group 2d Guidebook; content is the same but pagination is different for other groups that also have a TAS element in their evaluations)

There is no link within the Guidebook to the TAS guidance document, nor is there access to it or even mention of it on the IMPACT website.

There are also strong elements of communication and support in the Guidebooks, however. On some pages of the guidebooks, there is a phone number and email address for questions concerning IMPACT. Also, most of the sections of the guidebooks are phrased as questions teachers might ask (e.g., “How will I receive feedback from my IMPACT observation?” [DCPS, n.d.-c, p. 8, Group 2d Guidebook]). Also, the guidebooks for classroom teachers that incorporate an Essential Practices component, which is scored using rubrics based on classroom observations, provide not only the rubric for the essential practices (e.g., “maximize student ownership of learning” [DCPS, n.d.-c, p. 21, Group 2d Guidebook]), but also include content-specific examples of how the standards can be enacted in English language arts, math, science, and social studies, as well as identify sample LEAP (LEarning together to Advance our Practice) professional development modules that link to the examples.

In summary, the Hawaii EES includes two-way communication through opportunities for input and feedback, personable and respectful language for teachers, and some culturally sustaining language. The strengths of the DC IMPACT system with respect to communication are the organization of guidebooks in anticipation of teacher questions, as well as content-specific examples and access to professional development modules for essential practices. These strengths counterbalance the detached tone and complexity of the guidebooks.

Compatible Goals

In the EES system, as previously described, 50% of the composite evaluation is comprised of the SLO (Student Success Plan [SSP] for a classroom teacher or School System Improvement Objective [SSIO] for non-classroom educators). The stipulations for SSPs is that they are “thoughtfully selected outcomes or standards that will reflect the most important desired learning,” are “specific to the source or subject and grade” taught, draw upon baseline data, and identify “instructional strategies to be utilized” (HIDOE, n.d.-b, p. 27). Teachers create one SSP per year and are required to get it approved by their supervisor. Behind these specifications, teachers have wide latitude regarding the focus of their SSP, the assessments and instructional practices they use, and whether they develop an individual SSP or a common SSP with colleagues who teach the same grade/course/subject. Because it promotes teacher learning and development (serving individual and organizational interests) yet provides evidence that can also be used for evaluation (organizational interest), the SSP is an example of a process which reflects a compromise between individual and organizational goals.

Within the IMPACT system, some teachers have a SLO component, known as Teacher-Assessed Student Achievement Data (TAS). These teachers have the opportunity to develop their own goals, but after this point, there is limited involvement by the teacher in assessing growth:

Please note that administrators must approve all assessments, targets, or weights selected for TAS goals. In the spring, achievement data for all assessments will be presented to administrators who, after verifying the data, will assign scores for each goal based upon the rubric. (DCPS, n.d.-c, p. 36 of Group 2d Guidebook)

Thus, while teachers set goals and choose assessments that will promote their learning (individual interests) as well as support the district goal for increases in student achievement (organizational interests), individual autonomy is circumscribed to some degree by requiring that others approve their SLOs and use a rubric to render a final determination of effectiveness.

In summary, the Hawaii and DC systems both utilize SLOs, yet the way in which these processes are implemented in practice has noteworthy implications for the degree to which SLOs reflect compatible goals. Hawaii's EES better reflects a balance between individual (personal growth) and organizational (accountability) goals.

Supportive Climate

The EES website provides a list of supports, including EES resources on the HIDOE intranet (accessible to teachers via username and password and not available to the public), such as: “videos, presentations, reference documents, Frequently Asked Questions and other communications” (para. 4). The PDE2 online system, in addition to being the platform for EES documentation, provides a search feature for professional development opportunities.

Throughout the ESS policy documents, there is language that reflects a growth-oriented, supportive climate, such as a description of EES as a system that “provides educators with quality feedback and support to improve their effectiveness with students, and informs professional development” (HIDOE, n.d.-a, para. 1) and “teacher quality is best supported within an organizational culture that embraces ongoing feedback and commits to continuous learning” (HIDOE, n.d.-b, p. i). Additionally, a value for differentiation is discursively constructed through statements such as, “Every teacher is unique, therefore support and development should not look exactly the same for everyone. It is imperative that teachers and administrators have opportunities for honest conversations focused on promoting continuous improvement” (HIDOE, n.d.-b, p. 9). Also, given the myriad forms of two-way communication and emphasis on “continuous improvement of design and implementation” (HIDOE, n.d.-a, para. 3) of EES, it appears that there is a growth-oriented expectation of not only teachers but of the EES system itself.

Within DCPS, there are elements of a supportive culture that reside mostly outside of IMPACT but may intersect with it. One example is the Essential Practice Video Library, a burgeoning collection of videos featuring DCPS teachers enacting excellent practices. Each video includes a voiceover play-by-play explanation of the exemplary practices featured in the video. In this way, teachers can see what great instruction looks like and makes DCPS expectations for teacher practice clear and tangible. The other key element of supportive climate is LEAP professional development (introduced in an earlier section). LEAP involves a weekly session of professional learning communities (LEAP teams) at their schools facilitated by content experts (LEAP Leaders). LEAP 90-minute seminars follow a sequence of new learning about a core instructional practice, lesson planning that incorporates the practice into an upcoming lesson, analysis of student work, and plans for how to respond to student data. In addition, teachers engage in LEAP coaching, which

can involve modeling and debrief, co-planning, and observation and debrief. These are powerful supportive practices, although they are not embedded in IMPACT.

Specific to IMPACT, the climate appears high stakes with asymmetrical levels of support. One example of this high-stakes climate is the deficit approach to the Core Professionalism component of each educator's evaluation, which includes: attendance, punctuality, compliance with policy, and respectfulness. Core Professionalism scores cannot add or contribute to one's overall rating; they can at best maintain one's total point value or, at worst, reduce it. Any rating of "slightly below standard" results in a 10-point deduction, and any rating of "significantly below standard" results in a 20-point deduction.

Additionally, there is language throughout the policy documents that speaks to a "performance-based culture" (DCPS, n.d.-e, para. 1), where "teachers should be held accountable for the achievement of their students" (DCPS, n.d.-a, p. 30), and a teacher's evaluation consists of a series of quantifications of the person's performance that is totaled and used to categorize that teacher into one of five performance levels. In this respect, a teacher is reduced to a number—a number that, by virtue of the complexity of the multi-component rating system with its series of scales and rubrics and scoring tables, leaves no room for context, nuance, or situational consideration.

The IMPACT system is, at its core, a system of reward and punishment. As discussed earlier, one year of an unsatisfactory rating results in immediate dismissal, as does two years of minimally effective and three years of developing ratings. There is an appeals process, but this appeals process, as written, evokes little sense of hope but rather a greater sense of doom. There is very little discursive evidence which suggests that the system values the importance of due process, fairness, or having one's voice heard.

However, for a teacher who complies and performs, the rewards are indeed rich. As discussed earlier, bonuses of up to \$25,000 and increases to base pay make high performance and compliance lucrative. Further, good soldiers of this system are:

celebrated at Standing Ovation, an annual gala hosted by the DC Public Education Fund. This star-studded celebration honors the achievements of DCPS' top teachers, publicly recognizing their outstanding work and awarding cash prizes. Grammy-winning recording artists such as John Legend, Wyclef Jean, and Roberta Flack have performed at this event to honor DCPS teachers. (DCPS, n.d.-e, para. 4).

Such a system of high-stakes reward and punishment is inherently antithetical to the tenets of self-determination, as it engenders controlling motivational orientations to the process. In doing so, it risks stripping participants of the reasons why behaviors are engaged in and why outcomes are sought. Rather, IMPACT's elaborate system of reward and punishment is reminiscent of Gramsci's (2003) concept of ideological hegemony through which the dominant power extracts consent (Maglaras, 2013) by manipulating beliefs, explanations, values, mores, and norms, such that the oppressive system becomes normalized, even validated, by its own internal logic.

In summary, Hawaii's EES provides a supportive climate through a number of differentiated supports for teachers and language that reflects a growth-oriented, supportive climate. DCPS provides a supportive climate to teachers through the Essential Practice Video Library and LEAP professional development and teacher leadership. Conversely, IMPACT's system of rewards and punishments undermines a supportive climate by promoting competition as well as an ideological hegemony that serves to condition teachers towards performance for its own sake.

Discussion

Summary

Overall, our analysis of the IMPACT and Hawaii EES systems yielded some findings of note across the elements of our integrated framework. First, EES and IMPACT have marked differences in the degree to which each policy is attendant to the mutual interests of stakeholders towards growth/improvement (in the former) and reward and punishment (in the latter). Second, both Hawaii's EES and DC's IMPACT both contain similar deployment of multiple measures that include teacher observation, student growth data, and measures of teacher professionalism, with some small, but important, differences. The IMPACT model does emphasize the calculation and use of teacher effectiveness scores from standardized test data and, while both systems employ SLOs, in the Hawaii system, teachers have the option to develop common SLOs across multiple teachers to promote collaboration. Hawaii's system of SLO develop better reflects a balance between individual (personal growth) and organizational (accountability) goals.

Third, the Hawaii EES includes two-way communication through opportunities for input and feedback, personable and respectful language for teachers, and some culturally sustaining language. The strengths of the DC IMPACT system with respect to communication are the organization of guidebooks in anticipation of teacher questions, as well as content-specific examples and access to professional development modules for essential practices. Finally, Hawaii's EES provides a number of differentiated supports for teachers and language that reflect a growth-oriented, supportive climate. DCPS provides a supportive climate to teachers through the Essential Practice Video Library and LEAP professional development and teacher leadership. Conversely, IMPACT's system of rewards and punishments may serve to undermine a supportive climate by promoting competition as well as an overall controlled-motivation orientation to the process.

Conclusions

Acknowledging and adjusting the dynamics of power. As Ryan and Deci (2017) point out, power inequities residing within the environment in which feedback occurs can have wide-ranging ramifications on how that feedback is received and used. One of the endemic challenges of teacher evaluation systems is that there are inequitable power relationships both at the policy level as well as the level of practice. At the policy level, top-down, accountability-based evaluation prioritizes the needs of the system and organization to classify, rate, and code individuals for the purpose of summatively assessing, dismissing, censuring, or rewarding teachers. This tendency was evident in particular with respect to the IMPACT program. While the rationale is under the guise of equity and fairness, this need necessarily entails that teachers assume a subordinate role in the process instead of participating as equals with input into what teacher evaluation should look like. In some sense, summative evaluation must be inherently unequal, for if the subjects of the evaluation were equal, the assumption is they would work to loosen expectations for success. Assumptions about how teachers would respond to a more equitable system breeds distrust among teachers and can often undermine the credibility of the process and the feedback generated by it.

Furthermore, at the level of practice, principals have historically served as evaluators. There is growing evidence that teachers find meaning and satisfaction in evaluation processes that are more open to teacher input, as well as those that are conducted by teaching peers, instructional coaches, or other expert teachers as opposed to the principal (Ford et al., 2018; Lavigne & Good, 2019, 2020; Smith & Kubacka, 2017). We believe there is an opportunity to better consider the role of veteran teachers, coaches, and mentors in the evaluation process, not only because they often possess the knowledge and experience needed to support fellow teacher improvement, but also because

principals are already overworked (Lavigne & Good, 2015) and, as a result, might approach the evaluation process with more attention to efficiency than rich feedback (Lavigne & Good, 2020).

Similarly, the inherent power dynamics of many teacher evaluation systems stifle two-way, meaningful communication about what works and what is not working with teacher evaluation and how to fix it. We found stark differences in the language used which suggested that too great an emphasis on summative components can also shape how communication flows from different policy actors, sending implicit messages about the role teachers are to play in the process.

The ultimate purpose of reducing power inequities and re-centering teachers as key actors in the evaluation system is to ensure that feedback gets used not just for effectiveness judgments, but also for improving teaching. Because a majority of teachers enter the profession for altruistic reasons (Lortie 1975; Rosenholtz 1991), it is not surprising that a substantial portion of what motivates teachers to improve is the desire to see their students grow and thrive. Performance data will always be inherently interesting—people like to see that they have done well—but this feeling is typically temporary (Ryan & Deci, 2017). Feedback which points the way to better teaching and learning on the part of students will, in the long term, sustain teachers' intrinsic motivation for the work (Ford et al., 2017; Ryan & Weinstein, 2009), and ultimately feed their desires to improve their practice so that they can help students grow and thrive (Ford, 2018).

Supporting teachers' psychological needs as learners. An overall climate of support is needed for teachers who are involved in meaningful evaluation work. Resources are important, but, because of the challenging and messy nature of teaching, these should also be coupled with peer support, intensive coaching, and modeling to support teachers' psychological need to feel competent in their ability to face new teaching challenges (Farrell, 2014; Gabriel & Woulfin, 2017; Marsh et al., 2006; Marsh et al., 2010). Competence must also be balanced with the latitude to apply learning to practice in a way teachers find meaningful. Autonomy, also a key psychological need for teachers, allows teachers to experience success through their application of knowledge to problems of interest through self-determined action (Ford et al., 2017). Autonomy, however, is not simply freedom without structure; structure is necessary for true autonomy to be realized (Ryan & Deci, 2017).

Within teacher evaluation systems, an important structural component is the clarity which exists with respect to evaluation expectations, standards of success, and roles and responsibilities (Delvaux et al., 2013; Kelly et al., 2008). Clear standards can also facilitate perceptions of the evaluation as an authentic and fair assessment of practice (Delvaux et al., 2013; Lavigne, 2014), which is an important predictor of whether or not teachers will use evaluation feedback or not (Ford, 2018). Accountability policy serves some of the needs of organizations for control and certainty, but typically fails to be mindful of the needs of individuals as learners. Neoliberal approaches to teacher evaluation are based on the foolhardy assumption that certainty in these educational processes or outcomes can be achieved, but this assumption is in direct contravention to what educational philosophers and researchers have known about education for decades—teaching and learning are inherently uncertain, messy endeavors (Cohen, 2011; Lavigne & Good, 2019, 2020; McDonald, 1992).

An obsession with certainty comes, paradoxically, with hidden costs. In the pursuit of uniformity and certainty, we disrupt natural human predilections towards self-determination and growth and instead produce individuals who simply follow rules, policies, and procedures for their own sake—in the balance, gaining very little for their own benefit. We believe it is entirely possible to create more humane systems of evaluation that better appreciate the inherent tension between summative and formative goals—between individual and organizational needs—and seek to strike a better balance between the two. Such a balance will necessarily require constant vigilance to be maintained, because the tension between these two purposes is always present.

Implications for Future Teacher Evaluation Policy, Practice, and Research

Based on the findings from our analysis of the Hawaii and DC teacher evaluation systems, we offer the following recommendations that move beyond overreliance on attribution of teacher effectiveness to changes in students' standardized test scores:

- 1) Build from a theory of action that centers teacher growth and improvement instead of reward and punishment. Focus on building teacher excellence, as opposed to sorting teachers into “keep” and “discard” piles.
- 2) Attend to *how* multiple measures are used to evaluate teachers. While multi-measure evaluation systems are now common for teachers, the examples of EES and IMPACT demonstrate that evaluation systems can have the same components but use them differently to greatly different effects. As such, policymakers must work to incorporate these components in ways that best promote teacher self-determination and informational significance of the feedback, such as: a) giving teachers substantive autonomy over establishing SLOs while holding teachers to high expectations (for rigorous assessments and ambitious growth targets) and b) allowing teachers to collaborate on SLOs.
- 3) Use student growth measures that are based on standardized test scores in novel ways—and then study the effects of these practices. The Hawaii EES system requires teachers to reflect on student growth percentile data but does not use it to directly calculate teacher effectiveness. Similarly, such data can be used in other novel ways: a) as an indication that a teacher may need additional support/coaching, reflecting reciprocal accountability by leaders for teacher growth; b) as an indication that a teacher could be tapped to serve as a teacher leader and receive additional investment (e.g., professional development on pedagogy and coaching); and/or c) as a source for teacher reflection tied to goal-setting (which could then be connected to an SLO).
- 4) Engage in continuous monitoring and improvement of teacher evaluation systems, drawing on feedback from teachers as well as other stakeholders in non-trivial ways. Disrupt asymmetrical systems of power that yield only nominal influence to teachers. Provide feedback loops that inform changes and yield some positions of influence on evaluation design teams to teachers.
- 5) Invest in providing teachers with frequent, specific, timely, and actionable feedback (e.g., from instructional coaches and informal teacher leaders) that is meaningful to them and that they can apply to their practice. This requires reciprocal accountability on the part of schools. Conversely, teacher professionalism measures should incorporate assessments of teachers' responsiveness to feedback.
- 6) Make explicit and clear expectations for teacher excellence through tools such as the DCPS Essential Practice Video Library and LEAP ongoing, team-based professional development. Again, this requires reciprocal accountability by policymakers.

We recognize the tension between making evaluation personally meaningful and maintaining systems that allow for reasonable comparison. Such comparison should not be normative—across teachers—but rather comparison to high standards that are clear and explicit.

While there is a growing body of research on teacher evaluation systems, there is more work to be done. The field needs research on the most effective ways to use standardized-test based student growth measures, as well as research on how teachers perceive and respond to various teacher evaluation systems. As mentioned previously, as people enact policy, they invariably change it (Cohen, 1990; Tyack & Cuban, 1995), just as their sensemaking is informed by policy discourses (Coburn, 2005; Coburn & Woulfin, 2012; Cohen & Hill, 2001; Rigby, 2015). Additional research is also needed into how labor markets enable and constrain teacher evaluation practices.

Within the context of ESSA, the time is ripe for policymakers to capitalize on the shift of control over teacher evaluation from the federal to state to make important changes to teacher evaluation systems. The goal should be to achieve harmony between the demands of school organizations and the general public for educator accountability with the needs of individual teachers to feel supported in their growth and development as practitioners.

References

- Akcan, S., & Tatar, S. (2010). An investigation of the nature of feedback given to pre-service English teachers during their practice teaching experience. *Teacher Development*, 14(2), 153-172. <https://doi.org/10.1080/13664530.2010.494495>
- Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. Routledge.
- Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS-EVAAS) in the Houston independent School District (HISD): Intended and unintended consequences. *Educational Policy Analysis Archives*, 20(12). <https://doi.org/10.14507/epaa.v20n12.2012>
- Bitler, M., Corcoran, S., Domina, T., & Penner, E. (2019, November). Teacher effects on student achievement and height: A cautionary tale (Working Paper No. 26480). National Bureau of Economic Research. <https://doi.org/10.3386/w26480>
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, 42(2), 231-268. <https://doi.org/10.3102/00028312042002231>
- Close, K., Amrein-Beardsley, A., & Collins, C. (2018, June). *State level assessments and teacher evaluation systems after the passage of the Every Student Succeeds Act: Some steps in the right direction*. National Education Policy Center, University of Colorado.
- Coburn, C. E. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy*, 19(3), 476-509. <https://doi.org/10.1177/0895904805276143>
- Coburn, C. E., & Woulfin, S. L. (2012). Reading coaches and the relationship between policy and practice. *Reading Research Quarterly*, 47(1), 5-30. <https://doi.org/10.1002/RRQ.008>
- Coens, T., & Jenkins, M. (2002). *Abolishing performance appraisals: Why they backfire and what to do instead*. Berrett-Koehler Publishers, Inc.
- Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, 12(3), 311-329. <https://doi.org/10.3102/01623737012003311>
- Cohen, D. K. (2011). *Teaching and its predicaments*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674062788>
- Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works*. Yale University Press. <https://doi.org/10.12987/yale/9780300089479.001.0001>

- Copland, F. (2010). Causes of tension in post-observation feedback in pre-service teacher training: An alternative view. *Teaching and Teacher Education*, 26(3), 466-472. <https://doi.org/10.1016/j.tate.2009.06.001>
- Copland, F. (2011). Legitimate talk in feedback conferences. *Applied Linguistics*, 33(1), 1-20. <https://doi.org/10.1093/applin/amr040>
- Cosner, S. (2011). Teacher learning, instructional considerations and principal communication: Lessons from a longitudinal study of collaborative data use by teachers. *Educational Management Administration & Leadership*, 39(5), 568-589. <https://doi.org/10.1177/1741143211408453>
- Crouse, K., Gitomer, D. H., & Joyce, J. (2016). An analysis of the meaning and use of student learning objectives. In K. K. Hewitt & A. Amrein-Beardsley (Eds.), *Student growth measures in policy and practice: Intended and unintended consequences of high-stakes teacher evaluations* (pp. 203-222). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-53901-4_11
- Curry, K. A., Mwavita, M., Holter, A., & Harris, E. (2016). Getting assessment right at the classroom level: Using formative assessment for decision making. *Educational Assessment, Evaluation and Accountability*, 28(1), 89-104. <https://doi.org/10.1007/s11092-015-9226-5>
- Danna, R. (2004). Diving into data analysis. *Journal of Staff Development*, 25(3), p. 24-27.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15. <https://doi.org/10.1177/003172171209300603>
- Datnow, A., & Park, V. (2014). *Data-driven leadership*. Jossey-Bass.
- Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, 34(2), 267-297. <https://doi.org/10.1002/pam.21818>
- Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G., & Van Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and Teacher Education*, 36, 1-11. <https://doi.org/10.1016/j.tate.2013.06.011>
- District of Columbia Public Schools (DCPS). (n.d.-a). *IMPACT Guidebook: Group 1*. Retrieved from <https://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/1--Teachers-Grades-4-with-Individual-Value-Added-Student-Achievement-and-Student-Survey-Data.pdf>
- DCPS. (n.d.-b). *IMPACT Guidebook: Group 2a*. Retrieved from <https://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/2a--Early-Childhood-Education-Teachers.pdf>
- DCPS. (n.d.-c). *IMPACT Guidebook: Group 2d*. Retrieved from <https://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/2d--Small-Group-Intervention-and-Grades-1-2-Teacher-LEAP-Leaders.pdf>
- DCPS. (n.d.-d). *IMPACT Guidebook: Group 10*. Retrieved from <https://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/10--Counselors.pdf>
- DCPS. (n.d.-e). IMPACT: The DCPS evaluation and feedback system for school-based personnel. Retrieved from <https://dcps.dc.gov/page/impact-dcps-evaluation-and-feedback-system-school-based-personnel>
- DCPS. (n.d.-f). LEAP: Teacher professional development. Retrieved from <https://dcps.dc.gov/page/leap-teacher-professional-development>

- Eckert, J. (2016). Measuring student growth: A view from the classroom. In K. K. Hewitt & A. Amrein-Beardsley (Eds.), *Student growth measures in policy and practice: Intended and unintended consequences of high-stakes teacher evaluations* (pp. 73-93). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-53901-4_4
- Farrell, C. C., & Marsh, J. A. (2016). Metrics matter: How properties and perceptions of data shape teachers' instructional responses. *Educational Administration Quarterly*, 52(3), 423–462. <https://doi.org/10.1177/0013161X166638429>
- Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher*, 43, 100-107. <https://doi.org/10.3102/0013189X14521864>
- Firestone, W. A., & Donaldson, M. L. (2019). Teacher evaluation as data use: What recent research suggests. Advance online publication. *Educational, Assessment, Evaluation and Accountability*. <https://doi.org/10.1007/s11092-019-09300-z>
- 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>
- Ford, T. G., Urick, A., & Wilson, A. (2018). Exploring the effect of supportive teacher evaluation experiences on U.S. teachers' job satisfaction. *Educational Policy Analysis Archives*, 26(59), 1-36. <http://doi.org/10.14507/epaa.26.3559>
- Ford, T. G., Van Sickle, M. E., & Fazio Brunson, M. (2016). The role of “informational significance” in shaping Louisiana elementary teachers' use of high-stakes teacher evaluation data for instructional decision making. In K. K. Hewitt & A. Amrein-Beardsley (Eds.), *Student growth measures in policy and practice: Intended and unintended consequences of high-stakes teacher evaluations* (pp. 117-135). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-53901-4_6
- Ford, T. G., Van Sickle, M. E., Clark, L. V., Fazio-Brunson, M., & Schween, D. C. (2017). Teacher self-efficacy, professional commitment and high-stakes teacher evaluation (HSTE) policy in Louisiana. *Educational Policy*, 31(2), 202-248. <https://doi.org/10.1177/0895904815586855>
- Gabriel, R., & Woulfin, S. (2017). *Making teacher evaluation work: A guide for literacy teachers and leaders*. Heinemann.
- Gitomer, D., Crouse, K., & Joyce, J. (2015). *A review of the DC IMPACT teacher evaluation system*. Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. The National Academies Press. Retrieved from: <https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse160397.pdf>
- Gramsci, A. (2003). From Selections from the Prison Notebooks. In V. Hodgkinson & M. W. Foley (Eds.), pp. 190-202. *Civil society reader*. University Press of New England.
- Grissom, J. A., & Youngs, P. (2016). *Improving teacher evaluation systems: Making the most of multiple measures*. Teachers College Press.
- Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, 26(1), 5-28. <https://doi.org/10.1007/s11092-013-9179-5>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hawaii Department of Education (HIDOE). (2019). *Educator Effectiveness System orientation* (online video). Retrieved from <https://vimeo.com/345581612>

- HIDOE. (n.d.-a). Educator Effectiveness System. Retrieved from <http://www.hawaiipublicschools.org/TeachingAndLearning/EducatorEffectiveness/EducatorEffectivenessSystem/Pages/home.aspx>
- HIDOE. (n.d.-b). *EES manual for evaluators and participants: 2019-2020 Educator Effectiveness System*. Honolulu, HI: Author. Retrieved from <https://www.hawaiipublicschools.org/DOE%20Forms/Educator%20Effectiveness/EESManual.pdf>
- HIDOE. (2010). Hawaii Race to the Top Application. Retrieved from <https://www2.ed.gov/programs/racetothetop/phase2-applications/hawaii.pdf>
- HIDOE. (2013, July 29). Hawaii's Race to the Top grant cleared from high-risk status. Retrieved from <http://www.hawaiipublicschools.org/ConnectWithUs/MediaRoom/PressReleases/Pages/Hawaii's-Race-to-the-Top-grant-cleared-from-high-risk-status.aspx>
- Hewitt, K. K. (2015). Educator evaluation policy that incorporates EVAAS value-added measures: Undermined intentions and exacerbated inequities. *Education Policy Analysis Archives*, 23(76). <http://doi.org/10.14507/epaa.v23.1968>
- Hill, R. & Rapp, L. (2012). School-based coaches plant seeds of learning: A districtwide approach to data analysis promotes job-embedded learning and improved teacher practice. *Journal of Staff Development*, 33(4), 36-39.
- Holloway, J., & Brass, J. (2017). Making accountable teachers: The terrors and pleasures of performativity. *Journal of Education Policy*, 33(3), 361-382. <https://doi.org/10.1080/02680939.2017.1372636>
- Huguet, A., Farrell, C. C., & Marsh, J. A. (2017). Light touch, heavy hand: Principals and data-use PLCs. *Journal of Educational Administration*, 55(4), 376-389. <https://doi.org/10.1108/JEA-09-2016-0101>
- Ingersoll, R., Merrill, L., & May, H. (2016). Do accountability policies push teachers out? *Educational Leadership*, 73(8), 44-49.
- Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher*, 44, 105-116. <https://doi.org/10.3102/0013189X15575517>
- Kane, T. J., Kerr, K. A., & Pianta, R. C. (2014). *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching Project*. Jossey-Bass. <https://doi.org/10.1002/9781119210856>
- Kelly, K. O., Ang, S. Y. A., Chong, W. L., & Hu, W. S. (2008). Teacher appraisal and its outcomes in Singapore primary schools. *Journal of Educational Administration*, 46, 39-54. <https://doi.org/10.1108/09578230810849808>
- Koehler, M. (2016, October 31). DC's teachers have a new tool for success. Will it work? *Greater Washington*, Retrieved from <https://ggwash.org/view/43361/dcs-teachers-have-a-new-tool-for-success-will-it-work>
- Lane, J. L. (2019). Maintaining the frame: Using frame analysis to explain teacher evaluation policy implementation. *American Educational Research Journal*, 57(1), 5-42. <https://doi.org/10.3102/0002831219848689>
- Lavigne, A. L. (2014). Exploring the intended and unintended consequences of high-stakes teacher evaluation on schools, teachers, and students. *Teachers College Record*, 116(1), 1-29. Retrieved from: <http://www.tcrecord.org/Content.asp?ContentId=17294>
- Lavigne, A. L., & Good, T. L. (2015). *Improving teaching through observation and feedback: Beyond state and federal mandates*. Routledge. <https://doi.org/10.4324/9781315777030>

- Lavigne, A. L., & Good, T. L. (2019). *Enhancing teacher education, development, and evaluation: Lessons learned from educational reform*. Routledge. <https://doi.org/10.4324/9781315630892>
- Lavigne, A. L., & Good, T. L. (2020). Addressing teacher evaluation appropriately: A research brief for policymakers. *APA Division 15 Policy Brief Series*, 1(2), 1-7.
- Learning Policy Institute. (2017). *What's the cost of teacher turnover?* Author. Retrieved from <https://learningpolicyinstitute.org/product/the-cost-of-teacher-turnover>
- Longo-Schmid, J. (2016). Teachers' voices: Where policy meets practice. In K. Kappler Hewitt & A. Amrein-Beardsley (Eds.), *Student growth measures in policy and practice* (pp. 49–71). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-53901-4_3
- Lortie, D. (1975). *Schoolteacher: A sociological analysis*. Chicago: University of Chicago 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>
- Mandinach, E. B. (2012). A perfect time for data-use: Using data-driven decision making to inform practice. *Educational Psychologist*, 47(2), 71–85. <https://doi.org/10.1080/00461520.2012.667064>
- Maglaras, V. (2013). Consent and submission: Aspects of Gramsci's theory of the political and civil society. *SAGE Open*, January-March, 1-8. <https://doi.org/10.1177/2158244012472347>
- Marsh, J. A., & Farrell, C. C. (2015). How leaders can support teachers with data-driven decision making: A framework for understanding capacity building. *Educational Management Administration & Leadership*, 43(2), 269–289. <https://doi.org/10.1177/1741143214537229>
- Marsh, J. A., McCombs, J. S., & Martorell, F. (2010). How instructional coaches support data-driven decision making: Policy implementation and effects in Florida middle schools. *Educational Policy*, 24, 872–907. <https://doi.org/10.1177/0895904809341467>
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education* (RAND occasional paper #OP-170-EDU). RAND. Retrieved from: http://www.rand.org/pubs/occasional_papers/OP170.html
- Marshall, C., & Rossman, G. (2016). *Designing qualitative research* (6th ed.). SAGE.
- McDonald, J. P. (1992). *Teaching: Making sense of an uncertain craft*. Teachers College Press.
- Members of the DC State Board of Education. (2018). *Teacher and principal turnover in public schools in the District of Columbia*. Author. Retrieved from <https://sboe.dc.gov/sites/default/files/dc/sites/sboe/publication/attachments/SBOE%20Teacher%20Turnover%20Report%20-%20FINAL.pdf>
- Miles, K. H., Pennington, K., & Bloom, D. (2015). *Do more, add more, earn more: Teacher salary redesign lessons from 10 first-mover districts*. Center for American Progress/ERS. Retrieved from <https://eric.ed.gov/?id=ED560060>
- Mokhtari, K., Thoma, J., & Edwards, P. (2009). How one elementary school uses data to help raise students' reading achievement. *The Reading Teacher*, 63(4), 334-337. <https://doi.org/10.1598/RT.63.4.10>
- National Education Policy Center [NEPC]. (2019, July 11). *Newsletter: This teacher evaluation trend was the reform du jour. Now states are backing away*. Author. Retrieved from: <https://nepc.colorado.edu/sites/default/files/publications/Newsletter%20teacher%20eval.pdf>
- Paufler, N. A., & Sloat, E. F. (2020). Using standards to evaluate accountability policy in context: School administrator and teacher perceptions of a teacher evaluation system. Advanced online publication, *Studies in Educational Evaluation*, 64, 1-15. <https://doi.org/10.1016/j.stueduc.2019.07.007>

- Putman, H., Ross, E., & Walsh, K. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. National Council on Teacher Quality. Retrieved from <https://www.nctq.org/publications/Making-a-Difference>
- Reddy, L. A., Dudek, C. M., Peters, S., Alperin, A., Kettler, R. J., Kurz, A. (2018). Teachers' and school administrators' attitudes and beliefs of teacher evaluation: A preliminary investigation of high poverty school districts. *Educational Assessment, Evaluation, and Accountability*, 30, 47–70. <https://doi.org/10.1007/s11092-017-9263-3>
- Rice, J. K., & Malen, B. (2016). When theoretical models meet school realities: Educator responses to student growth measures in an incentive pay program. In K. Kappler Hewitt & A. Amrein-Beardsley (Eds.), *Student growth measures in policy and practice* (pp. 29–47). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-53901-4_2
- Rigby, J. G. (2015). Principals' sensemaking and enactment of teacher evaluation. *Journal of Educational Administration*, 53(3), 374-392. <https://doi.org/10.1108/JEA-04-2014-0051>
- Rosenholtz, S. J. (1991). *Teachers' workplace: The social organization of schools*. Teachers College Press.
- 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.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.
- Ryan, R. M., & Weinstein, N. (2009). Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. *Theory and Research in Education*, 7(2), 224-233. <https://doi.org/10.1177/1477878509104327>
- Ryan, S. V., Nathaniel, P., Pendergast, L. L., Saeki, E., Segool, N., & Schwing, S. (2017). Leaving the teaching profession: The role of teacher stress and educational accountability policies on turnover intent. *Teaching and Teacher Education*, 66, 1-11. <https://doi.org/10.1016/j.tate.2017.03.016>
- Saldaña, J. (2013). *The coding manual for qualitative researchers*. SAGE.
- Schildkamp, K., Poortman, C., Luyten, H., & Ebbeler, J. (2017). Factors promoting and hindering data-based decision making in schools. *School Effectiveness and School Improvement*, 28(2), 242–258. <https://doi.org/10.1080/09243453.2016.1256901>
- Semke, H. D. (1984). Effects of the red pen. *Foreign Language Annals*, 17(3), 195-202. <https://doi.org/10.1111/j.1944-9720.1984.tb01727.x>
- Sloat, E., Amrein-Beardsley, A., & Holloway, J. (2018). Different teacher-level effectiveness estimates, different results: inter-model concordance across six generalized value-added models (VAMs). *Educational Assessment, Evaluation and Accountability*, 30(4), 367-397. <https://doi.org/10.1007/s11092-018-9283-7>
- Smith, W. C., & Kubacka, K. (2017). The emphasis of student test scores in teacher appraisal systems. *Educational Policy Analysis Archives*, 25(86), 1-29. <https://doi.org/10.14507/epaa.25.2889>
- Stringer, P. M. (2009). Capacity building for school improvement: A case study of a New Zealand primary school. *Educational Research for Policy and Practice*, 8(3), 153–179. <https://doi.org/10.1007/s10671-009-9073-6>

- 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)
- U.S. Department of Education. (2013). Race to the Top: Hawaii report year 2: School year 2011-2012. Retrieved from
<https://www2.ed.gov/programs/racetothetop/performance/hawaii-year-2.pdf>
- Vásquez, C. (2004). "Very carefully managed": Advice and suggestions in post-observation meetings. *Linguistics and Education*, 15(1-2), 33-58.
<https://doi.org/10.1016/j.linged.2004.10.004>
- Waite, D. (1993). Teachers in conference: A qualitative study of teacher-supervisor face-to-face interactions. *American Educational Research Journal*, 30(4), 675-702.
<https://doi.org/10.3102/00028312030004675>

About the Authors

Timothy G. Ford

University of Oklahoma

tgford@ou.edu

<http://orcid.org/0000-0003-2975-4191>

Dr. Ford examines the role of policy and school leadership in shaping teacher and leader working conditions and in supporting teachers' and leaders' psychological needs as learners.

Kim Kappler Hewitt

University of North Carolina at Greensboro

kkhewitt@uncg.edu

<http://orcid.org/0000-0002-5127-9625>

Dr. Hewitt research focuses on leadership preparation; instructional leadership, equity and social justice, and sustaining reform. She serves as director of the Principal Preparation for Excellence and Equity in Rural Schools (PPEERS) program, which involves a partnership with 12 rural districts.

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 63

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

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**, **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