

2022

Shaping Educator Sensemaking in Complex Systems? Policy-Directed Teacher Evaluation Models as Boundary Objects

Kelley M. King

University of North Texas, kelley.king@unt.edu

Noelle A. Paufler

Clemson University, npaufle@clemson.edu

Rachel L. Biritz

Clemson University, rlebron@clemson.edu

Ryan M. Smits

University of North Texas, ryan.smits@unt.edu

Follow this and additional works at: <https://digitalcommons.library.umaine.edu/jes>



Part of the [Educational Administration and Supervision Commons](#), and the [Teacher Education and Professional Development Commons](#)

Recommended Citation

King, K. M., Paufler, N. A., Biritz, R. L., & Smits, R. M. (2022). Shaping Educator Sensemaking in Complex Systems? Policy-Directed Teacher Evaluation Models as Boundary Objects. *Journal of Educational Supervision*, 5 (1). <https://doi.org/10.31045/jes.5.1.5>

This Intersection of Supervision and Evaluation is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Journal of Educational Supervision by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.

Shaping Educator Sensemaking in Complex Systems? Policy-Directed Teacher Evaluation Models as Boundary Objects

Journal of Educational Supervision

88 – 116

Volume 5, Issue 1, 2022

DOI: <https://doi.org/10.31045/jes.5.1.5>

<https://digitalcommons.library.umaine.edu/jes/>

Kelley M. King¹, Noelle A. Paufler², Rachel L. Birtz², and Ryan M. Smits¹

Abstract

This study examined a state-wide, policy-directed teacher evaluation model implemented across public schools and educator preparation programs. Such models are grounded in a theory of action that situates teacher learning within social relationships, yet does not account for the complexity of systems. Results challenge policy's implicit theory that an evaluation model can function as a boundary object to create a common understanding of good teaching and positively impact teacher professional practice. We found contradictory evidence that the model served as a boundary object that facilitated shared sensemaking as mediated understandings of good teaching collided with expectations in classroom contexts.

Keywords

teacher evaluation; supervision; educational policy; systems theory; teacher preparation; complex adaptive systems; sensemaking

¹ University of North Texas

² Clemson University

Corresponding Author:

Kelley M. King (College of Education, Teacher Education and Administration, University of North Texas, Matthews Hall 204-M, Denton, TX, USA 76201)
email: kelly.king@unt.edu

Introduction

In recent decades, policymakers have supported the development and implementation of teacher evaluation systems posited to (a) measure teacher effectiveness for purposes of accountability and (b) support teacher professional learning and growth. We refer to these systems as Performance Evaluation and Professional Growth (PE/PG) models (Mette et al., 2019). Such models include observations and ratings of professional practice, conferencing between evaluators and teachers, and, potentially, measures of student learning. As posited in policy, PE/PG models foster professional learning through relationships, mentorship/coaching, and communication while defining and measuring and holding educators accountable for quality teaching for practitioners across complex systems. Existing studies broadly examine PE/PG practices and impacts in Early Childhood through 12th-grade (EC-12) in the United States (e.g., Amrein-Beardsley, 2014; Derrington & Campbell, 2015; Hazi, 2018; Lavigne & Good, 2019; Robertson-Kraft & Zhang, 2018) and other countries including Chile (Taut & Sun, 2014), China (Liu & Zhao, 2013), Spain (Cuevas et al., 2018), and the United Kingdom (Hopkins et al., 2016).

These models, with policy-directed definitions of teacher quality, now operate across organizations and levels of the profession, including EC-12 schools and educator preparation programs (e.g., Paufler et al., 2020a). In response to critiques of educator preparation as lacking coherence across the pre-service through in-service years (Brandon & Derrington, 2019), some states have created a PE/PG model that prescribes a standardized view of teacher quality for both pre-service and in-service teachers. Less attention has been given to research on the evaluation of clinical teachers (CTs), also known as student teachers or teacher interns (Burns & Badiali, 2015; Burns et al., 2016; Nolan & Hoover, 2010), particularly from a systems perspective (Cochran-Smith et al., 2014).

This paper reports on the initial implementation of a statewide PE/PG model across both EC-12 public schools and preparation programs in Texas. We present data from clinical supervisors (hereafter supervisors) and CTs associated with one large university-based preparation program. Participants experienced the initial implementation of the model as part of the CTs' final semester clinical teaching experience in EC-12 schools. The purpose of the study was to help us better understand the impact of PE/PG models in complex systems—specifically, the evaluation of CTs who operate across the borders of preparation programs and EC-12 schools. We explored the implicit theory of action we posit underlies these policies, namely that a PE/PG model can function as a boundary object that creates a common understanding of quality teaching as developed by the state.

This study was guided by one overarching question: How does this PE/PG model shape sensemaking about teaching in complex systems? and two specific sub-questions:

- In what ways does this PE/PG model create or fail to create common understandings of quality teaching within and across communities (preparation program and EC-12 schools)?
- What is the perceived impact of this PE/PG model among differently situated individuals within broader systems?

Conceptual Framework

These PE/PG models, implemented across settings, are intended to hold all educators, including CTs, accountable while supporting their professional growth through evaluation and supervision practices. Policy-directed PE/PG models are grounded in an implicit theory of action that situates teacher learning within and across communities of practice. Social learning theory and associated empirical research support the idea of professional learning broadly as participation in the practices of social communities within and across organizations (e.g., Dinsmore & Wenger, 2006; Wenger, 1998). In particular, Wenger's (1998) social learning framework involves understanding that "knowledge creation is social, produced through meaningful dialogue and conversations that occur within communities" (Patton & Parker, 2017, p. 352).

Wenger understands learning to be rooted in social learning systems, which he calls communities of practice. Accordingly, Wenger (2010) notes that communities of practice display many characteristics of systems in general, including emergent structure, complex relationships, and dynamic boundaries. In Wenger's (1998) model, *individual learning* entails engaging in practices of a community, *community learning* requires refining practice and "ensuring new generations of members," and *organizational learning* "is an issue of sustaining the interconnected communities of practice through which an organization knows what it knows and thus becomes effective and valuable as an organization" (pp. 7-8).

However, conceptual models of teacher evaluation, as defined in federal and state policy and developed and implemented in practice, do not explicitly intersect with the conceptual frameworks for such learning, e.g., communities of practice and social learning theory (Levine, 2010; Palmeri & Peter, 2019). Such models have been criticized for conflating supporting teacher growth through supervision and evaluation of effectiveness (Mette & Riegel, 2018; Palmeri & Peter, 2019). Further, the role of these models in social learning within and across educational organizations remains under-theorized (Levine, 2010; Mette, 2019). Wenger's framework includes several concepts of particular relevance to this study: communities of practice, sensemaking in community, and boundaries.

What Is a Community of Practice?

Wenger (1998) understands shared practice to be "the source of coherence" for the specific type of community he calls the community of practice (p. 72). In a Wengerian sense, the three critical dimensions of practice are mutual engagement, a joint enterprise, and a shared repertoire. Distinguishing a community of practice from a group, team, or network, he argues that community of practice membership is not based on social category, the flow of information across a network of individuals, or geographic proximity. Rather, *mutual engagement* as a practice in a community "exists because people are engaged in actions whose meanings they negotiate with one another" (Wenger, 1998, p. 73). Accordingly, a *joint enterprise* "is the result of a collective process of negotiation that reflects the full complexity of mutual engagement," (p. 77-78) which is defined by participants and extends beyond just a stated goal to create mutual accountability.

Explaining that communities of practice are not self-contained but rather develop in context, Wenger (1998) notes that the community of practice enterprise is defined by individual members based on their own position in the broader system and the pervasive influence of the employing institution. Jointly pursuing an enterprise fosters the development of a *shared repertoire* of resources for negotiating meaning, which “gives coherence for the medley of activities, relations, and objects involved” in the community of practice (Wenger, 1998, p. 82). Although often heterogeneous, “elements of the repertoire . . . gain their coherence” based on their belonging “to the practice of a community pursuing an enterprise” (p. 82). Individuals can simultaneously participate in multiple communities of practice, which can potentially create various forms of continuity among them, even without active effort to sustain those connections (Wenger, 1998). Accordingly, CTs, their supervisors, and mentor teachers are part of overlapping communities of practice nested within the broader educational system of EC–12 and the preparation program.

Sensemaking in Community

Members of a community of practice “join forces in making sense of addressing challenges they face individually or collectively” (Wenger et al., 2011, p. 9). A community or network creates value when participants engage in information sharing, learning from each other’s experiences, helping others with challenges, co-creating knowledge (sensemaking), and stimulating change (Wenger et al., 2011). In this case, participation in educational processes (clinical teaching) is effective in fostering learning (sensemaking) because “there is a match between knowing and learning, between the nature of competence and the process by which it is acquired, shared, and extended” (pp. 101-102).

Boundaries, Boundary Objects, and Boundary Spanning

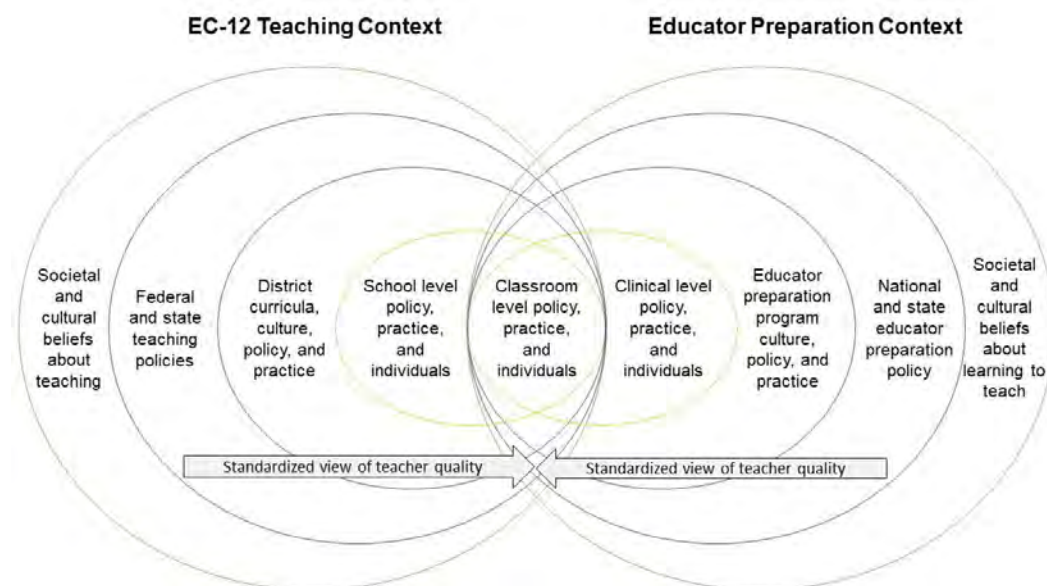
As defined by Wenger (1998), *boundary objects* are “artifacts, documents, terms, concepts, and other forms of reification around which communities of practice organize their interconnections” (p. 105). Wenger (1998) noted that, “Leigh Star coined the term boundary object to describe objects that serve to coordinate the perspectives of various constituencies for some purpose” (p. 106). A boundary object is adaptable so that various contexts can use and understand the object in a way that aligns with their community, yet “robust enough to maintain a common identity across sites” (Star & Griesmer, 1989, p. 393). “When a boundary object serves multiple constituencies, each has only partial control over the interpretation” (Wenger, 1998, pp. 107-108). Thus, boundary objects are “nexus of perspectives” (Wenger, 1998, p. 108).

A boundary object functions as a way to build common understandings about intersecting practices. When professionals operate within and across systems, they engage in boundary crossings. Education researchers have noted the ways that clinical experiences span organizational boundaries (Anagnostopoulos et al., 2007; Montecinos et al., 2015). Individuals involved in clinical experiences are members of the university community while also connected to (and operating within) the school community as brokers or “boundary spanners” (AACTE Clinical Practice Commission, 2018, p. 11).

We posit that policymakers assume PE/PG models present a reified understanding of quality teaching intended to cascade through systems, shaping sensemaking (see Figure 1). The intended

purpose of implementing the PE/PG model across settings would be to use it as a boundary object around which educators adopt a common understanding of quality teaching and adapt their teaching accordingly.

Figure 1. *Systems Theory Model of Education Contexts*

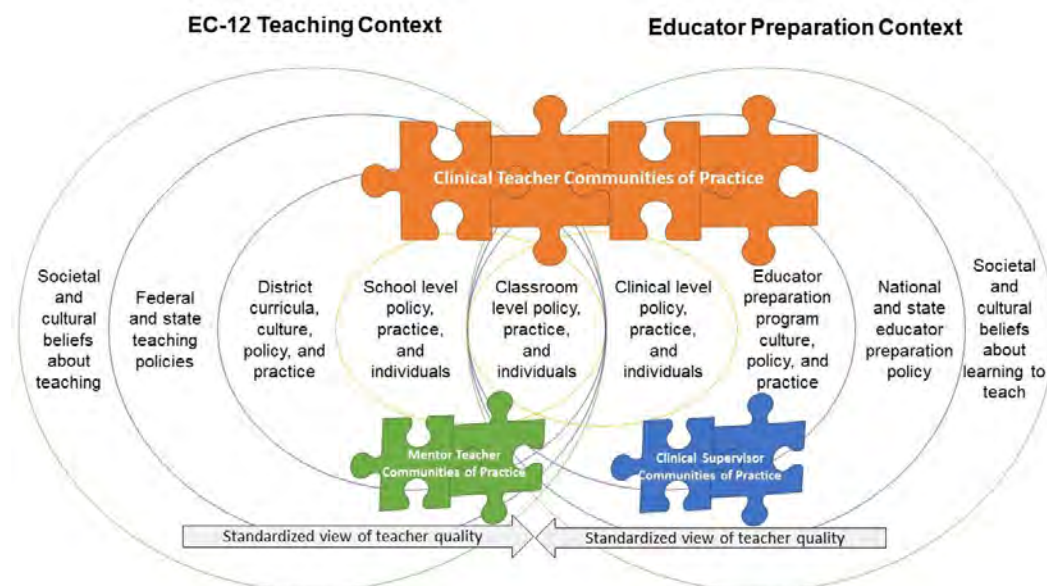


Note. Adapted from “Cascading, Colliding, and Mediating: How Teacher Preparation and K-12 Education Contexts Influence Mentor Teachers’ Work,” by R. Roegman and J. Kolman, 2020, *Journal of Teacher Education*, 71(1), p. 109 (<https://doi.org/10.1177/0022487119850174>). Copyright 2019 by the American Association of Colleges for Teacher Education.

Understanding Pre-service Teacher Development in Complex Systems

Educator preparation happens in overlapping contexts and systems (including preparation programs and EC-12 schools), necessitating an approach that recognizes “all parts of a system and [how] changes in one part of the system affect the context as a whole” (Cochran-Smith et al., 2014, p. 107). For researchers studying educator preparation, this requires that “teacher education actors, organizations, and processes at multiple levels can be fruitfully conceptualized as complex systems” (Cochran-Smith et al., 2014, p. 107) which include individuals (e.g., CTs and their supervisors), classrooms, schools within districts, and preparation programs (see Figure 2). Thus, pre-service teacher learning in clinical settings is supported by “providing student teachers with access to more knowledgeable others” who can “structure a strong support system that encourages and reflects on effective teacher education practices” (Bates et al., 2009, p. 90).

The negotiation of meaning that occurs between pre-service teachers and more knowledgeable others—sensemaking—is key to learning in clinical settings (Canipe & Gunckel, 2020). Sensemaking involves social learning: the way “people concerned with identity in the social context of other actors” extract and interpret cues (Weick et al., 2005, p. 409). Learning occurs through this structuring of experiential cues within social contexts (Holt & Cornelissen, 2014, p. 525). Thus, learning is contextual and, in many cases, polycontextual (Engeström et al., 1995).

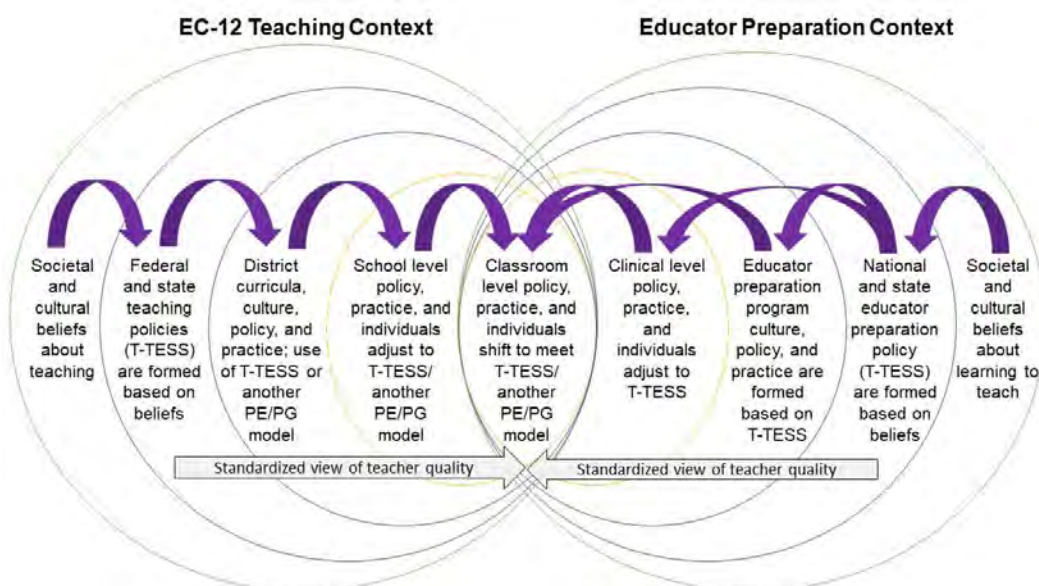
Figure 2. *Systems Theory Model of Education Contexts with Multiple Communities of Practice*

Note. Adapted from “Cascading, Colliding, and Mediating: How Teacher Preparation and K-12 Education Contexts Influence Mentor Teachers’ Work,” by R. Roegman and J. Kolman, 2020, *Journal of Teacher Education*, 71(1), p. 109 (<https://doi.org/10.1177/0022487119850174>). Copyright 2019 by the American Association of Colleges for Teacher Education.

Evaluation Within and Across Complex Systems

In their research on mentor teachers, Roegman and Kolman (2020) used the terms cascading, colliding, and mediating to describe how “individuals, policies, and beliefs influence the work”, and their research maps interactions within and across these systems (p. 113). Briefly, *cascading* refers to the “process by which an action occurring within one system successively triggers or initiates something in another system” (p. 113). *Colliding* “refers to the common occurrence of disagreement between messages, values or practices both within and across systems and contexts” (p. 113). *Mediating* refers to “the ways in which individuals within systems work to interpret ideas and policies through their own lens” (p. 113). Thus state-level policy directives (e.g., PE/PG models) can be understood to drive cascades, collisions, and mediations throughout levels of multiple systems (e.g., EC-12 and preparation program).

The state’s attempt to implement the same model across multiple organizations within the system suggests a theory of action that understands education as a continuous, interconnected system (see Figure 3). However, little research addresses evaluation within the complex settings and systems of clinical teaching, even as policy directives have focused on common systems of evaluation across settings. Examining the role of PE/PG models, in this particular case, offers the ability to understand “the dynamic and non-linear relationships between individuals, curricula, policies, and institutions within and across the systems and contexts” (Roegman & Kolman, 2020, pp. 109-110) and provides a basis for supporting system-wide learning.

Figure 3. *Cascading Systems Theory Model of Education Contexts*

Note. Adapted from “Cascading, Colliding, and Mediating: How Teacher Preparation and K-12 Education Contexts Influence Mentor Teachers’ Work,” by R. Roegman and J. Kolman, 2020, *Journal of Teacher Education*, 71(1), p. 109 (<https://doi.org/10.1177/0022487119850174>). Copyright 2019 by the American Association of Colleges for Teacher Education.

Methods

This exploratory study is the third case embedded in a larger, multiple-case study. The multiple-case study examined the implementation and impact of new teacher evaluation systems in two EC-12 school districts and one university-based preparation program. The first study examined school administrator and teacher perceptions of the implementation of a new teacher evaluation system in a large, suburban, fast-growth school district in the southwestern United States before the passage of the Every Student Succeeds Act of 2015 (Paufler, 2014; Paufler & Sloat, 2019). The second study explored perceptions of a new PE/PG model, the Texas Teacher Evaluation and Support System (T-TESS; Texas Education Agency [TEA], 2019), in one high school in a large, suburban, fast-growth district (Paufler et al., 2020b). This study centered the perceptions of CTs and their supervisors in a large, university-based preparation program in Texas, regarding T-TESS after the first year of implementation.

In this study, the focus was on the impact of changing policy and practice mandates on the perceptions and experiences of CTs and their supervisors. In particular, we sought to understand whether this model fostered common understandings of quality teaching among participants and across organizational boundaries. Case study allows us to apply and construct conceptual models that reflect real-world complexity (Yin, 2018)—here, the overlapping policies, practices, institutional and organizational structures, and communities of practice in which the CTs and supervisors operate.

Setting

The research setting is a university-based preparation program in Texas with approximately 600 initial certification completers annually. In the 2016-2017 academic year, Texas replaced an existing model, Professional Development and Appraisal System, with T-TESS. Public school districts were required to adopt T-TESS or a locally developed aligned model that must be approved by the state, beginning in 2016-2017. Beginning in the 2017-2018 academic year, Texas mandated that preparation programs train supervisors on T-TESS. This encouraged preparation programs not using T-TESS to adopt the rubric. In preparation, the preparation program field director in this study received state-mandated training and then conducted local training for the preparation program's supervisors. T-TESS represents a departure from its predecessor which essentially functioned as a "checklist" for observing classroom instruction. Instead, T-TESS incorporates a rubric based on the Danielson (2013) Framework for Teaching. The rubric consists of 16 dimensions across four domains: planning, instruction, learning environment, and professional practices and responsibilities (TEA, 2016). As designed by the state, T-TESS ratings included improvement needed, developing, proficient, accomplished, and distinguished. However, the preparation program removed the highest rating (distinguished) and added an unsatisfactory rating that would be used to identify pre-service teachers who would not be recommended for certification.

Participants

We invited all supervisors who had evaluated CTs using T-TESS in fall 2018 and spring 2019 and all CTs who were evaluated in spring 2019 to participate. We used a mixed-methods design with a quantitative survey nested within the broader case study (Yin, 2018). We do not report demographic data of survey, interview, and focus group participants to protect participant confidentiality. No personally identifiable information, individual teacher evaluation results, or student-level achievement data was collected.

Data Collection

Survey

Participants responded to parallel online surveys in fall 2018 (supervisors, $n=22/32$, 68.8%) and spring 2019 (CTs, $n=76/331$, 23.0%). The surveys were developed by one of the researchers (Paufler, 2014; see also Paufler & Sloat, 2019). The original survey instruments were developed based on data collected during extensive interviews with teachers and supervisors, reviewed by an external panel of experts, and administered to almost 1,500 teachers ($n=1,051/1,444$). The survey instruments included both open- and closed-ended items, all descriptive, which precluded the calculation of reliability (Paufler, 2014; see also Paufler & Sloat, 2019). To ensure that the survey instruments reflected the context in this study, we met with the preparation program field director in advance to review the instruments and based on feedback received, removed a limited number of non-applicable items and slightly modified word choice in a few items to reflect the language used in T-TESS and in the preparation program's trainings and resources. In total, the adapted supervisor and CT surveys had 18 and 17 closed-ended, descriptive items, respectively, and four open-ended items.

Interviews and Focus Groups

Multilevel nested samples (Onwuegbuzie & Collins, 2007) were used for interviews and focus groups. In spring 2019, we conducted interviews with the field director and supervisors who indicated willingness to participate on the survey ($n=7$). Additionally, we conducted two focus groups and one interview with elementary ($n=2$), middle ($n=9$), and secondary ($n=1$) CTs at the end of clinical teaching in spring 2019. The interview and focus group protocols included 15 questions related to perceptions of the purpose of T-TESS, measuring teacher quality in general and using T-TESS, the impact of T-TESS on professional practice, and improving T-TESS implementation.

Data Analysis

We analyzed interview and focus group data (Yin, 2018) and open-ended survey items using reflexive thematic analysis (Braun & Clarke, 2006; 2019; Bryne, 2021). Several rounds of coding were conducted using ATLAS.ti. Wenger's social learning theory (1998) provided sensitizing concepts as we conducted the qualitative analysis. According to Charmaz (2003), sensitizing concepts are "those background ideas that inform the overall research problem" and that serve as "points of departure" for data analysis (Charmaz 2003, as cited in Bowen, 2006, p. 14). These included, but were not limited to, communities of practice, sensemaking, organizations, identity, and concepts related to boundaries. Additional cycles of coding involved synchronous and asynchronous researcher reflection as we developed consensus on codes and themes.

Separately, we analyzed close-ended survey responses using descriptive statistics in JMP and the IBM Statistical Package for the Social Sciences. We computed descriptive statistics (means and standard deviations) using respondents' numerical response to sets and series of Likert-type survey items and rank-ordered participant responses to determine frequency. While we calculated Spearman's intercorrelation and logistic regression coefficients and found statistical significance for some analyses, we are only reporting results for a subset of descriptive survey items here. Since there was no random assignment and no random selection, associations can only be made for the participants in this study. We designed the study so that readers might make naturalistic, not statistical, generalizations from the findings (Stake & Trumbull, 1982).

Findings

In our analysis, we integrated findings across qualitative and quantitative data. Here, we report findings related to two superordinate themes: *Making Sense of T-TESS versus Making Sense of Teaching* and *Perceptions of T-TESS Impact*. *Making Sense of T-TESS versus Making Sense of Teaching* refers to the way that both CTs and supervisors were making sense of T-TESS. At the same time, CTs were making sense of teaching as it was practiced in their school setting, taught in their university coursework, and evaluated via T-TESS. As we will further discuss throughout this section, there were parallels and differences in perceptions of T-TESS across differently situated individuals.

Both CTs and supervisors made sense of T-TESS within multiple, overlapping communities. CTs operated not only as pre-service teachers within the EC-12 school setting but also as students within the preparation program. How CTs made sense of teaching was contextualized, primarily based on their mentor teacher's knowledge and practices. They strategized with their mentor teacher to understand how to model quality teaching for observations. When contextualized practices aligned with expectations related to T-TESS, CTs generally accepted T-TESS as valid. However, when their supervisor's understanding of quality teaching vis-a-vis T-TESS collided with the contextualized practices of the EC-12 setting or with their own identity, CTs objected to having to perform inauthentically for evaluation. CTs noted this performativity collided with their beliefs about centering students. When supervisors operated together as a community of practice, their sensemaking focused less on contextualized understanding of quality teaching and more on establishing a common understanding of T-TESS rubric descriptors and performance levels in order to calibrate ratings. Findings related to CT and supervisor perceptions on making sense of T-TESS versus making sense of teaching, and the impact of T-TESS are presented in Table 1. Below we discuss the subordinate themes that comprise the two broad superordinate themes.

Table 1
Integrated Findings for CTs and Supervisors Regarding T-TESS

Theme	Clinical Teacher (CT) Perceptions	Clinical Supervisor Perceptions
Making Sense of T-TESS versus Making Sense of Teaching	<p>A. The majority of CTs (66%) believed T-TESS covers all/most of the important characteristics of quality teaching. Some noted it did not account for context.</p> <p>B. Aspects of T-TESS were perceived as inauthentic, including early submission of detailed lesson plans and an artificially limited window for observation. Demands of T-TESS inappropriately drove instruction.</p> <p>C. Some CTs believed T-TESS was not objective, but was mediated by their supervisor's understanding. They believed supervisors were experienced teachers with a general understanding of quality teaching. However, their ratings represented one person's perspective.</p> <p>D. Sensemaking focused on understanding supervisor expectations.</p> <p>E. CTs believed that T-TESS rubric included criteria beyond the scope of clinical teachers, particularly learning environment (Domain 3) and professional practices and responsibilities (Domain 4).</p>	<p>A. Nearly all supervisors (90%) believed T-TESS covers all/most of the important characteristics of quality teaching.</p> <p>B. Supervisors believed they could apply T-TESS objectively to ensure reliable and fair ratings. They desired additional training for themselves and even more so for CTs.</p> <p>C. Some supervisors believed that scores were inappropriately influenced by state requirements that scores be concentrated at the middle of the scale or by supervisor's belief that CTs should show consistent growth.</p> <p>D. Sensemaking focused on calibration and ratings.</p> <p>E. Supervisors believed that T-TESS rubric included criteria beyond the scope of CTs, particularly learning environment (Domain 3) and professional practices and responsibilities (Domain 4).</p>
Perceived Impact of T-TESS	<p>A. CTs had divergent views on T-TESS's impact on teaching. Some cited reflection on their teaching practices and dialogue with their supervisor as promoting growth. However, some CTs viewed the need to perform to their supervisors' mediated understanding of T-TESS as detracting from their focus on their EC-12 students.</p> <p>B. More than half of CTs (55%) believed T-TESS had no real impact on student academic achievement and learning.</p>	<p>A. Supervisors generally believed that T-TESS positively impacted CTs' and their own professional practice. They most frequently cited clarity/focus on the important aspects of quality teaching, opportunities for dialogue, and reflection on teaching practices as promoting CTs' growth.</p> <p>B. Most supervisors (63%) believed T-TESS had a positive impact on student academic achievement and learning.</p>

Clinical Teachers Making Sense of T-TESS versus Making Sense of Teaching

Theme 1 for CTs comprises three subthemes: *Sensemaking in Communities of Practice*, *Lack of Common Understanding Across Systems*, and *Tensions and Incongruities*.

Subtheme: Sensemaking in Communities of Practice

There was a wide range of understanding of T-TESS for CTs, who had first seen the rubric and procedures during training at the beginning of their clinical teaching semester. In focus groups, some CTs seemed not to recognize the term T-TESS, but quickly figured out that we were referring to the evaluation practices. Their reported understanding of T-TESS was mixed. A CT noted: "Now we're at the end and I still haven't figured it out." Others believed that the combination of T-TESS training and practice they received was adequate and expressed that it made sense to them. For example, a CT reported: "To me, it was very user-friendly. I was able to understand it after my first T-TESS evaluation, I was able to follow...how the scores are formatted and all the technicalities of it."

CTs described making sense of T-TESS within overlapping communities of practice, which included their supervisor and mentor teacher. For example, a CT expressed appreciation for the support around T-TESS that was provided by the preparation program: "Having the experience here while I was still a student, while I could still ask for help, while I can still be, 'I don't understand this, I feel lost, I feel scared, and I have the support system here.'" CTs also expressed that they engaged in making sense of T-TESS by working with their mentor teachers. However, in some placements, mentor teachers had little knowledge of T-TESS, due to the district using a locally developed model. Thus, when mentor teachers had questions, the CTs did not know the answers and had to go to their supervisors. Some CTs suggested that attending T-TESS information sessions with mentor teachers would give the latter "clearer expectations and [help them] understand the evaluations more."

As noted above, CTs generally understood the purpose of T-TESS and expressed appreciation for the support/supervision built into the PE/PG model. Most CTs ($n=61/76$, 80.3%) believed the primary purpose for evaluating the professional practice of CTs *should be* to help improve the quality of their professional practice. Of the remaining responses, the majority believed the primary purpose should be to determine whether CTs have the teaching competency necessary for certification ($n=11/76$, 14.5%). However, CTs had differing perceptions on the preparation program's purpose for evaluating CTs in reality: to help improve the quality of their professional practice ($n=35/76$, 46.1%), to determine whether CTs have the teaching competency necessary for certification ($n=19/76$, 25.0%), mostly to comply with state legislation ($n=18/76$, 23.7%), and to hold them accountable for their practices/performance ($n=4/76$, 5.3%).

As indicated in Table 1, most believed the rubric represented most, if not all, components of quality teaching. Nearly all CTs expressed that they appreciated the focus on growth and goal setting. With respect to feedback, they particularly valued written comments over numerical ratings. One CT suggested that the numerical ratings provided little actionable information: "It's a lot of numbers... so I'm just like, 'Er da da da da.'" Another CT discussed the struggle of being given the highest rubric score without verbal feedback: "Cool, I got another four, that's

great, but where's my comment?... I need to know what they think I should grow on.” Dismissing even positive information in her performance ratings, this CT sought out the detailed comments:

It doesn't matter if I got proficient or accomplished, it's the little tiny comments next to it saying 'I like this because you did this and this.' Being a little more detailed, not just giving me a number... That's the thing I value most, is the little comment section.

Although CTs valued the detailed feedback, they expressed frustration about delay between and perceived discrepancies across the verbal feedback they received during post-conferences immediately after their observations, their rubric scores, and written rubric comments. CTs believed the written feedback on the rubric more truly reflected the supervisors' perceptions, what they “actually thought,” and they used written commentary to make sense of the ratings.

At times, however, CTs struggled to understand the ratings and verbal feedback, and this undermined their acceptance of the T-TESS model. For example, a CT suggested that ratings were selected to artificially show growth from the beginning of the year to the end, with the same lesson receiving higher scores later in the semester: “Because the second one was just like it . . . and she loved my second thing. I didn't change anything, it was literally the same lesson plan.” This perceived lack of reliability undermined CTs' acceptance of feedback. Further, CTs resisted critical feedback that was given before a relationship with the supervisor had developed. A CT explained, “We meet her that one time and then you put your lesson plan in [the online system]. It gets ripped apart... I went into my first evaluation, I was shaking, I never feel like that in my classroom.” Another related the importance of establishing a relationship:

What she had to say was so beneficial for me, [but] going back and having to readdress it again online, you don't see her face in that moment, you just see the negative comment... It wasn't supposed to be negative but you read it that way because you felt defeated at the moment.

Subtheme: Lack of Common Understanding Across Systems

Although they valued connections with and feedback from their supervisors, CTs noted tensions, discrepancies, and misunderstandings in the way T-TESS was implemented across settings. A common complaint was the lack of detailed discussion unpacking T-TESS and how to perform to its standards in coursework:

We don't really discuss T-TESS in general ... we never really just went over the rubric. It's in the handbook. Okay ... but having that conversation, okay if you do this, or this is how this goes, and we want you to be able to do this, and then we go from there so we have an understanding of how we're supposed to be performing.

CTs wondered how to earn high ratings, particularly in the domains that evaded observation. One CT described being acknowledged by her school for volunteering her time after school, while being ranked developing on T-TESS domain 4 by her supervisor, “How am I developing? I

volunteered my track time, I helped out with the dance which is why they gave me this balloon.... Not everyone has the time to do this.... How do you test domain 4, at all?"

Supervisors Do Not Understand CTs' Classrooms. As CTs moved into classrooms, the lack of common understanding from the university to the classroom led to conflicting expectations and a sense of being caught in the middle. An elementary CT noted, "I know what the T-TESS is asking me to do, and then I also know what this district is looking at." Similarly, a CT expressed that even though they agreed with their supervisor's feedback, district requirements dictated lesson content and pacing, shaping their ability to respond: "I appreciate your feedback, but it's really not helpful because these aren't variables I can control." This CT continued, "It's not my classroom, it's not my curriculum, and they are my lesson plans, but at the same time the whole [team], we're teaching the same lesson.... There are a lot of parameters that aren't under my control."

When they struggled to understand T-TESS feedback from their supervisor, CTs worked with their mentor teachers to understand the comments and strategize how to improve. At times, this fostered joint sensemaking. A CT noted working with their mentor teacher to improve:

I looked at the negative feedback, and I talked to my mentor teacher, and she gave me that confidence boost that was like, okay she said this, how about we work on this specifically in our next lessons and then see where it goes in our next evaluation.

However, sometimes, mentor teachers simply rejected the evaluation feedback: "My first mentor teacher and my supervisor both said that I talk too fast. I go to my second mentor teacher, and he's like 'No, you talk perfectly normal speed.'"

Mentors Do Not Understand T-TESS. Further complicating their position negotiating overlapping communities, some CTs noted that their mentor teachers were unfamiliar with the rubric and observation process. Although all districts in Texas were required to implement a PE/PG model aligned to T-TESS for teacher evaluation, not all mentor teachers were familiar with the process or instruments, and they did not have experience evaluating other teachers. This left CTs struggling to coordinate across the preparation program and the classroom while feeling pressure from both sides. The following focus group exchange captures two CTs' frustration:

CT1: Can you please explain this to my mentor teacher, because there's no connection there. They don't understand. I don't understand. They don't understand. Then my supervisor comes in and she knows what's going on.

CT2: Then they get kind of crabby with us.

As the above quotation suggests, lack of common understanding of T-TESS potentially undermined relationships on which CTs rely for learning in communities of practice.

At times, this extended to CTs discounting the professional knowledge of their mentor teachers when it comes to meeting T-TESS expectations. A CT noted, "My mentor teacher honestly has no idea what to look for when I'm in the classroom." CTs believed that additional mentor teacher

training on T-TESS would foster common understanding across the system and relieve some of their stress: “It’s like those sessions that we sit through as student teachers, if our mentor teachers were there, and they’re hearing the same things that we’re hearing, then they would have clearer expectations, and understand the evaluations more.” CTs were making sense of teaching based on what they learned in their coursework, what they saw and practiced in their clinical settings, and through T-TESS. Often, these competing understandings collided.

Subtheme: Tensions and Incongruities

Although they accepted T-TESS as representative of quality teaching in theory, CTs contrasted the T-TESS rubric’s construct of quality teaching or “what they’re wanting and what it should be” with their understanding of “what is actually reality,” teaching real students in real classrooms. They noted several areas of tension between their sensemaking around T-TESS and their contextualized sensemaking of teaching. Collisions between T-TESS and their developing sense of teaching led CTs to critique what they saw as inauthenticity in T-TESS. This included time-limited observations, “preconferences” that consisted of submitting a lesson plan online well before the lesson, and the supervisors’ lack of contextual understanding. A CT interpreted their supervisor’s expectations and adjusted accordingly, but not in ways they believed served students:

We choose lessons that don't allow for a lot of differentiation of instruction because...I know [the supervisor] wanted direct instruction. She required us to have at least 25 minutes of it, which was an absurd thing to require somebody when you are also dictating [that on] the day we are meeting. So now I'm having to re-plan entire units around that, and it's absurd and detrimental to the students.

Another CT described lacking a personal connection to their supervisor:

She doesn't really care for me. When she's giving me my criticism, okay, I'm going to cater strictly to you because I know how you feel. I'm going to do my best to be the teacher you want me to be and how you see classes are supposed to go.

However, like others, this CT did not express belief that this was an accurate picture of their teaching or something they would practice outside of observations.

CTs described their efforts to respond to the tensions between observations and their contextualized practice in terms of “gaming T-TESS.” This referred to the way CTs performed for their supervisors during observations in tension with their normal practice of prioritizing students. A CT explained, “My first evaluation I was catering specifically towards my kids, and then after the evaluation, I said okay, ‘strictly for my supervisor.’” A CT described the disparity between how they teach generally versus during evaluations:

I just learned how to play off of her and so, of course I'm going to get a good score because I know exactly what you want. This isn't how I would teach.... I'm just playing to you.

The same CT indicated that they believed this performativity undermined the validity of evaluations. They argued observations should capture “actually how I am in the classroom” as opposed to “this is how I think I should be.” Further, this CT emphasized perceived rewards or consequences tied to T-TESS as a primary motivating factor, “I want a good score on this because my principal is going to look at this and I want a job.” Another CT explained the desire to get a good rating on the purportedly objective T-TESS rubric in a way that highlights inherent tensions, “Because we are scored on [an] objective point system, we are simply trying to find a way to beat the system and meet the expectations of our evaluator.” This CT believed that in performing for T-TESS evaluations “we destroy our own creativity.” Thus, as they described it, CTs were not using T-TESS to make sense of quality teaching; they were temporarily modifying their practice based on perceptions of their supervisors’ preferences and understandings of T-TESS. Sensemaking became a game.

Supervisors Making Sense of T-TESS Versus Making Sense of Teaching

Making Sense of T-TESS Versus Making Sense of Teaching was also a theme for supervisors. For supervisors, this theme consisted of two subthemes that paralleled the first and third subthemes above: *Sensemaking in Communities of Practice* and *Tensions and Incongruities*. Supervisors understood the purpose of T-TESS as “normaliz[ing] the types of teachers we’re producing and how we want them to teach.” In general, they valued having guidelines for observations and the potential for consistent, unbiased feedback to CTs. Supervisors found T-TESS to be a useful tool to “have some kind of baseline observation of teachers” such that “everybody’s being evaluated in the same manner. You have a little bit of consistency and it’s not just how you felt that day.” Thus, T-TESS evaluation was understood as “based on some protocol that has meaning and purpose behind it.” On surveys, supervisors expressed that the primary purpose the preparation program evaluates the professional practice of CTs is to help improve the quality of their professional practice ($n=12/21$, 57.1%) or to determine whether CTs have the teaching competency necessary for certification ($n=7/21$, 33.3%), which was similar to the perception of CTs. When asked what supervisors thought the primary reason for evaluating the professional practices of CTs *should be*, most supervisors ($n=14/21$, 66.7%) believed T-TESS was used to help improve the quality of CTs’ professional practice. Some supervisors noted that there are multiple reasons to evaluate CTs.

Unlike CTs, in general, supervisors did not struggle to understand the T-TESS rubric. Based on survey results, all supervisors who responded ($n=21/21$, 100%) agreed or strongly agreed that they had been well trained in the use of the T-TESS rubric to evaluate CTs. A supervisor noted that training on the rubric clarified the focus of the PE/PG model:

I think [consistency] helps me stay focused on specific areas they want me to pay attention to. Otherwise, I may be more interested in focusing on other things. But now that I have to cover these areas, it keeps me in line with what the state thinks is really important.

Further, supervisors believed that the rubric was a comprehensive model of quality teaching, even if “extremely wordy and extensive.” When asked to what extent components of the rubric incorporate the important characteristics of a good/effective teacher, the majority of supervisors

($n=19/21$, 90.5%) expressed that T-TESS covers all/most of the important characteristics of good/effective teaching. The remaining supervisors ($n=2/21$, 9.5%) believed that T-TESS covers some of the important characteristics. In interviews, they noted how specific components of the rubric accurately represent requirements of practice, “Domain two [instruction] is actually the instruction and to me that's extremely important. Obviously if you don't have domain three [learning environment] down fairly well, you're going to have some trouble with domain two.” However, supervisors noted that there are a few important attributes/indicators still missing that should be added, such as relationships with students and the physical classroom environment, which go beyond the descriptions of the rubric. Despite some comments regarding missing elements, in general the findings suggest that, compared to CTs, supervisors did not experience the tensions between quality teaching as defined by T-TESS and localized practices embedded within specific classrooms and districts.

Subtheme: Sensemaking in Communities of Practice

Like CTs, supervisors worked within communities of practice to make sense of T-TESS. Also, like CTs, they agreed that T-TESS presented a comprehensive model of quality teaching. They described building an understanding of T-TESS through communication with preparation program administrators, other supervisors, and CTs.

With Preparation Program Administrators. Supervisors strongly believed the preparation program communicated *adequately* ($n=11/21$, 52.4%) or *very well* ($n=10/21$, 47.6%) regarding T-TESS. After a full day of state-directed, university-delivered T-TESS training, supervisors generally felt prepared to use T-TESS. A supervisor described the training as an “intense day” during which, “we saw videos, we calibrated scoring, we discussed it, we did a lot of group work, we read through it, we broke it apart. Deconstructed, reconstructed.” This “intense day” purportedly created a common understanding so supervisors can use the instrument in a consistent manner. Indeed, supervisors recognized the need for consistency, what the state called “calibration,” in order to maintain fidelity in the implementation of T-TESS across sites and supervisors. Otherwise, as one supervisor explained, “sometimes people...default to what they think they know is best practice” rather than relying on the rubric and training. Therefore, supervisor sensemaking focused on consistent, reliable scoring using the T-TESS rubric as the standard of best practice. All supervisors who responded to the survey ($n=21/21$, 100.0%) agreed or strongly agreed that they were able to “evaluate clinical teachers in an objective and unbiased manner.”

With Other Supervisors. Despite believing they were able to objectively evaluate CTs based on the understanding of T-TESS developed through training, in interviews, supervisors frequently spoke of the need for more time to work together to calibrate scoring to ensure that the ratings all CTs received were fair. Thus, although supervisors reported receiving adequate training on the T-TESS rubric from the preparation program, two-thirds ($n=14/21$, 66.7%) agreed or strongly agreed that they would like more training, presumably to calibrate ratings and increase reliability. In the qualitative data, supervisors repeatedly expressed the challenges of consistent scoring, at times challenging, at times working toward the possibility of objectivity: “I know there's consistency in when we calibrate when we do some sampling together, but after that we really don't have much discussion the rest of the year.” A supervisor noted of practicing in

training sessions, “just being forced to rank something, see that I was along the lines of everybody else, helped.” Likewise, another expressed, “I would like to see us go through more of the calibration times and then really get into good, deep conversation about why we scored the way we did.” Thus, in conversations with other supervisors, supervisors described using T-TESS as a boundary object, shaping a communal sense of quality teaching as defined by the rubric, with reliable rating as the focus of these conversations.

With CTs. Supervisors also talked about T-TESS shaping conversations with their CTs about teaching. Mediating between T-TESS’s understanding of quality teaching and CTs as individuals, other supervisors utilized the T-TESS process in ways intended to help their CTs make sense of their own teaching style, presumably still as measured by T-TESS ratings: “My goal is to help them be the best they can be, help them develop their style that's going to work in a classroom, suggest ideas for them to, that will relate to their style.” Based on survey results, the vast majority of supervisors ($n=18/21$, 85.7%) agreed or strongly agreed that T-TESS would be significantly improved by providing CTs with additional training. A supervisor believed that T-TESS would shape how CTs perceive others’ teaching as they observe it in their clinical settings, directing “what they should be seeing when they're observing their cooperating [mentor] teacher, but also when they go out to visit others.” In this sense, the T-TESS model of quality teaching would become “that little voice in their head.” In other words, T-TESS could be used as a framework to guide thinking and conversations about teaching beyond formal evaluations. This data indicates that supervisors understand the potential of T-TESS to shape CTs’ internalized conceptions of quality teaching.

Subtheme: Tensions and Incongruities

Like CTs, supervisors’ sensemaking around T-TESS stumbled on tensions and incongruities inherent to the model and its theory of action. Supervisors generally perceived T-TESS as a comprehensive model of quality teaching that supported CT growth and produced fair and accurate evaluations. However, they frequently noted their own struggles to reconcile their belief that T-TESS could be an objective measure despite threats to validity and reliability that became evident as they implemented the model.

Validity. The vast majority of supervisors believed that the ratings CTs received were accurate and fair. For example, 95.2% ($n=20/21$) agreed or strongly agreed that their CTs received a rating that accurately represented their professional performance. However, like CTs, supervisors discussed struggling with elements of T-TESS they understood as undermining validity. Supervisors ($n=19/21$, 90.5%) overwhelmingly indicated in their survey responses that T-TESS incorporates all/most of the important characteristics of quality teaching but said in interviews that they grappled with using the PE/PG model as a holistic picture of CTs’ abilities. One supervisor explained this difficulty: “The human element is really hard to express on a series of sheets. We have to have something. I get it.” This supervisor continued, “When I have my conversations with them, we cover those things. We spend a lot more time talking about the other stuff.”

Supervisors also noted tensions inherent to T-TESS. They viewed T-TESS as a component of measuring quality teaching, but not the only measure. For example, one supervisor considered

“relationships with kids” and “the classroom environment” as something that “telegraphs this kind of unspoken message about the teacher and what’s happening in the room.” They noted that this “goes beyond classroom management, which the T-TESS focuses on.” Attention to those details, they suggested, could provide a more holistic measurement of quality teaching.

Conversely, T-TESS includes dimensions that CTs are not in control of, for example, “how the classroom is set up.” One supervisor noted that CTs might have “inherited a great situation” but when they did not, “They can’t go in and change all that.” While supervisors acknowledged that there were things outside of the CTs’ control, they nevertheless were required to provide scores for those dimensions. Supervisors noted that these tensions could be mitigated. For example, multiple observations were helpful in determining whether the CT was a good teacher because “you’re seeing the different snapshots.” Even with multiple observations, however, they mentioned that T-TESS does not capture “those little things of their personality and how they are with the kids.” Thus, when asked to reflect on T-TESS as a measure of quality teaching, supervisors generally noted that T-TESS missed some aspects and captured factors related to broader contexts that were neither under CTs’ control nor related to their abilities.

Reliability. In addition to expressing some concerns regarding validity, supervisors raised concerns about reliability. In interviews, supervisors, like CTs, reported that the usefulness of T-TESS is “subjective and open to interpretation.” This contradicts the survey data where all supervisors ($n=21/21$, 100.0%) indicated that they were able to evaluate clinical teachers in an objective and unbiased manner. One supervisor indicated that T-TESS could measure the most important aspects of quality teaching, but:

It depends on the supervisors and it depends on the clinical teachers, how prepared they are. This is definitely a very comprehensive and intensive instrument, and just as with all the instruments prior to this . . . it’s as useful an instrument as the person who uses it.

These subjectivities result in inconsistencies. As one supervisor explained, “there is a difference between the other supervisors . . . two of us would like to grade a little bit higher than we’ve been told to grade, and it’s very hard for us, so, yeah, there are inconsistencies.” Supervisors suggested these differences could be mitigated with “serious long-term training for inter-rater reliability.” Though reliably rating CTs across multiple contexts and sites was a point of major concern for supervisors during interviews about the overall effectiveness of T-TESS.

Ostensibly, one purpose of T-TESS is to address what policymakers perceive as a skew toward rating all teachers highly. T-TESS evaluators were instructed to rate “rock solid” teaching at the middle (i.e., proficient) of the 5-point scale, presumably to create a normal distribution for accountability purposes. Some supervisors struggled with this enforced down-scaling. One reported consciously scoring teachers lower than they thought was warranted: “I want to give them a higher score probably more than the observation instrument says to, so I’ve learned to sit on my fingers and stay at developing.” The state directive to rate at the middle of the scale made it “easier to grade,” but was unfair to CTs because when “they’re really good . . . there’s no real way to say it if ‘proficient’ is supposed to be such a rock star.”

Like CTs, supervisors also recognized that evaluations could be based on factors outside what they observed. They acknowledged that some supervisors “would not give [CTs] the highest

ratings the first observation even though they aced it.” Rather, supervisors felt pressure to return evaluations that showed progress and were hesitant to give high ratings early in the semester. A supervisor explained: “You see this trajectory that goes up. . . the [CTs] are getting better after each one. There's nothing wrong with that, but I don't believe that all [CTs] would have the same trajectory.” Thus, supervisors experience tension between how they want to score and how they believe they are expected to score. According to one supervisor, “calibration is an attempt to get us to balance that out,” rather than scoring with subjectivities that could lead to inflated/deflated or prefabricated ratings. Most supervisors suggested that more training and repeated calibrations are the best tools to implement T-TESS in the most objective manner possible. This sat in tension with a recognition that subjectivities could never fully be calibrated away.

Clinical Teachers' Perceptions of Impact

Although participants valued communication built into the PE/PG system, perceptions of T-TESS's capacity to capture the complexity of learning in the classroom was mixed. When surveyed, most CTs ($n=52/73$, 71.2%) agreed or strongly agreed that T-TESS accurately captured their impact on improving student motivation, attitudes, and engagement. However, some CTs ($n=30/73$, 41.1%) disagreed or strongly disagreed that T-TESS adequately takes into account the influence of student background and characteristics. In focus groups, CTs expressed concerns that T-TESS was not authentic, but rather “just one person's opinion” based on limited observations. CTs indicated that they “catered” to their supervisor, not students, during observations. Further, CTs noted that supervisors did not always understand the classroom environment set by the mentor teacher and were not always welcome in the classroom. One indicated, “I know my supervisor stresses out my mentor teacher, because he's very laid back and chill. He's like ‘Is she going to be in the classroom?’”

Subtheme: Perceived Impact on CT Professional Practice

T-TESS feedback served as a guide for CTs, informing them of their strengths and weaknesses. A CT explained that the discrete categories helped them understand this:

Getting my evaluation back and seeing, again, how it is all broken down into specific domains . . . it kind of explains how you are doing in each area of teaching... your supervisor's own words give you a better idea of what you have to work on and what you are doing well in the classroom.... It just shows where my strengths lie, and . . . where I could put more work in.

While CTs may find aspects of evaluations inauthentic and performative, they still indicated that they used the feedback they received to alter future teaching practices. One CT remarked, “It helped me not just guide for the next lesson I teach, but also guide for the next day.” In addition to recognizing strengths, CTs used the feedback and the rubric to improve upon identified weaknesses, citing changes to their teaching as evidence of improvement over the course of their clinical teaching experience. A CT noted: “I learned what my major weakness was. From my beginning lesson to my last lesson, and my rubric I got back, I've improved and I'm soaring a lot higher than I was at the very beginning.” While this does not show how CTs altered their teaching, it does indicate that they used feedback to identify and address weaknesses.

Despite a majority of CTs believing that T-TESS accurately captured their impact on improving student motivation, attitudes, and engagement, when asked on the survey about the overall impact of T-TESS on student achievement, most CTs believed that T-TESS had no real impact ($n=41/75$, 54.7%), while a sizeable minority believed there to be a generally positive impact ($n=33/75$, 44%). Only one CT ($n=1/75$, 1.3%) believed T-TESS had a generally negative impact on student achievement. The student who responded that there was a generally negative impact commented, “I think that our mentor teachers should have more of a say of how we are evaluated, being that they are with us more often and see more lessons.” Based on open-ended, qualitative survey data, like the response of this CT, some CTs appeared to have been confused on whether this question was asking about the students in their classrooms or their achievement as teacher education students.

Supervisors’ Perceptions of Impact

In response to questions on the survey and interview protocol, supervisors reported their perceptions of T-TESS’s impact on their own and CT’s professional practice as well as its impact on student achievement.

Subtheme: Perceived Impact on Supervisor Professional Practice

Based on survey data, the vast majority of supervisors ($n=17/21$, 81.0%) believed T-TESS had a generally positive impact on their own professional practice. No supervisors described the impact of T-TESS on their practice as negative in survey responses. In an interview, one supervisor explained in greater detail the ways T-TESS positively impacted their professional practice, making them “more aware of different things” they “really hadn’t considered before.” The result was that they were “thinking about some of those categories” for their faculty role. This supervisor perceived that “observ[ing] somebody else” and “talk[ing] to them about their practices” makes them “more reflective” about their “own practices.” For example, considering the question, “how would I feel if somebody said that to me?” helped them be “a better communicator.” This self-reflection “rolls down to how you talk with your own students and your own colleagues.” Comments like this support the state’s theory of action that posits that a model of quality teaching can cascade through systems to at least partially shape sensemaking around quality teaching, at least for supervisors.

Subtheme: Perceived Impact on CT Professional Practice

Supervisors generally found T-TESS to have a positive impact on CTs’ understanding of quality teaching, and ultimately, their practice. Based on their survey responses, the majority of supervisors ($n=16/21$, 76.2%) believed T-TESS had a generally positive impact on the professional practice of CTs. Most supervisors ($n=15/21$, 71.4%) agreed that T-TESS both provided clarity and focus on important aspects of good/effective teaching and created dialogue, communication, and discussion about good/effective teaching practices with their supervisors, mentor teachers, or school administrators. Many supervisors ($n=13/21$, 61.9%) reportedly do not believe CTs are focusing or narrowing their teaching to fit T-TESS. Accordingly, supervisors

utilized the rubric to center conversations around effective teaching to guide CTs in reflection to alter their teaching, in turn creating a positive impact on their practice.

As we saw from the CT data, the clinical experience inadvertently taught CTs how to navigate the evaluation system with inauthenticity. Supervisors acknowledged the performative aspect of evaluations, but sometimes understood this as a positive. One supervisor considered that “knowing how to be evaluated” is “really important” for teachers’ career-long professional practice. This supervisor argued that knowing “how [to] craft the lesson that hits all the little boxes” is what “allows [one] to continue teaching.” For example, she noted, one might include technology in an observed lesson even if it is not necessary. By this logic: “If your job depends on you doing well on this T-TESS, then you need to be able to perform well for the T-TESS.”

In summary, supervisors appeared to be of two minds regarding the value of T-TESS. Although they described T-TESS as an effective measure of teaching quality, they also viewed it as an instrument riddled with inconsistencies that is unable to capture the holistic quality of teaching. This tension is also coupled with acceptance that there will always be performative aspects to evaluations that undermine PE/PG models’ potential for creating a common understanding of what quality teaching is. Despite their concerns about T-TESS expressed in interviews, supervisors’ survey responses indicated most believed T-TESS had a generally positive impact on student achievement ($n=12/19$, 63.2%), while a minority perceived no real impact ($n=7/19$, 36.8%). No supervisors reported that T-TESS had a generally negative impact on student achievement. Only one supervisor left qualitative feedback for this survey question: “Generally clinical teachers do not understand the descriptors.” Like CTs, there may have been a few supervisors who had confusion on whether the question was asking about classroom students or CTs.

Discussion

According to the inferred state theory of action, T-TESS and its associated understandings will *cascade* through systems to create common understanding about quality teaching (see Figure 3). In this study, we found contradictory evidence as to whether participants perceived that T-TESS served effectively as a boundary object that facilitated shared sensemaking about quality teaching within or across the communities in this first year of use. CTs and supervisors in this study generally agreed that T-TESS presented a comprehensive model of quality teaching in the ideal. However, our findings indicated problems with implementation: Both sets of participants expressed that T-TESS had not been introduced prior to clinical experiences and was not used in all districts. The T-TESS model of quality teaching was introduced only briefly at the beginning of the CTs’ clinical semester. As a result, CTs worked with their supervisor, who had limited understanding of classroom context, and their mentor teacher, who often had limited understanding of T-TESS, to attempt quality teaching as represented by T-TESS. Frequently, the CTs perceived this T-TESS-driven teaching as inauthentic. At the same time, CTs appreciated supportive feedback regardless of T-TESS alignment.

As the T-TESS model was applied in practice, it was *mediated* by both supervisors and CTs in ways that caused participants, particularly CTs, to question its validity and reliability. CTs were of two minds about the role of T-TESS: on one hand, they held that there is an ideal objective

evaluation that can be achieved, and, on the other hand, they believed that supervisors must understand the context of the teaching environment to provide feedback that enables growth and development. Without this contextual understanding, supervisors' (mediated) understanding of quality teaching sometimes collided with expectations in classroom contexts. These collisions led CTs to dismiss feedback as "one person's opinion." This suggests failure of the model to shape sensemaking around a standardized, decontextualized definition of quality teaching.

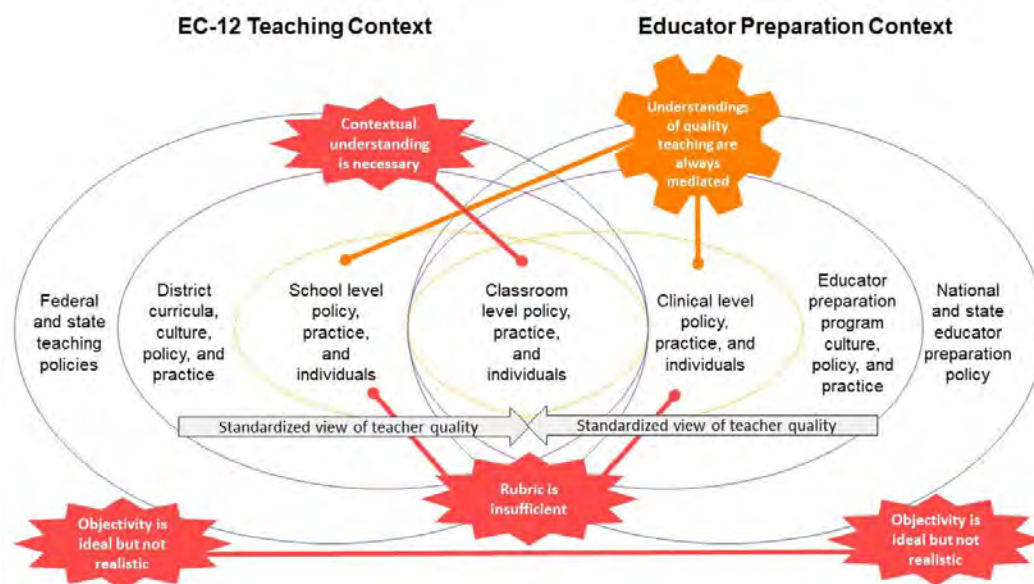
We posit that participants' conflicting beliefs derive from problems with the theory of action underlying T-TESS. Both supervisors and CTs believed that objectivity was ideal, particularly given the evaluative uses of T-TESS. However, when CTs spoke about "gaming" evaluations, they were performing to their supervisors' *mediated* understandings of quality teaching based on T-TESS. This conflict could explain the desire expressed by some CTs for an external evaluator: CTs expressed contradictory beliefs, indicating that supervisors would need to understand the classroom context to be objective, yet calling for multiple, external evaluators for T-TESS. Supervisors also acknowledged bias but suggested that they could be "cured" of subjectivity and trained to implement T-TESS objectively through continued calibrations. Yet, supervisors suggested this would require an unrealistic amount of training for them. Thus, participants' belief in objectivity contradicted the reality of the myriad ways T-TESS is necessarily mediated by supervisors and CTs.

Such tensions inherent in PE/PG models raise questions about the feasibility of the state's role in shaping teacher understanding while holding them accountable for teaching practices. Scholars have noted fatal flaws in the logic of PE/PG models, which are posited in policy to foster professional learning through building relationships, mentorship/coaching, and communication while also defining and measuring quality teaching for practitioners across systems (see, for example, King & Paufler, 2020; Mette et al., 2017). In EC-12 schools, Mette et al. (2017) noted tension "between the desired collaborative, trusting relationship and conflicting functions when the supervisor is also an administrator" (p. 710). In this study, when CTs "gamed" evaluations, they understood that they would not receive feedback conducive to growth. They were compelled to performativity because their future as a teacher depended on scoring well. Regardless of whether CTs "game" the evaluation to the subjectivities of their supervisor or the T-TESS rubric itself, "teachers are no longer encouraged to have a rationale for practice [or] account of themselves in terms of a relationship to the meaningfulness of what they do" (Ball, 2003, p. 222). Instead, teachers perform to what earns them the highest scores on their evaluation. In this study, CTs generally recognized that they were novices and desired authentic feedback, coaching, and mentoring. Paradoxically, many of these CTs held the belief that inauthenticity and performativity are inherent to high-stakes evaluations (Ball, 2003; Burns & Badiali, 2015; Hazi, 2018; Holloway & Brass, 2018).

The expectation of objectivity presents another tension inherent to the conflating of evaluation and support for growth. In complex systems, actors at a site necessarily will mediate artifacts to align with their beliefs and practices; if a policy dictates the use of an instrument, mediation and collisions will occur as different actors within and across communities attempt to make sense around it. Communities of practice are formed through this mediation as meanings are negotiated and understandings are built (Canipe & Gunkel, 2020). Thus, a boundary object cannot be inserted into a complex system and be assumed to direct learning and behavior within and across

communities of practice in a controlled and linear manner. Expert cognition in complex systems requires that “practitioners must move across boundaries to seek and give help, to find information and tools wherever they happen to be available” (Engeström et al., 1995, p. 332).

Figure 4. *Collisions and Mediations of T-TESS Across Systems*



Note. Red starbursts indicate collisions. Orange gear indicates mediation. Adapted from “Cascading, Colliding, and Mediating: How Teacher Preparation and K-12 Education Contexts Influence Mentor Teachers’ Work,” by R. Roegman and J. Kolman, 2020, *Journal of Teacher Education*, 71(1), p. 109 (<https://doi.org/10.1177/0022487119850174>). Copyright 2019 by the American Association of Colleges for Teacher Education.

Participant perceptions suggest that the PE/PG model served poorly as a boundary object within or across communities; the model did not seem to contribute to a common understanding of quality teaching. Rather, participants experienced *collisions* between the notion of objective, standardized definitions of quality teaching as defined by T-TESS and contextually situated understandings (see Figure 4). Building and navigating relationships between members of various communities of practice is complex (Martin et al., 2011). However, it is essential for developing common understandings and practices. As Martin et al. (2011) note, “redefining the nature of university-based teacher educators’ work adds further dimension to its complexity as processes of building and negotiating complex relationships are central to the work” (p. 308). Our research suggests the need for a better understanding of communities of practice and the role of PE/PG models as objects within and across boundaries. This requires (re)theorizing the role of growth-based evaluation in social learning and developing systemic organizational support for teachers across settings and career stages.

Conclusion

Policymakers may have adopted elements of social learning in their quest to improve teaching. However, despite recognizing the social nature of professional learning, these policies do not account for the complexity of systems. Rather, policymakers assume that systems operate mechanically, and interventions lead to linear, predictable results. In systems so conceived, “managers want workers to respond predictably to incentives and to accomplish goals defined by managers and to do this with little deviation” (McDaniel, 2007, p. 21). This is a focus on command, control, and planning (McDaniel, 2007). In the case of PE/PG models, rubrics outline ideal expectations for quality teaching, while evaluation of teaching controls outcomes—both in terms of teacher behavior and student achievement. Thus, in implementing common PE/PG models, policymakers address educational systems as “machines” rather than as “self-organizing systems, in which order emerges in a bottom-up fashion from the local relationships in which they are involved” (Ritter et al., 2004, p. 175).

The PE/PG policy theory of action assumes education is a system that can be commanded and controlled in a linear fashion. Our research suggests PE/PG models cannot simply be injected as boundary objects to enforce a common understanding of quality teaching across and within systems. In this study, both the evaluation and supervision functions of T-TESS were undermined by inevitable mediations and collisions. While there is considerable research suggesting the conflation of evaluation and supervision (e.g., Paufler et al., 2020a, 2020b; Burns & Badiali, 2015; Hazi, 2018; Mette et al., 2020), the state’s theory of action presupposes that the two can function simultaneously. Thus, we identify two faulty premises: the conflation of evaluation and supervision, as well as the attempts to eliminate mediations and collisions through better calibration, which perpetuates the faulty logic of command/control/plan.

Scholars of complex adaptive systems (e.g., Eoyang, 2006; McDaniel, 2007; Rowland, 2007) offer alternatives for managing human systems. McDaniel (2007), for example, proposes replacing the managerial goals of command, control, and planning with facilitating sensemaking, learning, and improvisation. Conceptualizing educational communities as complex adaptive systems necessitates alternative models of teacher development. Such new models could focus on sensemaking, learning, and improvisation. As members of multiple, overlapping communities of practice directly influenced by PE/PG models, educators are best situated to develop understandings of quality teaching that are context-sensitive and responsive to the dynamic, complex interactions of differently situated individuals across diverse educational settings.

References

- AACTE Clinical Practice Commission. (2018). *A pivot toward clinical practice, its lexicon, and the renewal of educator preparation* [Report]. American Association of Colleges for Teacher Education. <http://www.nysed.gov/common/nysed/files/cpc-aactecpcreport.pdf>
- Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: critical perspectives on tests and assessment-based accountability*. Routledge.
- Anagnostopoulos, D., Smith, E. R., & Basmadjian, K. G. (2007). Bridging the university school divide horizontal expertise and the “two-worlds pitfall.” *Journal of Teacher Education*, 58(2), 138-152. <https://doi.org/10.1177/0022487106297841>
- Ball, S. J. (2003). The teacher’s soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215–228. <https://doi.org/10.1080/0268093022000043065>
- Bates, A. J., Ramirez, L., & Dritz, D. (2009). Connecting university supervision and critical reflection: Mentoring and modeling. *The Teacher Educator*, 44(2), 90-112. <https://doi.org/10.1080/08878730902751993>
- Bowen, G. A. (2006). Grounded theory and sensitizing concepts. *International Journal of Qualitative Methods*, 5(3), 12-23. <https://doi.org/10.1177/160940690600500304>
- Brandon, J., & Derrington, M. L. (2019). Supporting teacher growth and assuring teacher quality. In M. L. Derrington & J. Brandon (Eds.), *Differentiated teacher evaluation and professional learning: Policies and practices for promoting career growth* (pp. 3-14). Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-16454-6>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589-597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Byrne, D. (2021). A worked example of Braun and Clarke’s approach to reflexive thematic analysis. *Quality & Quantity*. <https://doi.org/10.1007/s11135-021-01182-y>
- Burns, R., & Badiali, B. J. (2015). When supervision is conflated with evaluation: Teacher candidates’ perceptions of their novice supervisor. *Action in Teacher Education*, 37, 418-437. <https://doi.org/10.1080/01626620.2015.1078757>
- Burns, R., Jacobs, J., & Yendol-Hoppey, D. (2016). Preservice teacher supervision within field experiences in a decade of reform. *Teacher Education and Practice*, 29(1), 46-75.
- Canipe, M. M., & Gunckel, K. L. (2020). Imagination, brokers, and boundary objects: Interrupting the mentor-preservice teacher hierarchy when negotiating meanings. *Journal of Teacher Education*, 71(1), 80-93. <https://doi.org/10.1177/0022487119840660>
- Charmaz, K. (2003). Grounded theory: Objectivist and constructivist methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies for qualitative inquiry* (2nd ed., pp. 249-291). Sage.
- Cochran-Smith, M., Ell, F., Grudnoff, L., Ludlow, L., Haigh, M., & Hill, M. (2014). When complexity theory meets critical realism: A platform for research on initial teacher education. *Teacher Education Quarterly*, 41(1), 105-122.
- Cuevas, R., Ntoumanis, N., Fernandez-Bustos, J. G., & Bartholomew, K. (2018). Does teacher evaluation based on student performance predict motivation, well-being, and ill-being? *Journal of School Psychology*, 68, 154–162. <https://doi.org/10.1016/j.jsp.2018.03.005>
- Danielson, C. (2013). *The framework for teaching evaluation instrument*. The Danielson Group. <https://danielsongroup.org/products/product/framework-teaching-evaluation-instrument>

- Derrington, M. L., & Campbell, J. W. (2015). Implementing new teacher evaluation systems: Principals' concerns and supervisor support. *Journal of Educational Change*, 16(3), 305-326. <https://doi.org/10.1007/s10833-015-9244-6>
- Dinsmore, J., & Wenger, K. (2006). Relationships in preservice teacher preparation: From cohorts to communities. *Teacher Education Quarterly*, 33(1), 57-74.
- Engeström, Y., Engeström, R., & Kärkkäinen, M. (1995). Polycontextuality and boundary crossing in expert cognition: Learning and problem solving in complex work activities. *Learning and Instruction*, 5(4), 319-336. [https://doi.org/10.1016/0959-4752\(95\)00021-6](https://doi.org/10.1016/0959-4752(95)00021-6)
- Eoyang, G. H. (2006). Human systems dynamics: Complexity-based approach to a complex evaluation. In B. Williams & I. Imam (Eds.), *Systems concepts in evaluation: An expert anthology* (pp. 123-139). EdgePress.
- Hazi, H. (2018). Coming to understand the wicked problem of teacher evaluation. In S. J. Zepeda & J. A. Ponticell (Eds.), *The Wiley handbook of educational supervision* (pp. 183-207). John Wiley & Sons.
- Holloway, J., & Brass, J. (2018). Making accountable teachers: The terrors and pleasures of performativity. *Journal of Education Policy*, 33(3), 361-382. <https://doi.org/10.1080/02680939.2017.1372636>
- Holt, R., & Cornelissen, J. (2014). Sensemaking revisited. *Management Learning*, 45(5), 525-539. <https://doi.org/10.1177/1350507613486422>
- Hopkins, E., Hendry, H., Garrod, F., McClare, S., Pettit, D., Smith, L., Burrell, H., & Temple, J. (2016). Teachers' views of the impact of school evaluation and external inspection processes. *Improving Schools*, 19(1), 52-61. <https://doi.org/10.1177/1365480215627894>
- King, K. M., & Paufler, N. A. (2020). Excavating theory in teacher evaluation: Implementing evaluation frameworks as Wengerian boundary objects. *Education Policy Analysis Archives*, 28(57). <https://doi.org/10.14507/epaa.28.5020>
- Lavigne, A. L., & Good, T. L. (2019). *Enhancing teacher education, development, and evaluation: Lessons learned from educational reform*. Routledge.
- Levine, T. H. (2010). Tools for the study and design of collaborative teacher learning: The affordances of different conceptions of teacher community and activity theory. *Teacher Education Quarterly*, 37(1), 109-130.
- Liu, S., & Zhao, D. (2013). Teacher evaluation in China: Latest trends and future directions. *Educational Assessment, Evaluation and Accountability*, 25(3), 231-250. <https://doi.org/10.1007/s11092-013-9168-8>
- Martin, S. D., Snow, J., & Franklin Torrez, C. A. (2011). Navigating the terrain of third space: Tensions with/in relationships in school-university partnerships. *Journal of Teacher Education*, 62(3), 299-311. <https://doi.org/10.1177/0022487110396096>
- McDaniel, Jr., R. R. (2007). Management strategies for complex adaptive systems sensemaking, learning, and improvisation. *Performance Improvement Quarterly*, 20(2), 21-41. <https://doi.org/10.1111/j.1937-8327.2007.tb00438.x>
- Mette, I. M. (2019). The state of supervision discourse communities: A call for the future of supervision to shed its mask. *Journal of Educational Supervision*, 2(2), 1-10. <https://doi.org/10.31045/jes.2.2.1>
- Mette, I. M., Aguilar, I., & Wieczorek, D. (2020). A thirty state analysis of teacher supervision and evaluation systems in the ESSA era. *Journal of Educational Supervision*, 3(2), 105-135. <https://doi.org/10.31045/jes.3.2.7>

- Mette, I. M., Fairman, J. C., Lech, P. L., & Frankland, M. C. (2019). *Principal and teacher perceptions of performance evaluation and professional growth (PE/PG) system implementation*. Maine Education Policy Research Institute.
- 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. <https://doi.org/10.1177/1555458918759696>
- Montecinos, C., Walker, H., & Maldonado, F. (2015). School administrators and university practicum supervisors as boundary brokers for initial teacher education in Chile. *Teaching and Teacher Education*, 49, 1-10. <http://doi.org/10.1016/j.tate.2015.02.011>
- Nolan, J., & Hoover, L. A. (2010). *Teacher supervision and evaluation: Theory into practice* (3rd ed.). John Wiley & Sons.
- Onwuegbuzie, A. J., & Collins, K. M. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, 12(2), 281-316.
- Palmeri, A. B., & Peter, J. A. Conflated constructs: Disentangling the educative and evaluative functions of preservice teacher supervision. *Journal of Educational Supervision*, 2(2). <https://doi.org/10.31045/jes.2.2.5>
- Patton, K., & Parker, M. (2017). Teacher education communities of practice: More than a culture of collaboration. *Teaching and Teacher Education*, 67, 351-360. <http://doi.org/10.1016/j.tate.2017.06.013>
- Paufler, N. A. (2014). *Investigating a teacher evaluation system: School administrator and teacher perceptions of the system's standards of effectiveness* [Doctoral dissertation, Arizona State University]. Arizona State University KEEP. <https://keep.lib.asu.edu/items/153078>
- Paufler, N. A., King, K. M., & Zhu, P. (2020a). Delivering on the promise of support for growth? Evaluator perceptions of a new state teacher evaluation system. *Journal of Educational Supervision*, 3(2). <https://doi.org/10.31045/jes.3.2.3>
- Paufler, N. A., King, K. M., & Zhu, P. (2020b). Promoting professional growth in new teacher evaluation systems: Practitioners' lived experiences in changing policy contexts. *Studies in Educational Evaluation*, 65. <https://doi.org/10.1016/j.stueduc.2020.100873>
- Paufler, N. A., & Sloat, E. F. (2019). Using standards to evaluate accountability policy in context: School administrator and teacher perceptions of a teacher evaluation system. *Studies in Educational Evaluation*, 64, 1-15. <https://doi.org/10.1016/j.stueduc.2019.07.007>
- Robertson-Kraft, C., & Zhang, R. S. (2018). Keeping great teachers: A case study on the impact and implementation of a pilot teacher evaluation system. *Educational Policy*, 32(3), 363-394. <https://doi.org/10.1177/08959048186637685>
- Roegman, R., & Kolman, J. (2020). Cascading, colliding, and mediating: How teacher preparation and K-12 education contexts influence mentor teachers' work. *Journal of Teacher Education*. 71(1), 108-121. <https://doi.org/10.1177/0022487119850174>
- Rowland, G. (2007). Performance improvement assuming complexity. *Performance Improvement Quarterly*, 20(2), 117-136. <https://doi.org/10.1111/j.1937-8327.2007.tb00444.x>

- Ritter, T., Wilkinson, I. F., & Johnston, W. J. (2004). Managing in complex business networks. *Industrial Marketing Management*, 33(3), 175–183. <https://doi.org/10.1016/j.indmarman.2003.10.016>
- Stake, R. E., & Trumbull, D. (1982). Naturalistic generalizations. *Review Journal of Philosophy and Social Science*, 7, 1–12.
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, ‘translations’ and boundary objects: Amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, 19(3), 387–420. <https://doi.org/10.1177/030631289019003001>
- Taut, S., & Sun, Y. (2014). The development and implementation of a national, standards-based, multi-method teacher performance assessment system in Chile. *Education Policy Analysis Archives*, 22(71). <https://doi.org/10.14507/epaa.v22n71.2014>
- Texas Education Agency (TEA). (2016). *Texas Teacher Evaluation & Support System rubric*. https://www.teachfortexas.org/Resource_Files/Guides/T-TESS_Rubric.pdf
- TEA. (2019). *Texas Teacher Evaluation & Support System*. <https://www.teachfortexas.org/>
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing the process of sensemaking. *Organization Science*, 16(4), 409–421. <https://doi.org/10.1287/orsc.1050.0133>
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.
- Wenger, E. (2010). Communities of practice and social learning systems: The career of a concept. In C. Blackmore (Ed.), *Social learning systems and communities of practice* (pp. 179–198). Springer-Verlag London.
- Wenger, E., Trayner, B., & de Laat, M. (2011). *Promoting and assessing value creation in communities and networks: A conceptual framework*. Ruud de Moor Centrum.
- Yin, R. K. (2018). *Case study research and applications: Design and methods*. Sage.

Author Biographies

Kelley M. King is an Associate Professor of Curriculum and Instruction and Assistant Department Chair for Initial Certification Programs at the University of North Texas. Her research interests include the history and politics of schooling and education reform, with particular focus on the lives of teachers and education for democratic citizenship. Prior to coming to UNT, she worked for ten years as a public school teacher in Texas.

Noelle A. Paufler is an Assistant Professor – P-12 and the Program Coordinator for the EdD in Education Systems Improvement Science at Clemson University. She has experience as a high school social studies teacher, district administrator, and applied researcher in high-need districts and schools. Her research interests include K-12 educational policy, specifically how educational leaders enact accountability policy into practice and its impact on teachers and students.

Rachel L. Biritz is a PhD student in the Educational Leadership, P-12 program at Clemson University in the Department of Educational and Organizational Leadership Development in the College of Education. Her current roles as Graduate Research Assistant and University Supervisor in Clemson University’s College of Education, as well as her P-12 teaching background, allow her to research and influence education law, policy, and reform.

Ryan M. Smits is a PhD student at the University of North Texas in the College of Education. His research interests are at the intersection of education technology, policy, and learning environments.