

## What Did We Learn from Race to the Top Teacher Evaluation Systems?

Deani Thomas, MEd  
Teacher  
Ames Middle School  
Ames Community School District  
Ames, IA

Douglas Wieczorek, PhD  
Assistant Professor  
Educational Leadership, Organizations, and Policy  
School of Education  
Iowa State University  
Ames, IA

### Abstract

We reviewed and synthesized 35 peer-reviewed empirical studies published between 2014-2018 that investigated district leaders', principals', and/or teachers' interpretations of and experiences with Race to the Top teacher evaluation systems. We analyzed evidence of how educators' prior experiences, beliefs, values, organizational contexts, and community needs may have contributed to a teacher evaluation policy-practice divide. Our findings revealed three main implications for district- and systems-level instructional supervision and evaluation practices: 1) leaders should develop and provide relevant, differentiated supports for principals and teachers to implement and refine teacher evaluation processes; 2) leaders should maintain focus on stakeholders' professional relationships and a sense of community; and 3) leaders should utilize teacher evaluation processes to foster a systemic culture of professional growth.

### Key Words

district leadership, teacher evaluation, education policy, instructional leadership, accountability, instructional supervision, policy implementation

The United States government's competitive Race to the Top (RTTT) program allocated and distributed over four billion dollars to 19 states between 2009 and 2016 (American Recovery and Reinvestment Act of 2009; United States Department of Education, 2016). RTTT leveraged federal resources and state-level accountability measures meant to improve teacher effectiveness through rigorous, data-driven teacher evaluation systems connected to student achievement measures—value-added measures (VAM) (Amrein-Beadsley & Holloway, 2017; Patrick, 2016).

Previous research demonstrates how local leaders and teachers interpret education policies through their previous experiences, and adapt policy mandates to respond to local, organizational, and contextual needs (Coburn, Hill, & Spillane, 2016; Datnow, 2006; Werts, Della Salla, Lindle, Horace, Brewer, & Knoepfel, 2013).

Despite RTTT policy intentions, scholars contend that educators' individual agency, policy ambiguity, and a hierarchical, loosely-coupled educational system often contributes to a policy-practice divide (Cohen, Moffitt, & Goldin, 2007; Huffman, Pankake, & Munoz, 2006; Matland, 1995; Spillane, Parise, & Sherer, 2011).

The purpose of this article is to review and synthesize empirical studies that investigated district leaders', principals', and/or teachers' interpretations of and experiences with RTTT teacher evaluation processes.

We wanted to know how school leaders and teachers interpreted and implemented state-level RTTT teacher evaluation mandates in their local contexts, which potentially created a teacher evaluation policy-practice divide. To conduct our review we applied one guiding

question: How did public school district leaders, building leaders, and/or teachers engage with United States' RTTT program teacher evaluation policies in the context of their beliefs, previous experiences, and local school community contexts?

We do not intend to discuss the merits of RTTT policies or VAM models of teacher evaluation; rather, our goal is to inform systems-level instructional leaders who are in a primary position to develop and implement teacher evaluation systems in their respective school district contexts under the Every Student Succeeds Act (ESSA) and beyond (Coburn, Toure, & Yamashita, 2009; Derrington & Campbell, 2015; Rorrer & Skrla, 2005).

Our analysis shows that educational actors, including teachers, principals, and superintendents, are not opposed or resistant to evaluation systems.

However, evaluation systems emphasize proving competency rather than using performance indicators to support continued growth. If teachers and principals shift their focus from improving their practices to proving their competence, their thinking about student learning might also shift.

There are potentially dangerous, albeit unintended, consequences associated with shifting the collective mindset of the educational system to focus strictly on proving results, rather than striving to continuously improve.

Our methods and analysis were framed by two areas of theory and research: 1) theories of action which framed RTTT teacher evaluation policies, and 2) constructivist perspectives of educators' interpretation of and engagement with education policies.

## **RTTT Teacher Evaluation Policy Theory of Action**

Implemented as an optional, state-level policy inducement, RTTT stated four broad school improvement goals to improve student learning outcomes:

1. Adopting standards and assessments that prepare students to succeed in college and the workplace and to compete in the global economy;
2. Building data systems that measure student growth and success, and inform teachers and principals how they can improve instruction;
3. Recruiting, developing, rewarding, and retaining effective teachers and principals, especially where they are needed most; and
4. Turning around our lowest-achieving schools. (United States Department of Education, 2009, p. 2)

A policy inducement is defined as “the transfer of money to individuals or agencies in return for certain actions” (McDonnell & Elmore, 1987, p. 134). As the primary focus of our inquiry, RTTT funded states were required to develop and implement teacher evaluation systems in all community school districts and local schools to address RTTT goal three (McGuinn, 2012).

RTTT teacher evaluation systems were characterized by summative judgements of teachers’ annual performance, which included “common quality criteria that enable[d] quantifiable comparisons or classifications across a group” of teachers (Mintrop, Ordonez, Coghlan, Pryor, & Madero, 2018). Teachers’ professional performance and instructional quality criteria were articulated and guided by instructional practice protocols and evaluative

scoring rubrics [for a comprehensive summary and review see Gilmour, Majeika, Sheaffer, & Wehby (2018)], which principals used as tools to document classroom observational evidence, collect and evaluate artifacts of teaching practice, and determine teachers’ professional competence (Bradford & Braaten, 2018). This approach leveraged data-informed evaluations and performance ratings to motivate district leaders, principals, and teachers to mutually develop and implement instructional changes at the classroom, school, and district levels (Firestone, 2014).

Our analysis focused on how educators’ personal values and beliefs, professional perspectives, individual opinions, and shared organizational priorities and cultures influenced teacher evaluation implementation at the local level (Spillane, 2012).

## **A Constructivist Perspective of Teacher Evaluation Policy**

We employed a constructivist analytical perspective (Lambert, Walker, Zimmerman, Cooper, Lambert, Garner, & Sazabo, 2002), which describes how leaders’ and teachers’ previous experiences, beliefs, and values influenced their agency to interpret and enact RTTT teacher evaluation policy mandates in local contexts (Bridwell-Mitchell & Sherer, 2017; Cobb & McClain, 2006; Moran, 2017).

Previous empirical and theoretical research has demonstrated how local school leaders and teachers are simultaneously subjected to, and also act as instruments of, education policy implementation (Honig, 2006). Our theoretical approach accounts for the influence of local districts’ organizational and community contexts (Rorrer, Skrla, & Scheurich, 2008), and explains how leaders and teachers interpreted and implemented federal and state education policies to suit local priorities and respond to their stakeholders’

needs (Coburn, 2001; Honig & Coburn, 2008; Schecter & Shaked, 2017). Our analysis describes leaders' and teachers' perspectives and actions which potentially contributed to a RTTT teacher evaluation policy-practice divide.

## Methods

We applied Hallinger's (2013) literature review model to conduct a five stage literature review synthesis and analysis. In the following sections we describe our procedures.

### Article search processes

In stage one, we searched the Educational Resources Information Center (ERIC) and EBSCO databases using the broad search term "Race to the Top" and refining the searches to include peer-reviewed publications from 2014-2018. We also searched the table of contents of each of twenty-six online education journal issues during that time frame. Our initial search yielded a total of 135 article titles.

### Data inclusion criteria and screening procedures

In stage two, we read each article abstract to identify empirical studies. Then we reviewed the article's background, introduction, and research questions to determine each study's purpose. We retained qualitative or quantitative studies that purposefully investigated district leaders, principals, or teachers expressed interpretations of, or observed experiences with, RTTT teacher evaluation systems. We screened each study for indicators of empirical rigor: a clearly defined theoretical or conceptual framework, descriptions of methods relevant to the design, and a discussion of results or findings in the context of current literature.

### Determination of article relevance

In stage three, we reviewed each study's findings or results, discussion, conclusion, and

implications sections, to determine if the study yielded empirical evidence that was relevant to our guiding questions and theoretical perspective. This was important particularly for qualitative studies which often describe unanticipated, emergent findings or implications.

Our review includes a total of 35 studies. Of these, 12 investigated teachers, 15 investigated principals, and one focused solely on superintendents. Five studies included a combination of principals, superintendents, and teachers. Two articles did not specify a target sample group, but we retained them because leaders and teachers were included in the findings or discussion sections. In our sample, 24 studies used qualitative methods, six used quantitative methods, and five used mixed methods.

### Data analysis

In stage four, we applied a textual and content analysis methodology to organize, reduce, and analyze our data set into a literature review synthesis table format (Miles & Huberman, 1994). We collected text from each study's findings, discussion, conclusions, and implications sections and placed the data into Excel spreadsheets organized by participant type: district leaders, principals, and teachers. Within each sheet, we coded entries as priori evidence of educators' constructed thoughts, beliefs, values, conceptions, descriptions, or reflections regarding teacher evaluation processes.

### Second-stage data analysis and confirmation

In stage five, we completed a second round of open and inductive keyword coding within each participant group to develop coded categories across multiple studies (Taylor & Bogdan, 1998). These final coded categories provided us with contextual exemplars and descriptions of educators' interpretations and

experiences which were evident across multiple settings and cases. To ensure inter-rater validity and reliability in our analysis, we consulted with each other to determine thresholds of evidence, resolve contradictory evidence and provide a check on our subjectivity.

### **Methodological limitations**

Our review is constrained by several limitations. First, we only included published, peer-reviewed, empirical journal articles in our analysis. This criteria leaves out potentially significant research published as dissertations, book chapters, research reports, conference papers, or white papers. Second, we acknowledge potential errors in the search process and instances where we were not able to locate and review relevant articles.

Third, research about RTTT will continue to be published during the next 12 to 18 months due to the research publication lag time that exists in academia; additional evidence may be revealed in these studies which is not included here. Fourth, we acknowledge inherent researcher error and inconsistency in our coding and analysis. To the extent possible, we have mitigated these errors through transparent descriptions of our research and analysis procedures.

### **Findings**

Our goal was to provide research-based evidence of a RTTT teacher evaluation policy-practice divide, demonstrated through district leaders', principals', and teachers' experiences in local contexts. Our analysis and findings describe five thematic categories which were shared primarily between principals and teachers: 1) stakeholders' sensemaking of increased policy and professional demands; 2) stakeholders' perceptions of systemic reliability and consistency; 3) balancing accountability with teacher growth and learning; 4) teachers'

confidence and receptivity to new requirements; and 5) teachers' satisfaction and retention. Within each category we provided selective citations and examples from the 35 sampled research articles to demonstrate our claims.

### **Stakeholders' sensemaking of increased policy and professional demands**

Principals had to adapt to the new demands, and they attempted to integrate policy tools into their work, spent longer days working at school, took work home with them, and delegated administrative tasks to teacher leaders (Derrington & Campbell, 2015, 2018).

The increased demands on principals' time created by the new evaluation systems translated into principals spending less time in classrooms and interacting face-to-face with students and staff (Derrington & Campbell, 2014, 2015, 2017; Lavigne & Chamberlain, 2017). Evidence shows how principals value instructional leadership, and Derrington and Campbell (2017) found that over time, principals began to adjust aspects of new evaluation systems that they did not perceive as directly supportive of instructional leadership.

These adjustments were based on principals' professional knowledge and relationships with teachers and included modifying scoring rubrics based on contextual needs. Principals' efforts did not always translate into effective instructional leadership, and principals struggled to provide comprehensive feedback to all teachers (Reinhorn, Johnson, & Simon, 2017).

The critical importance to learning of relationships may be lost on policymakers, but it is not lost on principals or teachers. Learning requires vulnerability, vulnerability necessitates trust, and trust is founded upon quality relationships. Principals tended to focus their



attention on the formative aspects of the evaluation systems (Derrington & Campbell, 2017; Reinhorn et al., 2017) and emphasized the value of positive relationships (Robertson-Kraft & Zhang, 2016; Wieczorek & Theoharis, 2015). Collaboration and feedback, as achieved through pre- and post-observation conferences, were perceived by principals as highly beneficial (Donaldson & Papay, 2015; Reddy et al., 2018; Williams & Herbert, 2017).

Principals acknowledged the importance of positive relationships and shared a variety of approaches they used to nurture relationships such as reminding teachers of past successes, promoting unity, emphasizing the importance of the work to students and communities, and employing interpretive frameworks (Gawlik, 2017; Wieczorek & Theoharis, 2015). Reid (2017) found that some principals also attempted to preserve positive relationships with teachers by giving more favorable evaluation ratings, which called into question the reliability of the new evaluation systems.

### **Stakeholders' perceptions of systemic reliability and consistency**

Concerns over the reliability of new evaluation systems (Herlihy et al., 2014) were not surprising given that required training for principals was minimal (Derrington & Campbell, 2017).

Administrators were confused about their roles (Williams & Herbert, 2017), and principals who did receive training struggled most with tasks associated with establishing reliability in scoring observations (Lavigne & Chamberlain, 2017). Additional reliability concerns were related to perceived inconsistencies inherent in the evaluation systems. While some principals expressed that the evaluation frameworks could align expectations across districts (Wieczorek, Clark,

& Theoharis, 2018b), other principals and superintendents began to perceive and express concerns about the inconsistencies between schools and districts due to the subjective nature of the evaluation systems (Derrington, & Campbell, 2014, 2015, 2017). Derrington and Campbell (2017) provided a detailed overview of the ways in which principals' thinking evolved over time to reflect a combination of value for the evaluation tools and concerns about consistency between principals and across content areas.

The use of a single tool to evaluate all teachers was concerning to principals since the tools were more applicable to some content areas and grade levels than others (Derrington & Campbell, 2017, 2018; Wieczorek, Clark, & Theoharis, 2018a), and concerns about the use of a single tool were even more pronounced when only portions of a tool were prescribed.

For example, Williams and Herbert (2017) found in a study of ten Louisiana principals that principals were concerned that the adoption of only certain components of the Danielson framework had a negative impact on the scoring system because the adopted components did not apply equally well across all subject areas and grade levels.

An additional area of concern expressed by principals related to inconsistencies between student achievement data and observation data (Derrington & Campbell, 2017, 2018) that might be explained in part by changes teachers make to their instruction for evaluation observations (Ford, 2018; Williams & Herbert, 2017).

Teachers want to be perceived as competent and successful, and this desire becomes even more pronounced when their livelihood depends on perceptions of their competence. The requirements for effective

instruction outlined in evaluation tools such as rubrics might be ideal, but they might also be unrealistic for any human to implement on a continual basis. An emphasis on documentation for accountability purposes might be at odds with gaining insights to support professional growth (Shirrell, 2016).

Since teachers' livelihood under accountability systems is based largely on what they are observed doing, it is logical for teachers to attempt to do their best to make their performance match the expectations of evaluation tools regardless of whether it matches their daily practices or their beliefs about quality instruction (Ford, 2018; Williams & Herbert, 2018).

The issues of inconsistency and relevance, combined with the high-stakes nature of the evaluation systems, led principals to question the fairness of the new evaluation systems despite their acceptance of the systems' permanence (Derrington & Campbell, 2018; Flores & Derrington, 2017).

### **Balancing accountability with teacher growth and learning**

The stated mission of most schools is to promote civic, emotional, and cognitive development (Stemler, Bebel, & Sonnabend, 2011). Accountability-based evaluation systems, however, measure learning as a finite outcome and shift the focus from growth to ratings. This was reflected in principals' concerns that teacher evaluation systems interfered with teachers' growth as teachers became more focused on evaluation ratings than instruction (Wieczorek, Clark, & Theoharis, 2018b).

Principals and teachers alike valued learning and understood that data provide insights into learning, but they also understood that learning is complex and cannot be reduced

to numbers (Bradford & Braaten, 2018). Principals acknowledged that student achievement data were important, but they placed less importance on student achievement data than the evaluation systems might suggest (Lavigne & Chamberlain, 2017), which makes sense because estimates for student growth models are complicated by the prevalent nonrandom grouping and assignment of students (Dieterle, Guarino, Reckase, & Wooldridge, 2014). Wilcox and Lawson (2018) found support for these assertions through focus groups with 143 teachers who reported that they were preoccupied with the relationship between test scores and performance evaluations.

While not a lot is known from the available literature about teachers' thinking about new evaluation systems, the literature suggests that the interconnections between student achievement data, teacher evaluations, and instruction were not lost on teachers. For example, McDuffie et al., (2017) found through interviews with twenty-four middle school math teachers that the teachers were concerned about the use of student achievement data in evaluations and worried that state assessments would detract from quality instruction.

Due to the high-stakes nature of new evaluation systems and the influence of student achievement data on teacher evaluations, teachers tended to use evaluation frameworks as proxies for understanding Common Core State Standards (McDuffie et al., 2017; Stosich, 2017).

A potentially unfortunate and unintended consequence of accountability-based evaluation systems is the loss of confidence among educators in their own professional judgment (Bradford & Braaten, 2018). When teachers replace efforts to undertake the complex task of understanding

quality instruction with the menial task of conforming to checklists of instructional actions, rich educational experiences are sacrificed, and student learning suffers. For example, Ford (2018) found in a study of 32 Louisiana teachers that the teachers tended to focus their attention on making superficial instructional changes that they believed were most likely to improve their evaluation ratings. Interestingly, teachers and principals appeared to use evaluation tools in parallel ways.

While teachers were using the tools as proxies for professional growth based on genuine understanding of quality instruction, principals were using the tools as proxies for genuine understanding of quality feedback (Wieczorek, Clark, & Theoharis, 2018a). The ways in which teachers and principals used evaluation tools represents the shift described by Holloway and Brass (2018) from education actors conceiving of accountability apparatus as external to relying on accountability apparatus to define themselves as “transmitters of pre-determined standards and the ones responsible for delivering content correctly and objectively” (p. 378).

### **Teachers’ confidence and receptivity to new requirements**

While the reported actions of teachers seem to indicate that they care about evaluation systems, teachers generally expressed neutral attitudes about evaluation systems (Kowalski & Dolph, 2015; Reddy, Dudek, Peters, Alperin, Kettler, & Kurz., 2018). Novice teachers tended to be more receptive than veteran teachers (Robertson-Kraft & Zhang, 2016), and principals perceived and reported that veteran teachers experienced insecurities related to the new evaluation systems (Wieczorek & Theoharis, 2015).

Insecurities among veteran teachers made sense because they had rarely been

included in the evaluation process. Teachers were more receptive to evaluation systems when they believed that the process could improve their instruction, and they identified detailed feedback based on observations, collaborative communication, alignment with personal values, and an emphasis on professional growth as helpful to them in the evaluation process (Donahue & Vogel, 2018; Lawson et al., 2017; Reddy et al., 2018; Reinhorn et al., 2017; Robertson-Kraft & Zhang, 2016).

Teachers want to be respected as professionals, but reforms like new evaluation systems can come across as demoralizing when approached as something done *to* teachers rather than something done *with* teachers. Lawson et al. (2017) found that successful districts resisted the temptation to rely on a compliance-oriented, top-down approach to implementation and instead used collaborative communication to preserve teachers’ autonomy, which had a positive impact on teachers’ trust of school leaders and the evaluation process.

### **Teachers’ satisfaction and retention**

New teacher evaluation systems were designed with the intent of holding teachers accountable for effective instruction and potentially dismissing ineffective teachers from the profession.

While overall teacher retention patterns appeared to be unaffected by new teacher evaluations systems, new evaluation systems did appear to encourage ineffective teachers to leave, teachers with more “grit” to stay, and some retained teachers to strive to become more effective (Dee & Wyckoff, 2017; Robertson-Kraft & Zhang, 2016).

For example, Dee and Wyckoff (2017) found that the District of Columbia’s high-



stakes evaluation system, IMPACT, resulted in the replacement of many ineffective teachers as well as increased effectiveness among retained teachers, both of which resulted in student achievement gains. It is less clear whether evaluation systems like IMPACT led to the retention of the most effective teachers and whether teachers' job satisfaction was influenced by the systems.

Considering potential relationships between job satisfaction and teacher retention is important since job satisfaction likely has some influence over teachers' decisions to remain in the profession. Wright, Shields, Black, Banerjee, and Waxman (2018) found that while curricular and pedagogical autonomy were significantly lessened for teachers in RTTT states, job satisfaction did not differ among teachers in RTTT states and those in non-RTTT states; these findings did not isolate teacher evaluations as a variable in examining job satisfaction.

Koedel, Springer, and Tan (2017) found a correlation between job satisfaction and evaluation ratings with teachers who had higher ratings expressing higher levels of job satisfaction.

Of course, the correlation between evaluation ratings and job satisfaction could be explained in a variety of ways, and it seems unlikely that higher ratings alone directly cause teachers to be more satisfied.

### **Discussion and Implications for Practice**

A constructivist theoretical perspective situates leaders' and teachers' previous experiences, beliefs, values, and community concerns as significant influences on how they implement teacher evaluation systems.

The answers to our essential question, "What did we learn from RTTT teacher evaluation systems?" indicate evidence of a policy-practice divide in the context of leaders' and teachers' experiences at the local level. Systems-level school leaders need to acknowledge and understand how pre-existing professional cultures, individuals' previous experiences, and stakeholders' beliefs influence the teacher evaluation implementation process.

Based on our analysis, there are three main implications for practice that can inform systems leaders' efforts to develop, implement, or refine teacher evaluation processes in their local district contexts.

#### **Develop and provide differentiated support for implementation**

*Systems-level leaders should develop and implement supportive structures, procedures, and training to integrate teacher evaluation policies into organizational routines.*

Stakeholders need differentiated types of pedagogical and leadership support which includes specific resources and modes of preparation to effectively implement teacher evaluation processes.

Principals need sufficient support, training, and coaching to develop their understandings of and skills to supervise and evaluate instruction. Policy tools and instruments provide general, basic structures and guidelines, but systems-level leaders need to enhance or adapt tools to meet principals' and teachers' needs.

Despite the widespread use of instructional protocols and rubrics, they fall short to provide relevant, content-specific pedagogical support or guidance. Teachers

need to be essential partners in, and not a target of, the development and implementation of the teacher evaluation process as part of continuous improvement efforts.

Teacher evaluation policy instruments and procedures need to be relevant to teachers' content area pedagogy and grade level standards to support instructional meaningful change and professional development opportunities.

### **Maintain focus on professional relationships and a sense of community**

*Systems-level school leaders should consider the ways in which teacher evaluation systems can provide an opportunity to strengthen professional relationships, organizational culture, and climate.*

High-stakes teacher evaluation models can potentially trigger negative individual and organizational stressors which hurt professional culture.

Evidence demonstrates that principals and teachers rely on collaboration, trust, and relationships to develop and implement teacher evaluation processes at the local level.

Emotion and relationships drive the bus in a person-centered industry like education, so successful implementation of new initiatives require purposeful attention to stakeholders' emotional state and capitalizing on professional community and relationships.

The power of belief and emotion in the change process comes from authenticity in relationships.

Positive changes can be promoted when leaders emotionally support those they lead through the process of connecting their existing

beliefs and current emotions to proposed changes.

Teacher evaluation requires more than technical knowledge and efficient procedural implementation, and systems leaders should be cognizant of the affective and relational impacts on stakeholders' experiences.

### **Utilize teacher evaluation to foster a culture of professional growth**

*Systems-level school leaders need to emphasize teacher evaluation as an opportunity to develop a systemic, professional culture of growth and continuous improvement.*

Teacher evaluation driven by accountability, and professional growth, are sometimes viewed by stakeholders as incompatible goals.

Professional growth is a higher priority for stakeholders, and the extent to which principals and teachers believe the system is beneficial to practice, and subsequently to student learning, is essential.

Evaluation scores and ratings are not motivating factors to spur principals or teachers to change instructional practices, and are likely disruptive to collaborative, meaningful dialogue around professional growth.

Faced with new technical and cultural challenges, both principals and teachers did not want to get bogged down in the bureaucratic aspects of the policy, and focused on developing professional cultures which valued growth and improvement of practice.

Even in cases where teachers and principals knew the potential of negative professional consequences, they still valued the potential of the tools and system mandates to

improve practices and impact student learning. If principals and teachers do not believe the system can improve teaching at the classroom, school, or district levels, then they will simply ignore the policy, or treat the system as a compulsory obligation.

Absent system-level leadership bridging accountability policies with principals' and teachers' values, principals and teachers will not invest the time and energy to develop and implement meaningful changes to teaching and leadership.

### **Author Biographies**

Deani Thomas is a middle school special education teacher, instructional model, and mentor teacher leader in the Ames School District. She is a third-year doctoral student studying educational leadership and literacy education at Iowa State University. E-mail: [deani.thomas@ames.k12.ia.us](mailto:deani.thomas@ames.k12.ia.us) or [dtthomas@iastate.edu](mailto:dtthomas@iastate.edu)

Douglas Wieczorek is an assistant professor in the educational leadership, organizations, and policy program at Iowa State University. He teaches courses and conducts research on systems and building level leadership, specifically instructional leadership, leading school reform, and education policy implementation. E-mail: [dwieczor@iastate.edu](mailto:dwieczor@iastate.edu)

## References

- American Recovery and Reinvestment Act of 2009* (ARRA), Section 14005-6, Title XIV, (Public Law 111-5). Retrieved from <https://www2.ed.gov/policy/gen/leg/recovery/statutory/stabilization-fund.pdf>
- Amrein-Beardsley, A., & Holloway, J. 2017. Value added models for teacher evaluation and Accountability: Commonsense assumptions. *Educational Policy*, 1-27. DOI: 10.1177/0895904817719519
- \*Bradford, C., & Braaten, M. (2018). Teacher evaluation and the demoralization of teachers. *Teaching and Teacher Education*, 75, 49-59.
- Bridwell-Mitchell, E.N., & Sherer, D.G. (2017). Institutional complexity and policy implementation: How underlying logics drive teacher interpretations of reform. *Educational Evaluation and Policy Analysis*, 39(2), 223-247.
- \*Carrier, L. L., & Whaland, M. (2017). Left behind by policy: A case study of the influence of high stakes accountability policy on data-based decision making in one small, rural New Hampshire school. *Rural Educator*, 38(3), 12-26.
- Cobb, P., & McClain, K. (2006). The collective mediation of a high-stakes accountability program: Communities and networks of practice. *Mind, Culture, and Activity*, 13(2), 80-100.
- Coburn, C.E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145-170.
- Coburn, C.E., Hill, H.C., & Spillane, J.P. (2016). Alignment and accountability in policy design and implementation: The common core state standards and implementation research. *Educational Researcher*, 45(4), 243-251.
- Coburn, C.E., Toure, J., & Yamashita, M. (2009). Evidence, interpretation, and persuasion: Instructional decision making at the district central office. *Teachers College Record*, 111(4), 1115-1161.
- Cohen, D.K., Moffitt, S.L., & Goldin, S. (2007). Policy and practice: The dilemma. *American Journal of Education*, 113, 515-548.
- Datnow, A. (2006). Connections in the policy chain: The “Co-construction” of implementation in comprehensive school reform. In Honig, M.I. (Ed.), *New directions in education policy implementation*, pp. 105-124. Albany, NY: State University of New York Press.

- \*Dee, T., & Wyckoff, J. (2017). A lasting impact: High-stakes teacher evaluations drive student success in Washington, D.C. *Education Next*, 17(4), 58-66.
- \*Derrington, M. L., & Campbell, J. W. (2018). High-stakes teacher evaluation policy: US principals' perspectives and variations in practice. *Teachers and Teaching: Theory and Practice*, 1-17.
- \*Derrington, M. L., & Campbell, J. W. (2017). Teacher evaluation policy tools: Principals' selective use in instructional leadership. *Leadership and Policy in Schools*, 1-23.
- \*Derrington, M. L., & Campbell, J. W. (2015). Principal concerns and superintendent support during teacher evaluation changes. *ASSA Journal of Scholarship and Practice*, 12(3), 11-22.
- \*Derrington, M. L. (2014). Teacher evaluation initial policy implementation: Superintendent and principal perceptions. *Planning and Changing*, 45(1/2), 120-137.
- \*Dieterle, S., Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2014). How do principals assign students to teachers? Finding evidence in administrative data and the implications for value added. *Journal of Policy Analysis and Management*, 34(1), 32-58.
- \*Dodson, R. L. (2017). An analysis of principals' perceptions of the primary teaching evaluation system used in eight U.S. states. *International Journal of Education Policy and Leadership*, 12(5), 1-22.
- \*Donahue, E., & Vogel, L. R. (2018). Teacher perceptions of the impact of an evaluation system on classroom instructional practices. *Journal of School Leadership*, 28(1), 31-55.
- \*Donaldson, M. L., & Papay, J. P. (2015). An idea whose time had come: Negotiating teacher evaluation reform in New Haven, Connecticut. *American Journal of Education*, 122, 39-70
- \*Durand, F. T., Lawson, H. L., Campbell Wilcox, K., & Schiller, K. S. (2016). The role of district office leaders in the adoption and implementation of the Common Core State Standards in elementary schools. *Educational Administration Quarterly*, 52(1), 45-74.
- Firestone, W.A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher*, 43(2), 100-107.
- \*Flores, M., A., & Derrington, M.L. (2017). School principals' views of teacher evaluation policy: Lessons learned from two empirical studies. *International Journal of Leadership in Education*, 20(4), 416-431.
- \*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*, 1-33.



- \*Gawlik, M. A. (2015). On the frontline of CCSS implementation: A national study of factors influencing teachers' perceptions of teaching conditions and job satisfaction. *Journal of Educational Administration*, 53(2), 398-415.
- Gilmour, A.F., Majeika, C.E., Sheaffer, A.W., & Wehby, J.H. (2018). The coverage of classroom management in teacher evaluation rubrics. *Teacher Education and Special Education*, 1-14. <https://doi.org/10.1177/0888406418781918>
- Hallinger, P. (2013). A conceptual framework for systematic reviews of research in educational leadership and management. *Journal of Educational Administration*, 51(2), 126-149.
- \*Herlihy, C., Karger, E., Pollard, C., Hill, H. C., Kraft, M. A., Williams, M., & Howard, S. (2014). State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teachers College Record*, 116(1).
- \*Holloway, J., & Brass, J. (2018). Making accountable teachers: The terrors and pleasures of performativity. *Journal of Education Policy*, 33(3), 361-382.
- Honig, M.I. (2006). Complexity and policy implementation: Challenges and opportunities for the field. In Honig, M.I. (Ed.), *New directions in education policy implementation*, pp. 1-24. Albany, NY: State University of New York Press.
- Honig, M.I., & Coburn, C. (2008). Evidence-based decision making in school district offices: Toward a policy and research agenda. *Educational Policy*, 22(4), 578-608.
- Huffman, J.B., Pankake, A., & Munoz, A. (2006). The tri-level model in action: Site, district, and state plans for school accountability in increasing student success. *Journal of School Leadership*, 16, 569-582.
- \*Koedel, C., Li, J., Springer, M. G., & Tan, L. (2017). The impact of performance ratings on job satisfaction for public school teachers. *American Educational Research Journal*, 54(2), 241-278.
- \*Kowalski, T. J., & Dolph, D. A. (2015). Principal dispositions regarding the Ohio teacher evaluation system. *ASSA Journal of Scholarship and Practice*, 11(4), 4-20.
- Lambert, L. Walker, D., Zimmerman, D.P., Cooper, J.E., Lambert, M.D., Gardner, M.E., & Sazabo, M. (2002). *The constructivist leader* (2nd ed.). New York: Teachers College Press and the National Staff Development Council.
- \*Lavigne, A. L., & Chamberlain, R. W. (2017). Teacher evaluation in Illinois: School leaders' perceptions and practices. *Educational Assessment, Evaluation and Accountability*, 29, 179-209

- J. (2017). The role of district and school leaders' trust and communications in the simultaneous implementation of innovative policies. *Journal of School Leadership*, 27(1), 31-67.
- Matland, R.E. (1995). Synthesizing the implementation literature: The ambiguity-conflict model of policy implementation. *Journal of Public Administration Research and Theory*, 5(2), 145-174.
- McDonnell, L.M., & Elmore, R.F. (1987). Getting the job done: Alternative policy instruments. *Educational Evaluation and Policy Analysis*, 9(2), 133-152.
- \*McDuffie, A. R., Drake, C., Choppin, J., Davis, J. D., Magana, M. V., & Carson, C. (2017). Middle school mathematics teachers' perceptions of the Common Core State Standards for Mathematics and related assessments and teacher evaluation systems. *Educational Policy*, 31(2), 139-179.
- McGuinn, P. (2012). Stimulating reform: Race to the Top, competitive grants, and the Obama education agenda. *Educational Policy*, 26(1), 136–159.
- Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook*. Beverly Hills, CA: Sage.
- Mintrop, R., Ordonez, M., Coghlan, E., Pryor, L., & Madero, C. (2017). Teacher evaluation, pay for performance, and learning around instruction: Between dissonant incentives and resonant procedures. *Educational Administration Quarterly*, 54(1), 3-46.
- \*Moran, R. M. R. (2017). The impact of a high stakes teacher evaluation system: Educator perspectives on accountability. *Educational Studies*, 53(2), 178-193
- Muenich, J. A. (2014). A study of how secondary principals in Minnesota perceive the evaluation of their performance. *NASSP Bulletin*, 98(4), 380-309.
- Patrick, D. (2016). Problems with percentiles: Student growth scores in New York's teacher evaluation system. *AASA Journal of Scholarship and Practice*, 12(4), 32-49.
- \*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.
- \*Reid, D. B. (2017). Shared leadership: A comparative case study of two first year US principals' socialization around teacher evaluation policy. *Educational Management Administration and Leadership*, 1-14.
- \*Reinhorn, S. K., Johnson, S. M., & Simon, N. S. (2017). Investing in development: Six high-

performing high-poverty schools implement the Massachusetts Teacher Evaluation Policy. *Educational Evaluation and Policy Analysis*, 39(3), 383-406.

- \*Robertson-Kraft, C., & Zhang, R. S. (2016). Keeping great teachers: A case study on the impact and implementation of a pilot teacher evaluation system. *Educational Policy*, 1-32.
- Rorrer, A.K., & Skrla, L. (2005). Leaders as policy mediators: The reconceptualization of accountability. *Theory into Practice*, 44(1), 53-62.
- Rorrer, A.K., Skrla, L., & Scheurich, J.J. (2008). Districts as institutional actors in reform. *Educational Administration Quarterly*, 44(3), 307-358.
- Schechter, C., & Shaked, H. (2017). Leaving fingerprints: Principals' considerations while implementing education reforms. *Journal of Educational Administration*, 55(3), 242-260.
- \*Shirrell, M. (2016). New principals, accountability, and commitment in low-performing schools. *Journal of Educational Administration*, 54(5), 558-574.
- Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision making phenomena. *American Journal of Education*, 118(2), 113e141.
- Spillane, J.P., Parise, L.M., & Sherer, J.Z. (2011). Organizational routines as coupling mechanisms: Policy, school administration, and the technical core. *American Educational Research Journal*, 48(3), 586-619.
- Stemler, S. E., Bebell, D., & Sonnabend, L. A. (2011). Using school mission statements for reflection and research. *Educational Administration Quarterly*, 47(2), 383-420.
- \*Stosich, E. L. (2017). Principals and teachers "craft coherence" among accountability policy. *Journal of Educational Administration*, 56(2), 203-219.
- Taylor, S.J., & Bogdan, R. (1998). *Introduction to qualitative research methods: A guidebook and resource*, (3rd ed.). New York: John Wiley & Sons.
- United States Department of Education. (2009). *Race to the Top program executive summary*. Author: Washington: DC. Retrieved from <https://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- United States Department of Education. (2016). *Race to the Top Funds. Awards: Scopes of work decision letters*. Retrieved from <https://www2.ed.gov/programs/racetothetop/awards.html>
- Werts, A.B., Della Sala, M., Lindle, J., Horace, J.M. Brewer, C., & Knoeppel, R. (2013). Education stakeholders' translation and sense-making of accountability policies. *Leadership and Policy in Schools*, 12(4), 397-419.

- \*Wieczorek, D., & Theoharis, G. (2015). “We’re going to make lemonade out of lemons”: Urban principals, emotion, and Race to the Top implementation. *NASSP Bulletin*, 99(4), 281-303.
- \*Wieczorek, D., Clark, B., & Theoharis, G. (2018a). Principals’ instructional feedback practices during Race to the Top. *Leadership and Policy in Schools*.  
<http://www.tandfonline.com/doi/full/10.1080/15700763.2017.1398336>
- \*Wieczorek, D., Clark, B., & Theoharis, G. (2018b). Principals’ perspectives of a Race to the Top evaluation system. *Journal of School Leadership*, 28(5), 566-595.
- \*Wilcox, K. C., & Lawson, H. A. (2018). Teachers’ agency, efficacy, engagement, and emotional resilience during policy innovation implementation. *Journal of Educational Change*, 19(2), 181-204.
- \*Williams, K., & Herbert, D. (2017). Secondary school administrators' perceptions of Louisiana's Compass System. *AASA Journal of Scholarship and Practice*, 14(1), 19-30.
- \*Wright, K. B., Shields, S. M., Black, K., Banerjee, M., & Waxam, H. C. (2018). Teacher perceptions of influence, autonomy, and satisfaction in the early Race to the Top era. *Education Policy Analysis Archives*, 26(62), 1-24.
- \* = denotes a study included in the literature review analysis