Organization and School Leadership Faculty Publications and Presentations

College of Education and P-16 Integration

2020

A Thirty State Analysis of Teacher Supervision and Evaluation Systems in the ESSA Era

Ian M. Mette

Israel Aguilar
The University of Texas Rio Grande Valley

Doug Wieczorek

Follow this and additional works at: https://scholarworks.utrgv.edu/org_fac

Part of the Educational Leadership Commons

A Thirty State Analysis of Teacher Supervision and Evaluation Systems in the ESSA Era

Ian M. Mette[1], Israel Aguilar[2], and Doug Wieczorek[3]

Abstract

We analyzed teacher supervision and evaluation policy systems in 30 states since the passage of the Every Student Succeeds Act (ESSA) of 2015 in the United States (US). This qualitative study of state ESSA policy documents and legislation examined how teacher supervision and evaluation systems (TSES) models have been developed under ESSA, specifically regarding how the construction of TSES models conflated formative feedback with summative evaluation. Despite evolving federal-level and state-level education accountability policies spurred by No Child Left Behind (NCLB) in 2001, we argue that TSES systems are influenced by state-level historical political culture (Elazar, 1994; Fowler, 2013), workplace behaviorism (Hazi, 2019), decision-making structures (Hazi & Arredondo Rucinski, 2009; Ruff, 2019), and policy rationalism (Louis et al., 2008; Orr, 2007). Data were analyzed inductively (Wolcott, 2009) to investigate how 30 states developed TSES models and from this we analyze the messages conveyed about improvement. Thus, while ESSA intended to provide states and local districts with more political control to develop and implement TSES models, our analyses shows how ESSA has extended and reinforced state-level TSES policy development and reduced districts' local control and authority to supervise and evaluate instruction.

Keywords

teacher supervision and evaluation systems; supervision; political culture; technorationalism; workplace behaviorism

---

[1] University of Maine, Maine, USA
[2] University of Texas Rio Grande Valley, Texas, USA
[3] Iowa State University, Iowa, USA

Corresponding Author:
Ian M. Mette (School of Educational Leadership, Higher Education, and Human Development, College of Education and Human Development, University of Maine, 334 Merrill Hall, Orono, ME 04401, USA)
Email: ian.mette@maine.edu

Introduction

This study examines the potential role that state-level political cultures play on teacher supervision and evaluation system (TSES) development in the United States (US) by comparing TSES policy documents and legislation enacted after the passage of the Every Student Succeeds Act (ESSA) across 30 states. We sought to understand the ways in which states developed policy guidance for teacher supervision and evaluation practices in the beginning of the ESSA era. As such, this work analyzed how a set of 30 purposefully selected states applied the embedded policy rationalism of ESSA to subsequently implement TSES changes at the state and local levels intended to direct how administrators and teachers improve instructional practices. This article demonstrates how scholars can apply historical state political cultures, workplace behaviorism, and policy rationalism as a theory of action to investigate the entangled role these organizational influences play on TSES development and implementation. Ultimately, we argue that contemporary state-level conceptions of TSES models and policies under ESSA perpetuate inconsistent, ineffective, and confusing language and guidance for US educators that do not support existing empirical research on adult learning theories and professional growth practices.

The purpose of this research is to address three primary goals. First, it is important to analyze how TSES models have evolved under ESSA and conflated the related, yet separate purposes of supervision (formative instructional feedback) with the human resource processes of evaluation (summative evaluation determining ongoing employment). Second, by analyzing 30 states' contextualized political cultures during ESSA policy development and implementation, this article provides a contemporary analysis of how decision-making structures and the ideology of policy rationalism have influenced state-level TSES development throughout the US. Third, our analysis provides insight into the messages conveyed about improvement efforts through workplace behaviorism (Hazi, 2019), which in turn allows ESSA to continue to serve as the federal-level invisible hand that guides state-level TSES development in place of No Child Left Behind (NCLB). This paper addresses three research questions, specifically, 1) How are formative and summative feedback conflated in TSES models across the 30 purposefully sampled states?; 2) How do the 30 purposefully sampled states TSES inform a broad understanding of teacher evaluation policy across the US, particularly when considering the role of political culture on policy development?; and, 3) What are the emerging analyses about the application of workplace behaviorism in the 30 TSES models sampled, and what influence does this have on TSES model construction?

What is missing from the literatures on political culture, teacher supervision, and teacher evaluation, respectively, is empirical evidence that can determine the extent to which these ideologies, values, and beliefs potentially influence TSES development. To understand how state-level political culture influences TSES models which are embedded within these realities of localized politics, we developed an integrated conceptual framework that allows researchers, policymakers, and practitioners to analyze TSES development in a post-NCLB era. First, we offer an overview of how political culture, decision-making structures, policy rationalism, and workplace behaviorism potentially influence TSES policy development, as well as how these concepts affect local practice. Then, we provide a review of conceptual and empirical literature, specific to each part of our framework, and subsequently applied to our methods and analysis.

Background and Conceptual Framework

The passing of ESSA signaled a potential shift of the teacher evaluation landscape throughout the US, providing states with the authority to develop TSES models and define parameters for teacher effectiveness (ESSA, 2015). Now that state legislatures and policymakers are directing teacher supervision and evaluation policy and legislation, it is important to frame contemporary education policy analysis of supervision and evaluation through a state-level cultural and contextual lens. Previous research in political science, education policy and analysis, and educational leadership, demonstrates that state-level political culture and history has an influence on policy development, governance, and legislation (Elazar, 1994; Fowler, 2013; Louis, et al., 2008). However, no studies have examined the role of state political culture in the development or implementation of TSES model development. We argue that scholars in the field of teacher supervision and evaluation need to account for the influence of state-level and local-level political culture in order to determine how TSES models evolve, oftentimes contrary to established best practices and empirical findings, contributing to ongoing dissonance among teachers and administrators who implement systems of practice.

Despite a newfound level of control over TSES processes at the state level, scholars, practitioners, and policymakers have not developed consensus regarding how to effectively develop and implement models to improve teaching performance (Glanz & Hazi, 2019). Suggestions of how to improve teacher evaluation are in no small demand, such as the use of formative feedback not tied to summative evaluation, the use of self-reflection and self-evaluation, alternative measures (i.e. student surveys, parent surveys, etc.) of teacher feedback, and a clear bright line delineating student achievement from teacher evaluation (Mathis & Trujillo, 2016). At the same time, principals have expressed concerns regarding the implementation of state-level systems and the challenges of providing appropriate support for both teachers and principals to focus on growth mindsets (Derrington & Campbell, 2015). State-level systems model a consistent, technical approach to provide guidance for teachers and administrators in the processes of supervision and evaluation; however, local implementation is complex and requires more responsiveness to contexts and professionals' development.

The basis for administrators' concerns are founded in the respective, but often conflated and inconsistent, purposes and methods of teacher supervision and evaluation (Zepeda & Jimenez, 2019). For example, teacher evaluation can be useful for removing underperforming teachers (Grissom & Bartanen, 2018), however a much larger majority of teachers need a system that provides formative feedback which can be used to improve instructional practices (Mette et al., 2015; Stark et al., 2017). Even though ESSA gave states increased control over the development of TSES, many state models continue to implement teacher evaluation policies based on NCLB-era ideology using high-stakes test scores and performance ratings, which are approaches rooted in theories of workplace behaviorism (Hazi, 2019). Among scholars of teacher supervision and evaluation, what we do not know or understand is the role political culture plays in shaping states' agendas to develop and implement particular models of TSES. These state-level policy decisions and choices potentially influence local school administrators' and teachers' professional efficacy to improve instruction, and require further study. It is important to analyze how policy, political culture, decision-making structures, and policy rationalism have influenced how TSES evolved across states under ESSA.

Political Culture as a Theoretical Influence on TSES

In this study we examined how political culture potentially influenced TSES development and implementation during the beginning of the ESSA era. According to Elazar (1994) and Fowler (2013), historical political culture influences legislative decision-making structures (i.e. the level of local control) and policy rationalism (i.e. the amount of coordinated activity to solve social issues). Even though legislative priorities can ebb and flow over time in the context of state elections and evolving political rhetoric, the framework of political culture can be applied to investigate the enduring political perspectives based on geographic regions and historical cultural migration patterns. In turn, this can provide a useful tool to analyze the decisions of education policymakers within and across various states.

Elazar (1994) and Fowler (2013) provide detailed descriptions about the use of historical political culture as a theoretical framework that can be used to examine how and in what ways the behaviors, values, and beliefs of policymakers influence subnational policy decisions. Historic political culture is multidimensional and complex, specifically as it relates to history, migratory patterns of European immigrants in early America, historical political relationships, political attitudes, and ideologies, and is often associated with geographical boundaries (Louis et al., 2008). Defined as a "term to make more explicit and systematic much of the understanding associated with such long-standing concepts as political ideology, national ethos and spirit, national political psychology, and the fundamental values of a people," (Pye, 2015, p. 8) political culture has played an interesting role in educational policy development over the course of the last half of the 20th century and the first part of the 21st century. Starting with the National Defense Education Act of 1958, the US federal government became increasingly interested in mandating what should be taught in elementary and secondary schools across the country (Urban, 2010). Prior to this time, state political cultures greatly influenced the variation in policy that was passed at the state level and how this influenced practices at the local level. The influence of state political culture on policy development waxed and waned throughout the second half of the 20th century, but again began to take more of a back seat when *A Nation at Risk* was published. However, even with the passage and implementation of No Child Left Behind (NCLB, 2002), which arguably brought about some of the greatest federal control over educational practices, in theory state political culture research and analysis still offers important insights into the values, beliefs, and assumptions of stakeholders at the state level (Welch, 1993).

Policy development and initiatives vary from state to state, and are dependent on events throughout history, embedded ideologies within state-level politics, and cultural traditions that become expressed through policy development and legislation, all of which Elazar (1994) and Fowler (2013) conceptualize into three categories of either *moralistic*, *individualistic*, or *traditional*. These three categories serve as the framework of political culture in our analysis. Elazar's typology is a common and well-established framework that examines how dominant political cultures have created pervasive political patterns about the role of government and the impact it has on politics, the economy, and government spending (Fisher, 2016; Heck et al., 2014). Elazar's framework allows researchers "to examine the deeper influences shaping the character of political life in various parts of the United States" (Dincer & Johnston, 2016, p.

132). Fowler expands on Elazar's typology by considering societal values, distribution of power and resources, and the ever-changing political landscape in American politics (Heck et al., 2014).

Using Elazar's and Fowler's definitions, moralistic political cultures favor government bureaucracies that actively solve problems and protect the common good through enacting new laws, programs, or mandates. Individualistic political cultures value government activity as a social tool to grow and secure prosperity and efficiency. In individualistic political cultures the government is a system that functions to enhance economic prosperity and efficiency under certain limitations to prevent overreaching interventions, particularly in personal affairs. Finally, traditional political cultures support government activity, which maintains the existing structures, functions, and hierarchies of the social and political system. There is more value placed on deliberation and careful planning, to ensure order, while protecting a free market business model to respond to societal needs.

For example, Marshall et al. (1989) conducted a six state comparative study, which compared two states from each political culture category regarding the enactment of local and national education policies. Among other findings, their analysis revealed that absent a national-level policy agenda or initiative, states from each category diverged in terms of education policy activity and intentions. In a series of related publications, (Louis et al., 2006; Louis et al., 2008; Febey & Louis, 2008) researchers analyzed ten states' education policy activity and press and compared how both state- and local-level actors interpreted policy actions. Collectively, the authors found that the conception of political culture influenced how state and local leaders made sense of policies such as accountability and mandated testing differently across historical political culture categories. In Elazar's model, there are additional geographic considerations, where empirical research has substantiated how particular political cultures historically represent regions of the United States.

In studies completed outside the field of education, Mondak and Canache (2014) expanded upon Elazar's original framework and found subnational variation on the electorate's political culture, ideology, and characteristics across states. Other researchers have applied the theory of political culture to make predictions about policy development for social issues such as welfare reform (Mead, 2004), voting requirements (Hale & McNeal, 2010), and public policy discourse (Wilder, 2017). However, there is a gap in the literature as it relates to empirical evidence detailing how political culture influences educational policy development, specifically TSES development and implementation. Within the field of education, only one relatively recent research study has touched upon the potential broad, contextual influences of political culture on education policies and school improvement compared across states and regions of the US (Louis et al., 2010). As such, it is important to identify how, or if, political culture influences TSES development during the ESSA era, establishing a new line of inquiry to determine if and how ESSA goals have influenced TSES policy development at the state and local levels.

Based on this framework, we could realistically expect TSES models to look different across historical political culture categories. For example, it is reasonable to assume moralistic states would be likely to favor TSES models that address pedagogically sound instructional practices and culturally relevant practices based on political culture descriptors. Additionally, individualistic states might promote TSES models that focus on education systems in order to

develop students to be competitive in the global economy, but not necessarily through state mandated curricula requirements. It could also be expected that traditional states support the development of TSES models that are centralized, structured, and maintain order in a hierarchical political system.

History of Teacher Supervision and Evaluation Decision-Making Structures

Teacher quality matters, specifically the quality of teachers' instruction, and principals have a responsibility to provide supervision and evaluation of instruction (Rigby et al., 2017). Often, principals are not able to differentiate their own professional beliefs about supervision and evaluation due to state-level mandates that are determined by policymakers outside the local level. This policy-practice gap highlights the importance of understanding not only the implications for supervision and evaluation in an ESSA era, but also the policies and practices throughout the US.

A matter of debate, often publicly and through policy expansion, is the role of federal and state-level accountability policy about and around TSES development (Derrington & Campbell, 2018; Close, Amrein-Beardsley, & Holloway, 2017). Teacher supervision and evaluation has increasingly become a high stakes tool of the ongoing education accountability era, with principals becoming concerned about the role of supervision feedback to improve instruction, as well as the use of evaluation metrics to calculate teachers' effectiveness (Derrington & Campbell, 2018). However, recent scholarship has not fully addressed the distinct purposes of supervision and evaluation and how policies have contributed to the conflation of these two administrative processes. Scholars have established that teacher supervision should be a bottom-up, teacher-driven process based on individualized growth goals and used for professional development and formative feedback purposes (Glickman et al., 2018; Zepeda, 2017). Additionally, supervision scholars have made it clear through literature a clear bright line exists between the intentions of supervision and evaluation (Mette & Riegel, 2018; Glanz, 2018; Gordon, 2019; Zepeda & Ponticell, 2019). On the other side is the argument that teacher evaluation, based on summative feedback, should be a top-down, administrator-driven process based on student outcomes and used for ongoing employment purposes (Grissom & Bartanen, 2018; Hallinger et al., 2014; Robinson et al., 2008). While both sides have staunch philosophical supporters, at the heart of this debate is seemingly the push for school organizations to deliberately develop instructional improvement processes, which include various forms of feedback, direct coaching about performance, and the need to engage and collaborate in ongoing personal reflection (Kegan & Lahey, 2016; Knight, 2016). Although researchers have studied the resultant practitioner and school-level effects of supervisory and evaluative conflation, very little, if any research has attempted to determine a root cause of this issue as defined by policy culture, language, or intentions. We assert that the field has clearly defined the distinct purposes of supervision and evaluation and that practitioners have expressed professional uncertainty about how to separate them in the context of their practices, but we lack a holistic political analysis to understand how policymakers position these two critical functions of instructional improvement.

Despite the ongoing research and debates that surround TSES theory and practice, few studies or conceptual scholarship have discussed or examined the decision-making structures across

multiple states, specifically how different states develop, determine, and implement TSES. Acknowledging this empirical and conceptual limitation, the research field of teacher supervision and evaluation could benefit from developing a framework that conceptualizes and contextualizes the various influences on TSES development, particularly how NCLB, Race to the Top (RTTT), and the ESSA have acted as policy levers to influence professional policies and practices. Moreover, considering the decision-making structures that exist within states, often referred to as *centralized* or *decentralized*, may help to identify a state's level of control in determining a TSES model (Hazi & Arredondo Rucinski, 2009; Ruff, 2019).

Policy Technorationalism Influencing TSES Development

In addition to understanding decision-making structures across various state political cultures, it is also important to consider how policy rationalism, or the amount of coordinated activity to solve social issues, exists across various state systems. Louis et al. (2008) posit the use of *technorationalism*, a high level of coordinated activities by state policymakers, is often used in the attempt to comprehensively address school reform. Technorationalism applies the ideology that social issues and problems can be controlled precisely and orderly, often through policy efforts and mandates that relate to objective assessment of social outcomes, specifically through quantified data tied to accountability (Cheng, 2015; Orr, 2007). Understanding how TSES develop across states and political cultures is important, particularly in order to understand the role technorationalism plays in the often subjective art and science of applying instructional leadership in practice.

The influence of coordinated activities by policymakers to address TSES development is highlighted in the adoption of standardized TSES programs, such as those developed by Marzano, Danielson, and others (Mette & Riegel, 2018). These prepackaged TSES programs often merge the functions of supervision (formative feedback) and evaluation (summative feedback) into one outcome in an attempt to systematically produce growth in student achievement (Salo et al., 2015). However, these attempts to implement a technorational approach to TSES development have created conflation between the two functions and ultimately threaten supervision in favor of systematic teacher evaluation in the name of precise and objective assessment of teacher performance (Mette et al., 2017). Perhaps of greater concern, as prepackaged TSES systems continue to be favored and adopted by state policymakers, is the de-professionalization of instructional leadership that occurs (Hazi, 2019) as both principals and teachers lose agency to address instructional practices and innovation at the local level.

Workplace behaviorism. Embedded within technorational approaches within the field of teacher evaluation is an underlying and faulty assumption that negative feedback for adult learners will lead to increased motivation to improve practice (Drago-Severson, 2009). In teacher evaluation systems, this assumption is applied in practice that requires evaluators (usually principals) to identify weaknesses that will lead to instructional improvement, which in turn will increase student achievement scores (Ingersoll & Collins, 2019). The continued existence of popular evaluation systems (e.g. Danielson, Marzano, etc.) reinforces workplace behaviorism practices that attempt to place teachers on a continuum and standardize the evaluation process where only a very small percentage of teachers ever meet 'exemplar' ratings. What results is a system that demotivates and demoralizes teachers to improve their practice and trickles down,

failing to meet the learning needs of students because of the hyper-focus on test scores (Santoro, 2018).

Missing from policy is language that empowers teachers to take part in reflective practices and improvement efforts, and instead promotes hierarchy that suggests only principals or other evaluators are capable of giving feedback that leads to change. Analysis of language in TSES models is critical as they help educators deconstruct the misuse of evaluation instruments that are value-laden and only allow for a narrow interpretation of quality teacher (Hazi, 2010; Hazi, 2019). As Danielson (2011) admits, there is little to no agreement throughout the profession of what amounts to 'good teaching.' Herein lies the problem – that without a singular definition of what makes for good teaching (Glickman, 2015; Marshall, 2013), policymakers and practitioners alike cannot exercise rational control over the profession and the tight control of variables (Archer et al., 2016) will not improve outcomes of student achievement on standardized tests (Hazi, 2019).

In large part because of the accountability movement, many administrators end up spending more time focusing on technical aspects of school improvement rather than focusing on cultural and human development efforts (Waite & Nelson, 2005). Due to this reality, relational aspects of instructional improvement efforts are often ignored in place of technical efforts that attempt to drive immediate and measurable changes in student test scores (Nelson et al., 2008). Ylimaki (2012) posits the use of workplace behaviorism is ubiquitous to the point that even principals with the best intentions often resort to using data, efficiency, and productivity to evaluate teachers due to accountability pressures.

Depending on the extent to which policies are adopted, the favoring of technorational approaches to TSES development and implementation increasingly creates messages about 'the right way to teach' and 'the right way to evaluate teachers.' Examining how policies are developed, the influences that exist within the policies, and perhaps of most importance, the level at which decisions are made (e.g. the state or local level) is crucial to understanding the extent to which teachers are provided feedback about their instruction in the current ESSA era.

Methods

Using qualitative methods, we completed a three-stage document analysis of 30 states' TSES policy documents and ESSA documents from state departments of education (SDOE) to inductively describe, analyze, and interpret (Wolcott, 2009) publicly available archives related to state-level education policies. We purposely chose 30 states to analyze statutes and policies based on each state's political historical culture (Elazar, 1994; Fowler, 2013) to get a balanced sample of *moralistic*, *individualistic*, and *traditional* cultures, as these labels have been used to describe the historical political cultures of states and are often based on historical influences. Accounting for geographical diversity and population density, we chose the 30 states identified in Figure 1.

Figure 1: Historical political culture

Data Sources and Collection

We collected publicly available documents and separately previewed and read the context and content of policies within each state. In some cases, we also emailed or called state departments of education to clarify or to obtain information from websites that were not active. Data were collected from February 2019 through October 2019. After collecting states' documents, we utilized a shared cloud storage system to share and exchange policy documents and data. As a means to orient ourselves to the data, we met several times to discuss and reflect upon our initial review of the state documents to determine if the content provided sufficient data for analysis and also to determine initial reading and coding assignments divided equally among the authors. We collected and archived over 250 publicly available documents, all coming from state department of education (DOE) or state legislature websites. These documents included legislative documents, DOE regulations, memos on TSES implementation, archived state-level presentations, and state legislature technical reports studying TSES implementation.

Data Analysis

Using our initial data review and discussions as a first level of analysis, we deductively narrowed our analysis to analyze state-level policies, documents, and technical reports specifically related to two primary categories of our analytical framework: 1) teacher supervision as formative professional development, and 2) teacher evaluation as summative assessment of performance. We met via videoconference 14 times for a period of 1-2 hours at a time as a research team to comparatively code and reduce our analysis into these two categorical components to identify

language descriptors unique to each category (Creswell, 2013), as well deepen our policy analysis approach by creating initial lists of relevant codes (Miles, Huberman, & Saldaña, 2014). From the collected documents, we identified specific information about the TSES sources and feedback components, which we organized according to our two primary categories, which became the initial set of qualitative codes. For example, regarding teacher supervision we identified and coded policy and practice terms which were descriptive of legal code requirements, guidance, processes, purposes, or the materials of practice related to formative professional support to improve teachers' instruction. We included terms such as instructional supervision, teacher reflection, self-evaluation, coaching, professional development, portfolio development, peer to peer conferencing, teacher growth, teacher leadership, mentoring, and teacher improvement, of which were not described as related to summative evaluation or judgment of performance. Regarding teacher evaluation, we identified and coded policy and practice terms which were related to professional summative evaluation of teachers' instruction and judgment of their professional performance. We included terms such as instructional evaluation, summative evaluation, teacher ratings, evaluation labels or categories, and applications of rubrics for the purposes of providing categorical judgments of teachers' performance. Our coding discussions were complex because we had to account for terminology and concepts which were embedded in policy rationales and practices, many of which may not have been aligned with discussions reflected across the scholarly literature. This is where our secondary level of analysis proved critical our collaborative understandings of the data.

A secondary level of analysis of our two primary categories required reconciliation of terms that could be included in one or both categories, either supervision or evaluation, and was the focus of much of our discussions. These coded terms included instructional frameworks or models, student outcomes, student learning objectives, student learning goals, instructional standards, learning standards, performance standards, and other variations of terms related to student learning, student performance, and student achievement. For example, when we identified a state that used instructional frameworks, we needed to determine the contextual application of such models. These could be used at the state and local level to guide instructional standards and priorities, but was part of the supervision process. Or, some states converted the frameworks to instructional performance scoring rubrics that determined teachers' effectiveness as part of an evaluation. These terms described situations which provided differentiated application at the state and local levels, and formed a significant part of our comparative analysis of the data, presenting the conflation of supervision and evaluation practices. In these cases, some terms would be applied to supervision, evaluation, or to both supervision and evaluation where appropriate.

Once the data were deductively identified from the larger and original database of documents, we re-categorized the data into various table categories to assess the state-level policy models and language in an attempt to capture the gestalt of each state's dynamics and trends across the whole data set. Through each cycle of interpretation and re-interpretation, we also employed a cross-case check of each other's analysis to ensure interpretive consistency and trustworthiness (Saldaña, 2013). These cycles eventually helped us to refine and consistently interpret and analyze the data and reduce our data to workable categories of policy language aligned to our analytical framework.

Data Interpretation

At the final stage of data interpretation, we utilized our cross-case state analysis and verification procedures to apply our analytical framework of political culture, decision-making, and policy rationalism, to investigate the impacts on TSES development and implementation. As patterns and working hypotheses emerged from the analysis process of the collected artifacts, we articulated our interpretations into various categorical descriptors based on our framework. We applied the nature and intention of the coded categories and then cross-referenced the content within each state and each political culture category to determine a holistic analysis across all 30 states. At this point, we were able to observe patterns emerge among the states as a holistic sample, and then more acutely within political culture categories and geographic areas. We elaborate on a description of our evidence attributed to each identified theoretical category in the findings section below.

Subjectivity Statement

Here we briefly discuss our scholarly and experiential biases that are inherent influences on our analysis. Our conceptual framework and comparative methods helped us formulate a working heuristic that acted as an analytical bracket to clarify and identify our biases, which we checked throughout the analytical process. We each believe in the potential for teachers and administrators to drive innovation and school improvement at the local level, and are critical of reform efforts that mandate particular models, frameworks, and ratings systems to determine teacher effectiveness.

Albeit a lofty and improbable wish, we did think that states would make a policy break with the top-down federal accountability systems that dominated education since the inception of NCLB, especially within states that historically valued local control and limited government intervention. Although NCLB was inspired by education policy accountability systems developed in states such as Texas, under NCLB the systemic federal penalties applied to school districts for failure to perform at expected levels of achievement demonstrated an unprecedented level of instructional intervention and usurping of local control. Given existing evidence of how NCLB damaged educators' professional morale and disproportionately affected urban areas and communities of color, we saw ESSA as an opportunity for states and districts to construct means to serve their communities in contextually relevant and responsive ways. As such, we anticipated that NCLB style policies were not normalized and would not persist beyond NCLB's existence within various states' political cultures, whereas in some states, these top-down accountability policies were already inherent within existing political cultures and traditions and may continue in some form beyond NCLB and into ESSA.

While constructing their respective ESSA plans, we anticipated that states would rely on existing political cultures, traditions, values, and processes that may have preexisted NCLB to develop teacher supervision and evaluation systems at the state and district level. Through ESSA, our hope was that states would leverage their newfound ability to guide education policy and return the reins to local education agencies to determine how best to serve their respective school communities. In our minds, this meant that school districts would be provided the freedom to develop separate, yet related, formative supervision plans which focus on professional

development and growth, coupled with summative evaluation systems that determine teachers' professional and instructional effectiveness. These mutually developmental systems could be created locally and reflect communities' contextually relevant values and visions of teaching and learning, which we assume would simultaneously reflect elements of a state's political culture. Important for a qualitative study, our beliefs and scholarly knowledge focused this main point as a working heuristic and a starting point of comparison, debate, and reflection.

Findings

In this study we found four main findings. First, historical political culture does seem to play a role in TSES model development from state to state. Broadly speaking, states that have been labeled as individualistic or moralistic are more likely to use TSES models to grow and develop teachers using formative feedback structures and language that empowers LEAs rather than use statewide government activity to drive policy decisions. That said, an increasing number of moralistic and individualist states appear to be transitioning toward a more traditional political culture with state-wide policies that dictate TSES implementation regarding widespread use of student achievement data and growth models tied to teacher evaluation metrics.

Second, most states appear to conflate supervision (formative feedback) with evaluation (summative feedback) within their TSES models. There are exceptions to this, however, as five states do not have formative language in their TSES policy documents. Additionally, most states require self-reflection to be used as a data point in summative evaluations, which is an interesting concept for an evaluator to assess considering the internal nature of self-reflection. That said, there are five states (all individualistic or moralistic states) that do not require this to be part of the summative process.

Third, workplace behaviorism is used to shape and reinforce actions throughout the 30 states sampled, namely in the use of prepackaged teacher evaluation systems and the use of language to label teacher performance. Most states (20 out of 30) use some form of prepackaged TSES model, however of the 10 states that allow local educational agencies (LEA) the option to create their own model given a state framework, nine of those are moralistic/individualistic states. When examining the language used to label teacher performance, an overwhelming majority (63.3%) used Ineffective or Unsatisfactory to label low teacher performance and 73.3% used the labels Highly Effective, Distinguished, or Exemplary to identify high teacher performance. These patterns point to remnants of NCLB era policies that attempt to reward and punish teachers based on labeling performance outcomes, a framework that was supposed to have changed with ESSA. Additionally, only six states of the 30 states sampled do not use TSES ratings to remove teachers, and of those six, five are in individualistic or moralistic states.

Fourth, there is a clear pattern of technorational and centralized approaches to TSES development across the 30 states sampled. Twenty-one states require TSES models tie student outcomes to teacher evaluations, and of the nine that suggest (but do not require) or are neutral towards the use of student outcomes, eight are individualistic or moralistic states. When looking at the use of growth models with TSES development, 23 require the use of growth models (VAM, etc.), growth percentiles, or a combination of other metrics (including local decisions). Seven of the 30 states sampled use student learning objectives as a more moderate method of assessing student learning over time, six of which are individualistic or moralistic states.

Figure 2: Use of formative feedback as part of summative evaluation within TSES model
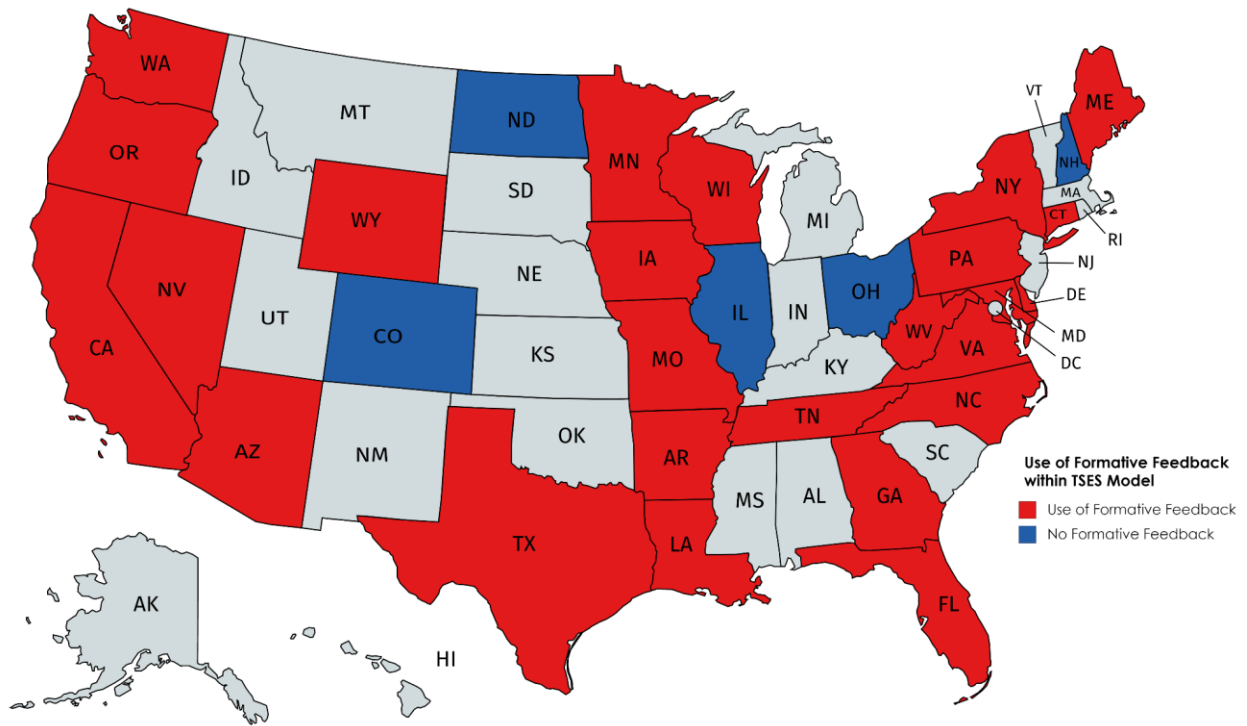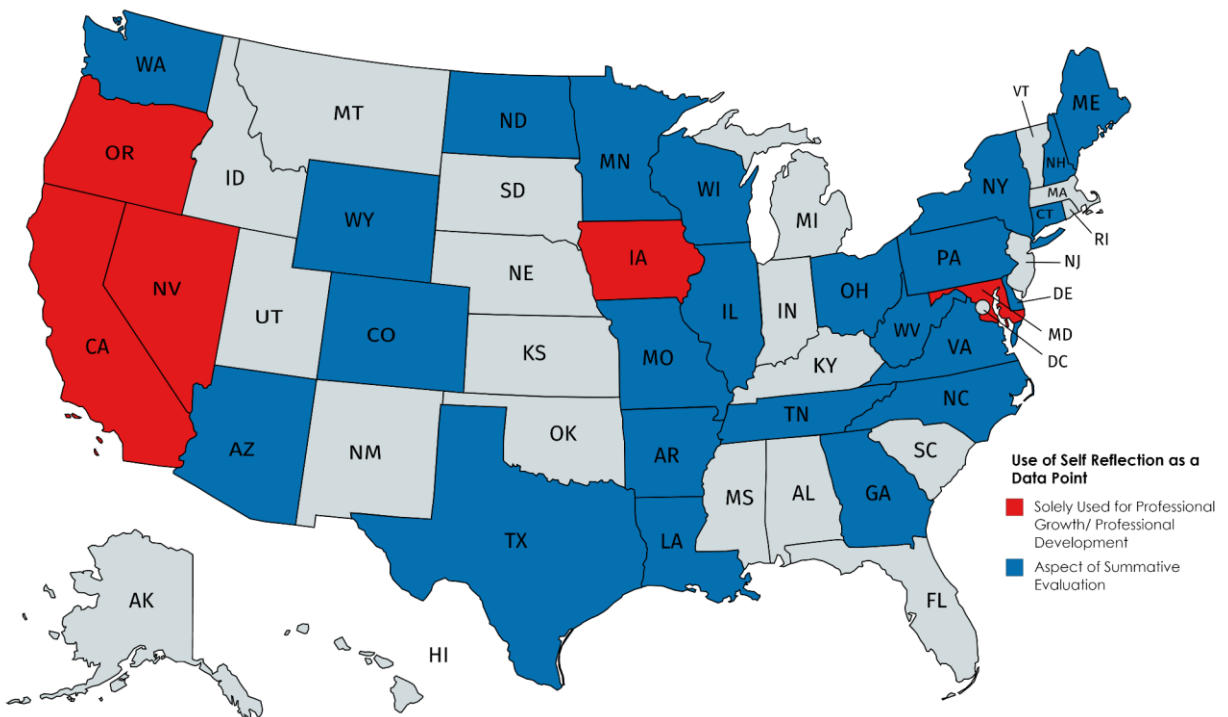

Figure 3: Use of self-reflection as a data point

Conflation of Formative with Summative Feedback

As discussed in the review of the literature, there are conflicting ideologies about the use of formative feedback being intertwined with summative feedback. At the heart of the debate is how and in what ways policy development is used to leverage changes to produce summative outcomes, which inherently are used to influence and inform formative feedback about instructional practices. Through this analysis, it is clear that throughout the 30 states sampled there is a conflation between the use of formative feedback (often defined as supervision) and summative feedback (what most refer to as evaluation) within TSES models. Of the 30 states sampled, 25 states (83.3%) embed formative feedback within the summative portion of the TSES models (see Figure 2). Specifically, these five states provide no language in their TSES models about the use of formative feedback being tied to summative evaluation. This provides a clear pattern of policy development across different regions of the country and across various historical political culture backgrounds. However, of the five states that do not conflate formative with summative evaluation (Colorado, Illinois, New Hampshire, North Dakota, and Ohio), all are either individualistic or moralistic.

Analyzing the variable of self-reflection as a data point, again, 83.3% of states (25 out of 30) require the use of self-reflection in summative evaluation. Five states (California, Iowa, Maryland, Nevada, and Oregon) only use self-reflection to either promote professional growth or professional development and do not use this as an aspect in the summative evaluation in their respective TSES models (see Figure 3). Interestingly, the five states that do not require self-reflection in the summative evaluation all come from individualistic or moralistic states, and none come from traditional states (see Table 1). Florida is the only state sampled that varies by county regarding the use of self-reflection to be assessed in summative evaluation. Again, the use of formative language, while included in TSES policy, often appeared to be conflated and used as part of summative evaluation tools and measurements.

Table 1: Percentage of moralistic/individualistic states using self-reflection in TSES

Self-Reflection Embedded in TSES Models	States	States Moralistic/ Individualistic	Percentage (out of 20 Moralistic/ Individualistic states)
Solely Professional Growth/Professional Development	CA, IA, MD, NV, OR	CA, IA, MD, NV, OR	25%
Aspect of Summative Evaluation	AR, AZ, CO, CT, DE, GA, IL, LA, ME, MN, MO, NC, ND, NH, NY, OH, PA, TN, TX, VA, WA, WI, WV, WY	CO, CT, DE, IL, ME, MN, MO, ND, NH, NY, OH, PA, WA, WI, WY	75%
Varies	FL		

Workplace Behaviorism and Use of Language

Increasingly, state TSES models appear to require local school districts to use state approved models. These prepackaged TSES models influence how LEAs are allowed to make decisions about teacher supervision and evaluation at the local level, which in turn creates a behavioristic view about the types of feedback teachers should be given in order to be considered 'acceptable.' As such, many TSES models serve as gateways for prepackaged evaluation systems determined at the state level. Figure 4 shows 20 out of 30 (66.7%) states sampled use some form of prepackaged TSES model. Of the remaining states, 10 allow for LEAs to create their own model given these adhere to a state framework. Nine of the 10 that allow LEAs the option to create their own TSES models are either individualistic or moralistic, and the tenth state (Texas), while it allows LEAs the option to create their own system, also does not provide legal support to school districts being sued by a teacher unless the district uses the Danielson model. Thus, while LEAs in Texas have the opportunity to develop their own model, the lack of legal support creates an incentive among practitioners to use the Danielson model. Table 2 provides more detailed information about prepackaged TSES models.

Figure 4: Requirements of TSES model development

Another factor contributing to workplace behaviorism in the 30 states sampled is the use of language to define TSES models. When looking at teacher performance, an overwhelming majority of states sampled (19 of 30; 63.3%) used the labels of 'Ineffective' or 'Unsatisfactory' to label low teacher performance. Two states (Pennsylvania and Virginia) used pejorative labels such as 'Unacceptable' and 'Failing,' however three states used no labels and instead simply delineated performance on a 1-4 scale (see Table 3). These three states that refused to label low performing teachers were either moralistic or individualistic states.

Table 2: Percentage of moralistic/individualistic state TSES requirements

Requirements Placed on TSES Development	States	States Moralistic/ Individualistic	Percentage (out of 20 Moralistic/ Individualistic states)
Predominantly Danielson & Marzano	AR, AZ, DE, FL, LA, ME, MN, NY, PA, TN, WA, WI, WV, WY	DE, ME, MN, NY, PA, WA, WI, WY	40%
Other models	CA, GA, MO, NC, NV, VA	CA, MO, NV	15%
LEA option using state framework	CT, CO, IA, IL, MD, ND, NH, OH, OR, TX,	CT, CO, IA, IL, MD, ND, NH, OH, OR	45%

Table 3: Labels used to identify low performing teachers

Descriptive Use of Labels to Identify Low Performing Teachers	States	Total	Percentage (out of 30 states analyzed)
Ineffective	AR, AZ, CO, DE, GA, LA, ME, MO, NH, NY, NV, OH, WY	13	43.3%
Unsatisfactory	FL, IL, MN, WA, WI, WV	6	20%
Other (Below Standard, Developing, Does Not Meet, Non-Proficient, Not Demonstrating, Rarely Effective, Significantly Below Expectations)	CT, NC, ND, OR, TN, TX	6	20%
Unacceptable, Failing	PA, VA	2	6.6%
No Labels Used	CA, IA, MD	3	10%

Additionally, 73.3% used the labels 'Highly Effective,' 'Distinguished,' or 'Exemplary' to identify high teacher performance. One state, Tennessee, used a value-laden label of 'Superior' to define high teacher performance. Again, the same three states refused to use a label to define high performing teachers and instead used a 1-4 scale. Table 4 provides more detailed information about the use of language and label to identify high performing teachers in the 30 TSES models analyzed.

Table 4: Labels used to identify high performing teachers

Descriptive Use of Labels to Identify High Performing Teachers	States	Total	Percentage (out of 30 states analyzed)
Highly Effective	AR, AZ CO, DE, FL, LA, ME, MO, NC, NH, NY, NV	12	40%
Exemplary	CT, GA, MN, ND, VA	5	16.6%
Distinguished	PA, TX, WA, WI, WV	5	16.6%
Other (Accomplished, Advanced Teaching Practices, Effective, Exceeds, Excellent, Highly Qualified, Innovating)	IL, OH, OR, WY	4	13.3%
Superior	TN	1	3.3%
No Labels Used	CA, IA, MD	3	10%

A third aspect of the use of workplace behaviorism in TSES models is the use of ratings to remove teachers. These patterns suggest a hangover from the NCLB era, where policies enacted were attempted to reward and punish teachers based on outcomes (see Figure 5). Five states can use TSES ratings to remove any teacher after two consecutive years (Colorado, Florida, Maine, North Dakota, and Wyoming), and five additional states allow the removal of teachers with low TSES ratings but do not have a specified timeline (Arizona, Georgia, Minnesota, Missouri, and Wisconsin). Only six states of the 30 sampled do not use TSES ratings to remove teachers, and five of these are individualistic or moralistic states (California, Connecticut, Maryland, New Hampshire, and Oregon).

Figure 5: Ability to use TSES rating to remove teacher
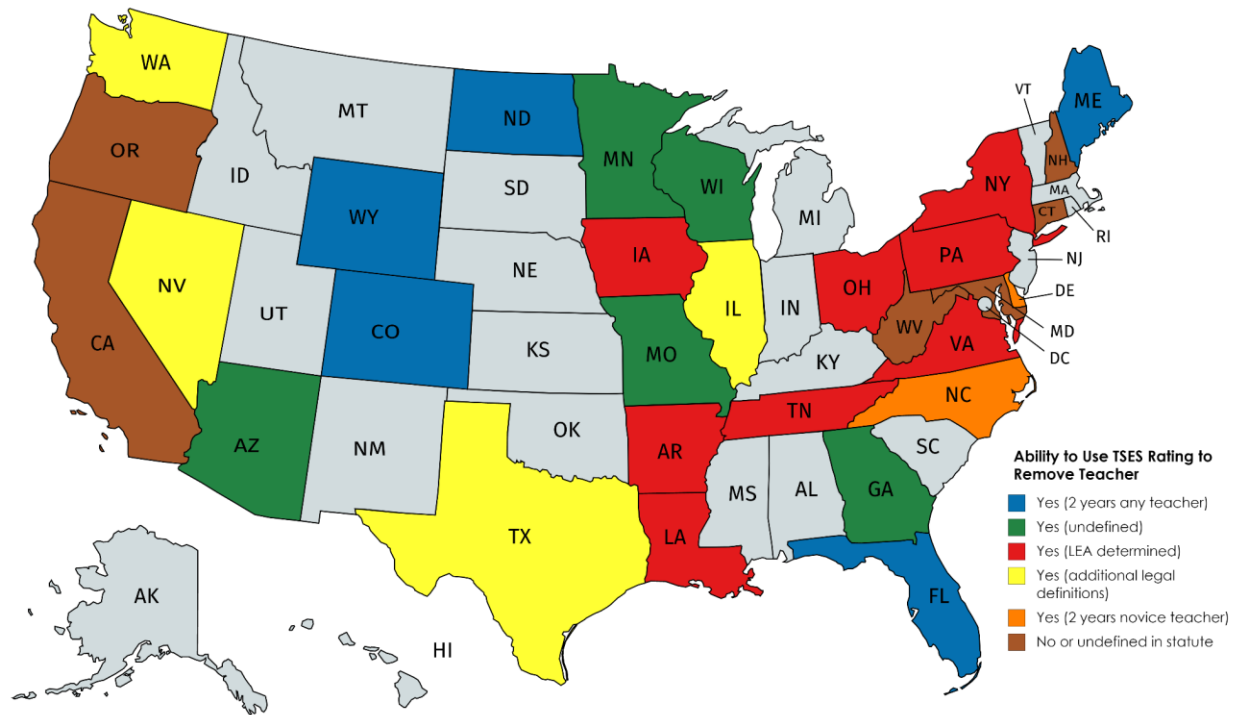
Technorational and Centralized Approach to TSES Development

The 30 state analysis conducted for this study points to highly coordinated activities by state mandated policies, which highlights a technorational and centralized approach to TSES development in the ESSA era. These approaches to state-level TSES development reveal the continuation of school reform efforts that attempt to precisely and methodically control teacher evaluation through the use of student assessment data and the attempt to produce quantifiable teacher performance data that is tied to accountability and human resource decisions. Thus, the use of student outcomes, the percentage of student outcomes tied to teacher ratings, and the use of growth models points to the continuation of NCLB era practices even in the new ESSA era. The only change is that these coordinated and technorational activities have shifted from being centralized at the federal level to being made at the state level.

Of the 30 states sampled, 21 require the use of student outcomes to be included in TSES model development. Six states suggest student outcomes should be used (Arkansas, California, Connecticut, Maine, Missouri, and North Dakota), while three states remain neutral about the use of student outcomes (Delaware, Iowa, and Oregon). Of the nine states that either suggest or are neutral about the use of student outcomes in TSES models, eight are either individualistic or moralistic states, and only one (Arkansas) is a traditional state. Figure 6 provides more details about the technorational and centralized approach to TSES model development across the 30 states analyzed. Figure 7 provides more evidence of technorationalism through requirements for TSES to use student outcomes as a percentage of a teacher rating score. Nine states (Arkansas, California, Connecticut, Iowa, Maryland, Maine, Missouri, North Dakota, and Oregon) either suggest or do not require the use of student outcomes tied to teacher evaluation – the remaining

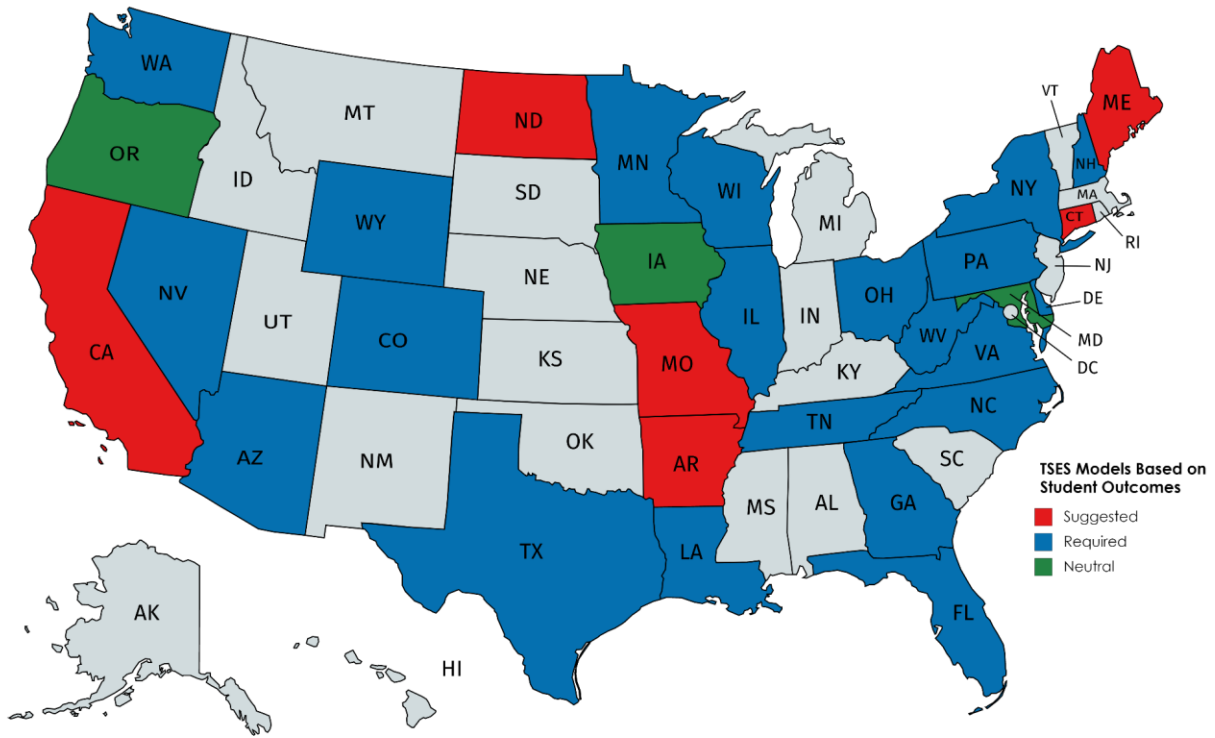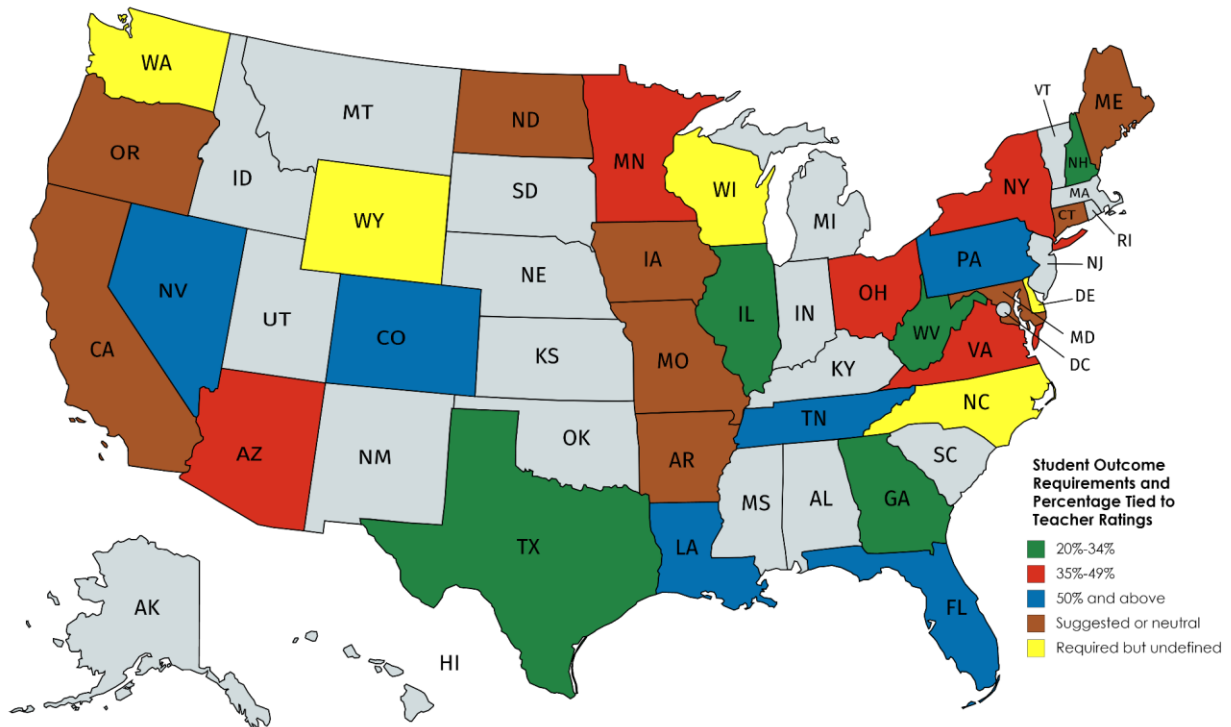*Figure 6: TSES models based on student outcomes*

Figure 7: Student outcome requirements and percentage tied to teacher ratings

21 states of the 30 analyzed do require student outcomes tied to teacher evaluation. Of the nine states that suggest or are neutral about tying student outcomes to teacher ratings, all but Arkansas are either individualistic or moralistic.

Figure 8: Use of growth models embedded in TSES
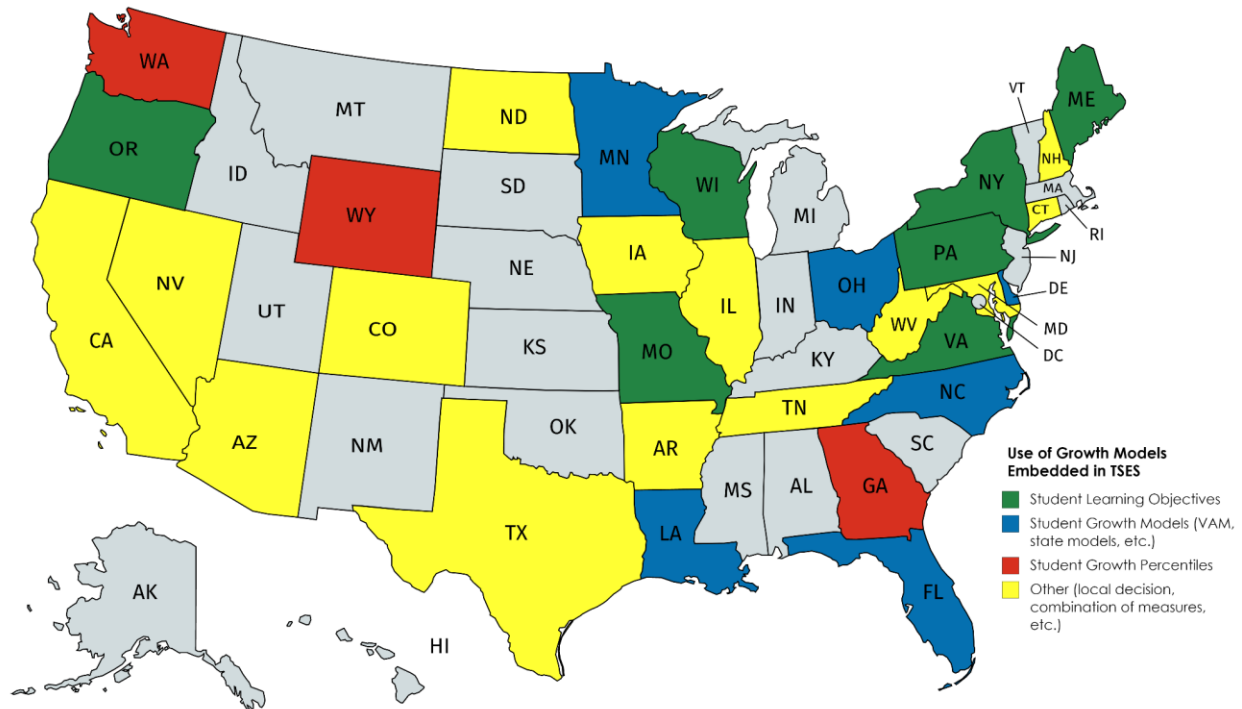
The required use of growth models also influences TSES development in the 30 states studied. While all states require some sort of student growth to be used as evidence in TSES model development, seven use student learning objectives rather than student growth models, student growth percentiles, or a combination of other measures (see Figure 8). Student learning objectives tend to focus more on the holistic development of students and allow for a student to meet the objectives with multiple interventions and additional support structures, whereas student growth measures and student growth percentiles are more of a standardized and mathematic evaluation process that do not account for various socioeconomic factors impacting student performance. Of the seven states that use student learning objectives, six (Maine, Missouri, New York, Pennsylvania, Oregon, and Wisconsin) are either individualistic or moralistic, while only one (Virginia) is traditional.

Discussion

The purpose of this research is to address three primary goals. First, it is important to analyze how TSES models have evolved under ESSA and conflate formative feedback with summative evaluation. Second, the analyses in this article provide insight into the messages conveyed about improvement efforts through workplace behaviorism (Hazi, 2019), which in turn allows ESSA to serve as an invisible hand that guides TSES development in place of NCLB. Third, by analyzing

30 states' individualized political cultures in the context of ESSA policy development and implementation, this article provides a contemporary analysis of how decision-making structures and the ideology of policy rationalism influence state-level TSES development throughout the US.

Based on our framework, we saw, to a certain extent, what we expected we might see in terms of differences between TSES models based on historical political culture. That said, the clustering and grouping of patterns was more dichotomous than we expected. State TSES models we found to be progressive were typically a mix of moralistic and individualistic states, and these were compared against states with more traditional political cultures. Moreover, we found some "creep" in traditional political culture to states that have historically been considered moralistic or individualistic. Specifically, we saw a larger percentage of all states using more centralized and technorational approaches in their TSES models than we expected to see. This finding reinforces the need for more research to identify how, or if, historical political culture influences TSES development during the ESSA era so that we can better understand how ESSA goals have influenced TSES policy development at the state and local levels. Our overall findings help to extend major research studies that already suggests the potential broad, contextual influences of political culture on education policies and school improvement compared across states and regions of the US (Louis et al., 2010).

For example, most of the 30 states in the sample appear to conflate supervision (formative feedback) with evaluation (summative feedback) within their TSES models. There are exceptions to this, however, as five states (all individualistic or moralistic states) do not have formative language in their TSES policy documents. Ironically, most states require self-reflection to be used as a data point in summative evaluations, which is an interesting concept for an evaluator to assess considering the internal nature of self-reflection. That said, there are five states (all individualistic or moralistic states) that do not require this to be part of the summative process. Whether by default or by design, TSES policy decisions at the state level appear to be increasingly centralized rather than decentralized. As ESSA policies move away from federal-level decisions about teacher evaluation there appears to be a power vacuum shift down to the state level. This prevents the ability for LEAs to make decisions at the local level, often ignoring local context or need.

Our research helps fill the gap in the literature that examines the role of state historical political culture in the development or implementation of TSES model development. Specifically, this research contributes to the literature that can help determine the extent to which ideologies, values, and beliefs potentially influence TSES development in the ESSA era. Additionally, the significance of this research rests in helping one better understand the effects on professional practice that historical political culture, decision-making structures, policy rationalism, and workplace behaviorism have at the local level. On the other hand, how individual school leaders have been interpreting evaluation policies and implementing them using their own discretion is also another facet of TSES development that is important to understand as part of what is also happening at the local level (Kraft & Gilmour, 2016; Reid, 2017).

At the local level, we have also seen the unexamined influence of historical political cultures, decision-making structures, policy rationalism, and workplace behaviorism manifest in de-

professionalization and loss of opportunity and agency for principals and teachers to address instructional leadership. For example, the "Widget Effect" or the act of treating evaluation like a rubber stamp with no significance for changing teacher performance (Weisberg et al., 2009) reminded scholars and practitioners that the role of teacher evaluation and capacity of leaders to carry out two distinct tasks (evaluation and supervision) is complex and counterproductive, especially in the ESSA era where teacher evaluations in their current form continue to resemble a rubber stamp with little effectiveness into the actual improvement of one's performance (Kraft & Gilmour, 2016; Reid, 2017).

While there continues to be an erasure of the formative feedback benefits of instructional supervision, the summative feedback leverage of workplace behaviorism of teacher evaluation becomes privileged. Others have addressed this 'wicked problem' (Mette et al., 2017; Hazi, 2019; Zepeda & Ponticell, 2019), but our study provides initial evidence of the ongoing decrease in instructional innovation and local-level capacity building due to the influence of prepackaged TSES systems (Marzano, Danielson, etc.). Similarly, Ylimaki (2012) argues that instructional leadership in a conservative era is challenging political work that requires a sensitivity to the relational aspect of the profession if change is to occur.

Furthermore, expecting principals and supervisors to use prepackaged TSES models only versus expecting them to account for the differences in teachers' needs and nuances in classrooms for the purpose of actual teaching improvement has historically served to disempower educators at the local level. This also further prevents the support of faculty through developmentally appropriate supervision and formative feedback, especially when prepackaged TSES models already conflate formative and summative feedback. According to Gordon (1992) we, as a field of scholars, have known this for some time now, but policy makers, scholars, and practitioners require a paradigm shift from the traditional supervision to the new supervision that entails empowerment of teachers for internal (self) accountability before external (evaluation) accountability. The perspective of 'the right way to teach' and 'the right way to evaluate' is a traditional ideology that has been thrust upon US principals and is problematic because of its technical challenges (i.e., time consuming) (Horng et al., 2010) and relational challenges (i.e., disregards pedagogical, personal, and social development) (Gordon, 1992). As Kraft and Gilmore (2016) write, "In practice, districts often hope to promote teacher development while also using evaluations for high-stakes accountability" (p. 711).

Another aspect that appears to further complicate TSES development is the use of workplace behaviorism. It is used throughout the 30 states sampled within the present study, namely in the use of prepackaged teacher evaluation systems and the use of language to label teacher performance. Thus, it is not surprising that TSES models in their current form within the 30 state sample are absent of clinical supervision processes. For example, most states (20 out of 30) use some form of prepackaged TSES model; however, of the 10 states that allow for local educational agencies (LEA) the option to create their own model given a state framework, nine of those are moralistic/individualistic states. When examining the language used to label teacher performance, an overwhelming majority (63.3%) used Ineffective or Unsatisfactory to label low teacher performance and 73.3% used the labels Highly Effective, Distinguished, or Exemplary to identify high teacher performance. To understand the meaning of these terms, it is important to understand that the language inherent in labels such as 'ineffective' and/or 'distinguished'

perpetuates a 'forced choice' and suggests that a desired teaching behavior is only achieved when something like a reward or condition is exchanged.

According to Foucault (1977), we live in world that is organized around 'regimes of Truth,' which control and manipulate individuals though discourses and norms. Thus, language has power and can subjugate a profession if policy is given too much power. Consider Foucault's panoptic metaphor, which showed that much like a panopticon, where a jailer stood and watched over prisoners to keep them aligned with a prison code of conduct, discourse also serves to align our identities, actions, and environments to prescribed norms. If we compare Foucault's conception of a panopticon, a prison structure used for surveillance and punishment of inmates, to that of a rubric with labels of 'effective/ineffective,' we can begin to see how coercion in the form of language and surveillance is maintained through a structure of labels that condition one or punish one into compliance.

In this study, the labels in many TSES models we examined in the 30-state sample do attempt to condition teachers to teach 'effectively' because all but six states use labels within TSES to remove a teacher. Without considering a teacher's thoughts, ability, needs and feelings in the process towards improving one's practice, teaching is essentialized and or treated entirely as a science, not an art. Additionally, only six of the 30 states sampled do not use TSES ratings to remove teachers, and of those six, five are in individualistic or moralistic states.

These patterns in use of language point to remnants of NCLB era policies that attempt to reward and punish teachers based on outcomes, a framework that was supposed to have changed with ESSA. It is important to note that while ESSA does require states to have a definition of what constitutes teacher ineffectiveness, states are not required to implement evaluation system. According to Pennington and Mead (2016), "culture eats policy for breakfast" (p. 27) or at least it did when the Obama administration rolled out RTTT. Because not all states participated in RTTT, nor were all interested in establishing a performance-based culture as a result of new policy, new teacher evaluation policies waned:

> [RTTT] reflected an assumption that policies could, by defining and measuring what good practice looked like, clarify to teachers what the district and state expected them to do in their classrooms every day and incentivize teachers to develop their own capacity to implement good practice. They also assumed that specific protocols for teacher evaluations would overcome a culture of reluctance to give hard feedback about performance—rather than that evaluators would find ways to bend protocols to avoid having hard conversations (p.27).

While changes in evaluation policy played a role in certain states, it alone did not transform TSES development as historical political cultures had already done.

Again, through our analysis we see the push to control education through ideologies found in more historical traditional political cultures, which perhaps should not be surprising in the current highly conservative US political environment. In addition, there is also a clear pattern of technorational and centralized approaches to TSES development across the 30 states sampled. Twenty-one states require TSES models tie student outcomes to teacher evaluations, and of the

nine that suggest (but do not require) or are neutral towards the use of student outcomes, eight are individualistic or moralistic states. When looking at the use of growth models with TSES development, 23 require the use of growth models (VAM, etc.), growth percentiles, or a combination of other metrics (including local decisions). Seven of the 30 states sampled use student learning objectives as a more moderate method of assessing student learning over time, six of which are individualistic or moralistic states. Thus, the mere presence of technorational ideology alone reinforces a 'one-size-fits-all' approach to TSES development, and again devalues the individualized supports needed for teachers to receive meaningful feedback and for principals to provide effective feedback to improve instructional practices.

Our analysis suggests current TSES models continue to reflect 'efficiency' and/or maintain NCLB ideologies, *particularly those aligned with historical traditional political cultures*, that then manifest in conflation of supervision to mean evaluation, as well as the movement to erase formative feedback. Perhaps of greatest concern is that many states have not taken advantage of the opportunity to rethink TSES given updated ESSA requirements, again, an outcome that can be tied to a spread in traditional ideologies and political culture. That said, there is a large amount of autonomy that is given back to the local level, which allows LEAs to control, if not be part of the conversation of, how instructional leaders develop teachers and are able to be developed to meet local needs. According to Lipsky's (2010) observation that policies are ultimately made by the "street-level bureaucrats" who implement them, local educators can push back against state-level interpretations of TSES models and processes across the US; however, the success of supervision as a whole depends on the capacity of local educators to implement reforms (Honig, 2006). Such was the case in the decentralized U.S. education system where local practice was often decoupled from central policy (Spillane & Kenney, 2012). Still, today, other variables at the local level also the extent to which educational leaders at the local who wish to practice supervision can exert control. Indirect forces and processes strip away autonomy from schools before educators can even challenge policy interpretations. According to Spillane et al. (2019), schools seek legitimacy and funding for survival, so school leaders are not in a position to challenge teacher evaluation policy interpretation. We recommend that district and school leaders work with teachers and teacher leaders to develop locally meaningful and developmentally focused teaching and learning improvement processes that meet the needs of their communities. Teacher evaluation policies provide standardized guidance, materials, and frameworks in some cases, but there is considerable latitude and control at the local level for leaders to operationalize relevant supports and developmental processes for their teachers.

Although our study does not address inherent causes of political shifts that occur at the state level, this study raises questions about the historical persistence and strength of state-level political cultures and their influence on future U.S. education policy. In some areas of the U.S. there are significant population demographic shifts that may potentially alter state-level political perceptions and legislative actions. We suspect over time that communities and states will develop and implement teaching and learning priorities that reflect local values and political ideologies, and not succumb to national standards or expectations which may not address unique needs and desires in a one size fits all manner. Even further, we would recommend that school leaders make more purposeful efforts to solicit feedback from their respective localized community stakeholders and craft teachers' performance expectations to reflect community

priorities, which may include aspects of service and outreach, community engagement, and other social practices that move beyond the classroom or school doors.

Standardized instructional and professional practice frameworks employed today do not include these types of community-oriented professional expectations or descriptions of teacher activity, thereby continuing to limit the potential impacts that teachers can make in their communities and in the field of education. These types of creative, and contextually relevant, solutions and professional practices could be transformative and position teachers as more integrated personnel and leaders in the community, moving beyond what we think of the prototypical, traditional teachers today. We hope that a grass-roots, bottom-up approach of activism at the community level could lead to greater levels of agency for parents, teachers, leaders, and school boards to engage with legislators and policymakers in collaborative ways to improve teaching and learning to best meet the needs of their children. To summarize, we continue to set our professional expectations too low, and limit the potential power and agency of teachers under the guise of efficiency, effectiveness, and political expedience.

Conclusion

While this study offers an analysis of only 3/5 of all state TSES in the US, it provides important preliminary data about various factors influencing TSES development at the state policy level. From this work, it does appear that there is a significant pattern of activity among certain state political cultures that informs a broad understanding of teacher evaluation policy across the US. For example, 75% of states with moralistic and individualistic political cultures use 'self-reflection' as a data point for summative evaluation. More specifically, 40% of the moralistic/individualistic states require Danielson and Marzano as a framework while 45% of the moralistic/individualistic states within our sample provide the option of using a state framework. The simple fact that only 15% of our sample states use other models suggests states with moralistic/individualistic political culture appear to be shifting to where government takes a more active or traditional role in control through policy. While more analyses would have to occur across the US to make this broad claim, it helps to initiate research that will inform how and to what extent state-level decisions about TSES development are made.

As faculty who prepare educational leaders in principal preparation programs, it is important to mention here that in spite of some states that treat teacher evaluation as if NCLB was still the law, *educational leaders* have agency at the local level to use supervision, *if taught*, and to recognize, *if taught*, the trends and socio-political ideologies that are remnants of NCLB. We place emphasis on referring to principals who enact supervision as educational leaders because we maintain that *educational administration* in its current form is aligned more with managing a building and the technical aspects associated with it than instructional supervision, which is about improving teaching (Waite & Nelson, 2005). Still, due to the overwhelming demand of accountability measures and standardized testing, principals end up spending much more time on the technical, not relational, aspects of the profession (Nelson et al., 2008). As a result, often, principals treat professional growth and professional development as synonymous with summative evaluation. While this may be understood as a form of educational malpractice when considering adult learners and the need to develop reflective stances regarding instruction, we as faculty in principal preparation also realize that we cannot assume each principal or assistant

principal in our sample or across the US has been taught/mentored/coached/supervised to serve a democratic society using standards for democratic supervision. Due to the fact it is important to develop leaders who will facilitate school improvement through supervision, we realize that it is important to also consider how, we, clinical supervision scholars within educational leadership prepare leaders. This research helps us accomplish that.

References

Archer, J., Cantrell, S., Holtzman, S., Joe, W., Tocci, C., & Wood, J. (2016). *Better feedback for better teaching: A practical guide to improve classroom observations*. Jossey-Bass.

Cheng, L. C. (2015). From subjectivity to objective evaluation: A techno-rationalist approach of assessment design for art and design education. *International Colloquium of Art and Design Education Research*, 421-430. http://dx.doi.org/10.1007/978-981-287-332-3_44

Close, K., Amrein-Beardsley, A., & Collins, C. (2018). *State-level assessments and teacher evaluation systems after the passage of the Every Student Succeeds Act: Some steps in the right direction*. National Education Policy Center.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Sage.

Dincer, O., & Johnston, M. (2016). Political culture and corruption issues in state politics: A new measure of corruption issues and a test of relationships to political culture. *Publius: The Journal of Federalism, 47*(1), 131-148. https://doi:10.1093/publius/pjw026

Danielson, C. (2011). *The framework for teaching evaluation instrument*. The Danielson Group.

Derrington, M. L., & Campbell, J. W. (2015). Implementing new teacher evaluation systems: Principals' concerns and supervisor support. *Journal of Educational Change, 16*, 305-326. http://dx.doi.org/10.1007/s10833-015-9244-6

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, 24*(3), 246-262. http://dx.doi.org/10.1080/13540602.2017.1421164

Elazar, D. J. (1994). *The American mosaic*. Westview Press.

Every Student Succeeds Act (ESSA) (2015). Pub. L. No. 114-95 § 114 Stat. 1177 (2015-2016).

Febey, K., & Louis, K.S. (2008). Political cultures in education at the state and local levels: Views from three states. In B.S. Cooper, J.G. Cibulka, & L.D. Fusarelli (Eds.), *Handbook of education politics and policy* (pp. 52-72). Routledge.

Fisher, P. I. (2016). Definitely not moralistic: Sate political culture and support for Donald Trump in the race for the 2016 republican presidential nomination. *PS: Political Science & Politics, 49*(4), 743-747. https://doi:10.1017/S1049096516001475

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. Pantheon Books.

Fowler, F. C. (2013). *Policy studies for educational leaders: An introduction* (4th ed.). Pearson.

Glanz, J. (2018). Chronicling Perspectives about the State of Instructional Supervision by Eight Prominent Scholars of Supervision. *Journal of Educational Supervision, 1*(1), 1-17. https://doi.org/10.31045/jes.1.1.1

Glanz, J., & Hazi, H.M. (2019). Shedding light in the phenomenon of supervision traveling incognito: A field's struggles for visibility. *Journal of Educational Supervision, 2*(1). http://dx.doi.org/10.31045/jes.2.1.1

Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (2018). *SuperVision and instructional leadership: A developmental approach* (10th ed.). Pearson.

Gordon, S.P. (1992). Paradigms, transitions and the new supervision. *Journal of Curriculum and Supervision, 8*(1), 62-76.

Gordon, S. P. (2019). Educational supervision. Reflections on its past, present, and future. *Journal of Educational Supervision, 2*(2), 27-52. https://doi.org/10.31045/jes.2.2.3.

Grissom, J. A., & Bartanen, B. (2018). Strategic retention: Principal effectiveness and teacher turnover in multiple-measure teacher evaluation systems. *American Educational Research Journal*, 1-42. https://doi.org/10.3102/0002831218797931

Hale, K., & McNeal, R. (2010). Election administration reform and state choice: Voter identification requirements and HAVA. *Policy Studies Journal, 38*(2), 281-302. http://dx.doi.org/10.1111/j.1541-0072.2010.00362.x

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. http://dx.doi.org/10.1007/s11092-013-9179-5

Hazi, H. M. (2010). *The statewide teacher evaluation instrument: Icon or fetish of teacher quality*. Paper presented at the annual meeting of the Council of Professors of Instructional Supervision.

Hazi, H. M. (2019). Coming to understand the wicked problem of teacher evaluation. In S. J. Zepeda & J. A. Ponticell (Eds.), *The Wiley handbook of educational supervision* (pp. 183-208). Wiley Blackwell.

Hazi, H. M., & Arredondo Rucinski, D. (2009). Teacher evaluation as a policy target for improved student learning: A fifty-state review of statute and regulatory action since NCLB. *Education Policy Analysis Archives*, *17*(5), 1-19. http://dx.doi.org/10.14507/epaa.v17n5.2009

Heck, R. H., Lam, W. S., & Thomas, S. L. (2014). State political culture, higher education spending indicators, and undergraduate graduation outcomes. *Education Policy, 28*(1), 3-39. http://dx.doi.org/10.1177/0895904812453996

Honig, M. I. (2006). *New directions in education policy implementation*. Suny Press.

Horng, E. L., Klasik, D., Loeb, S. (2010). Principal's time use and school effectiveness. *American Journal of Education, 116*(4), 491-523. http://dx.doi.org/10.1086/653625

Ingersoll, R. M., & Collins, G. J. (2019). Accountability, control, and teachers' work in American schools. In S. J. Zepeda & J. A. Ponticell (Eds.), *The Wiley handbook of educational supervision* (pp. 159-182). Wiley Blackwell.

Kegan, R., & Lahey, L. L. (2016). *An everyone culture: Becoming a deliberately developmental organization*. Harvard Business School Publishing.

Knight, J. (2016). *Better conversations: Coaching ourselves and each other to be more credible, caring, and connected*. Thousand Oaks, CA: Corwin.

Kraft, M. A. & Gilmour, A. F. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly, 52*(5), 711–753. http://dx.doi.org/10.1177/0013161X16653445

Lipsky, M. (2010). Street-level bureaucracy: Dilemmas of the individual in public service. Russell Sage Foundation.

Louis, K.S., Febey, K., Gordon, M., Meath, J., & Thomas, E. (2006). *Educational leadership in the states: A cultural analysis.* Center for Applied Research and Educational Improvement, University of Minnesota.

Louis, K. S., Thomas, E., Gordon, M. F., & Febey, K. S. (2008). State leadership for school improvement: An analysis of three states. *Educational Administration Quarterly, 44*(4), 562-592. https://doi.org/10.1177/0013161X08323858

Louis, K.S, Wahlstrom, K. L., Michlin, M., Gordon, M., Thomas, E., Leithwood, K., Anderson, S. E., Mascall, B., Strauss, T., & Moore, S. (2010). *Learning from leadership: Investigating the links to improved student learning, final report of research to the*

Wallace Foundation. Wallace Foundation Learning from Leadership Project: The University of Minnesota and The University of Toronto. Retrieved on January 23, 2020 from https://www.wallacefoundation.org/knowledge-center/Documents/Investigating-the-Links-to-Improved-Student-Learning.pdf

Marshall, K. (2013). *Rethinking teacher supervision and evaluation* (2nd ed.). Jossey-Bass.

Marshall, C., Mitchell, D., & Wirt, F. (1989). *Culture and education policy in the American states.* Falmer Press.

Mathis, W. J., & Trujillo, T. M. (2016). *Lessons from NCLB for the Every Student Succeeds Act.* National Education Policy Center.

Mead, L. M. (2004). State political culture and welfare reform. *Policy Studies Journal, 32*(2), 271-296. http://dx.doi.org/10.1111/j.1541-0072.2004.00065.x

Mette, I. M., Range, B. G., Anderson, J., Hvidston, D. J., & Nieuwenhuizen, L. (2015). Teachers' perceptions of teacher supervision and evaluation: A reflection of school improvement practices in the age of reform. *Educational Leadership Review, 16*(1), 16-30.

Mette, I. M., Range, B. G., Anderson, J., Hvidston, D. J., Nieuwenhuizen, L, & Doty, J. (2017). The wicked problem of the intersection between supervision and evaluation. *International Electronic Journal of Elementary Education, 9*(3), 709-724.

Mette, I. M., & Riegel, L. (2018). Supervision, systems thinking, and the impact of American school reform efforts on instructional leadership. *Journal of Cases in Educational Leadership, 21*(4), 34-51. http://dx.doi.org/10.1177/1555458918759696

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.

Mondak, J. J., & Canache, D. (2014). Personality and political culture in the American states. *Political Research Quarterly, 67*(1), 26-41. http://dx.doi.org/10.1177/1065912913495112

No Child Left Behind Act (NCLB) of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002).

Orr, S. (2007). Assessment moderation: Constructing the marks and constructing the students. *Assessment & Evaluation in Higher Education, 32*(6), 645-656. http://dx.doi.org/10.1080/02602930601117068

Pennington, C. and Mead, S. (2016). *For good measure? Teacher evaluation policy in the ESSA era.* Bellwether Education Partners. Retrieved from https://bellwethereducation.org/sites/default/files/Bellwether_ForGoodMeasure-GPLH_Final1216%20(1).pdf

Pye, L. W. (2015). Introduction: Political culture and political development. In L. W. Pye & S. Verba (Eds.), *Political culture and political development* (pp. 3-26). Princeton University Press.

Reid, D. B. (2017). U.S. principals' interpretation and implementation of teacher evaluation policies. *The Qualitative Report, 22*(5), 1457-1470. Retrieved from https://nsuworks.nova.edu/tqr/vol22/iss5/19

Rigby, J. G., Larbi-Cherif, A., Rosenquist, B.A., Sharpe, C. J., Cobb, P., & Smith, T. (2017). Administrator observation and feedback: Does it lead toward improvement in inquiry oriented math instruction? *Educational Administration Quarterly, 53*(3), 475-516. http://dx.doi.org/10.1177/0013161X16687006

Robinson, V. M. J, Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly, 44*(5), 635-674. http://dx.doi.org/10.1177/0013161X08321509

Ruff, R. R. (2019). State-level autonomy in the era of accountability: A comparative analysis of Virginia and Nebraska education policy through No Child Left Behind. *Education Policy Analysis Archives, 27*(6), 1-31. http://dx.doi.org/10.14507/epaa.27.4013

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Sage.

Salo, P., Nylund, J., & Stjernstrøm, E. (2015). On the practice architectures of instructional leadership. *Educational Management Administration & Leadership*, *43*, 490-506. http://dx.doi.org/10.1177/1741143214523010

Santoro, D. A. (2018). Is it burnout? Or demoralization? *Educational Leadership, 75*(9), 10-15.

Spillane, J. P., & Kenney, A. W. (2012). School administration in a changing education sector: the U.S. experience. *Journal of Educational Administration, 50*(5), 541-561. http://dx.doi.org/10.1108/09578231211249817

Spillane, J.P., Seelig, J.L., Blaushild, N.L., Cohen, D.K., Peurach, D.J. (2019). Educational system building in a changing educational sector: Environment, organization, and the technical core. *Educational Policy, 33*(6), 846-881. http://dx.doi.org/10.1177/0895904819866269

Stark, M. D., McGhee, M. W., & Jimerson, J. B. (2017). Reclaiming instructional supervision: Using solution-focused strategies to promote teacher development. *Journal of Research on Leadership in Education 12*(3), 215-238. http://dx.doi.org/10.1177/1942775116684895

United States. (1964). National defense education act of 1958, as amended by the 88th Congress. U.S.

Urban, W. J. (2010). *More than science and Sputnik: The National Defense Education Act of 1958*. The University of Alabama Press.

Waite, D., & Nelson, S. W. (2005). Educational leadership reconsidered. *La Revista Española de Pedagogía*, *63*(232), 389-406.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. The New Teacher Project.

Welch, S. (1993). *The concept of political culture*. New York: St. Martin's Press, Inc.

Wilder, M. (2017). Comparative public policy: Origins, themes, new directions. *Policy Studies Journal, 45*(1), 47-66. http://dx.doi.org/10.1111/psj.12200

Wolcott, H. (2009). *Writing up qualitative research.* Sage.

Zepeda, S. J. (2017). *Instructional supervision: Applying tools and concepts* (4th ed.). Eye on Education.

Zepeda, S. J., & Jimenez, A.M. (2019). Teacher evaluation and reliability: Additional insights gathered from inter-rater reliability analyses. *Journal of Educational Supervision, 2*(2). https://doi.org/10.31045/jes.2.2.2

Zepeda, S. J., & Ponticell, J. A. (Eds.) (2019). *The Wiley handbook of educational supervision*. Wiley Blackwell.

Author Biographies

Ian M. Mette is an associate professor of educational leadership and the program coordinator of the educational leadership program at the University of Maine. His research interests include teacher supervision and evaluation, school reform, and bridging the gap between research and

practice to inform and support school improvement efforts. Specifically, his work targets how educators, researchers, and policy makers can better inform one another to drive school improvement and reform policy. His first book, *A New Leadership Guide for Democratizing Schools from the Inside Out: The Essential Renewal of America's Schools*, coauthored with Carl Glickman, was published in spring 2020. You can follow him on Twitter (@IanMette) to see more of his work and interests.

Israel Aguilar is an assistant professor of organization and school leadership at the University of Texas Rio Grande Valley. Dr. Israel Aguilar earned a B.A. and M.A. in English literature from the University of Texas at Brownsville. He also holds a Ph.D. in education with a major in school improvement from Texas State University. In addition, he holds Texas educator licensure in the following areas: superintendent, principal, and classroom teacher. Dr. Aguilar's professional experience include serving as a school administrator in Dallas, and as a teacher in Austin and Brownsville.

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.