

2022

## The Restrictive Concepts of Teacher Evaluation and their Discourse Communities

Helen M. Hazi

West Virginia University, hmhazi@verizon.net

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



Part of the [Educational Administration and Supervision Commons](#)

---

### Recommended Citation

Hazi, H. M. (2022). The Restrictive Concepts of Teacher Evaluation and their Discourse Communities. *Journal of Educational Supervision*, 5 (1). <https://doi.org/10.31045/jes.5.1.3>

This Conceptual 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](mailto:um.library.technical.services@maine.edu).

# The Restrictive Concepts of Teacher Evaluation and their Discourse Communities

Journal of Educational Supervision

44 – 67

Volume 5, Issue 1, 2022

DOI: <https://doi.org/10.31045/jes.5.1.3>

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

**Helen M. Hazi<sup>1</sup>**

## **Abstract**

Teacher evaluation is a personnel practice in education and a field of study with multiple discourse communities, including a community in supervision. Many concepts from these discourse communities have influenced practice over time. In this article, I place teacher evaluation as a practice originating in the intersection of supervision and administration, describe its tumultuous relationship with supervision, and identify the many concepts that restrict its practice with examples of scholarship. This article is important “fieldwork” that scholars must periodically conduct on their niche to better understand its audiences, scope, and influences.<sup>2</sup> Examples of fieldwork include: analyses of scholarship, collections on a theme, handbooks, histories, interviews of scholars, memoirs, and genealogies of scholars. This article is a type of analysis of scholarship for the field of teacher evaluation, as Bolin (1988), Blumberg (1990), Garman (2020), Glanz (2018), Gordon (2019), Mette (2019), and Glanz and Hazi (2019) have done for the field of supervision. I am often asked, “Is there anything left to study about teacher evaluation?” This article may help scholars broaden their thinking about the many discourses of teacher evaluation as well as their own niche. Understanding the nature of the discourses helps a scholar navigate its writings and research, situate his/her contribution, and interpret the results of emerging research.

## **Keywords**

teacher evaluation; supervision; field of study; discourse communities; restrictive concepts

---

<sup>1</sup> West Virginia University

<sup>2</sup> This article is based on an earlier paper (Hazi, 2012) and is dedicated to the many scholars who have broadened my thinking. While these understandings and insights about teacher evaluation are uniquely mine, they are mutable. I hope readers may find themselves and some of their thinking herein and continue to contribute to the discourses of teacher evaluation.

## **Corresponding Author:**

Helen M. Hazi (Department of Curriculum & Instruction/Literacy Studies, College of Education and Human Services, West Virginia University, 608G Allen Hall, Morgantown, WV 26506, USA)  
email: [hmhazi@verizon.net](mailto:hmhazi@verizon.net)

## Introduction

Personal history is “virtuous,” according to Peshkin (1988), influencing how I define and view teacher evaluation and its relationship with other fields. It is disclosed to show how it shapes my thinking, research, and contribution. I grew up in a household with a father who believed unions fought for its workers and advised me never to cross a picket line. I was certified and hired first as a teacher and then as a school supervisor in a northeastern state that had collective bargaining. Knowing the importance of relationships with teachers and the limits of a quasi-administrator, I made my hiring as a supervisor contingent upon not being responsible for evaluation. The superintendent agreed, until, he said, a teacher was to be dismissed. Fortunately, I did not have to face that challenge. When I finished my doctorate with a dissertation about collective bargaining contracts, I took a job as a professor in Educational Administration. There I studied both teacher evaluation and supervision, recognizing their relationship. Along the way, I studied law and became an expert witness which contributed to my understanding of both the theory and practice of teacher evaluation and supervision.

Based on this history and four decades of readings, research, and experiences with teachers, I started a list of teacher evaluation concepts that were likely to be familiar to school practitioners. A concept was included if it was used in the development, delivery, or use of instruments to evaluate teachers, and if it appeared in litigation, collective bargaining contracts, training, or local policy. While the reader may view these concepts as jargon “beyond the reach of an uninitiated audience” (Becher & Trowler, 2001, p. 46), they are most likely to be common to practitioners. For example, the term “goal setting,” used in the post-conference, was included, but “fidelity of implementation” was omitted because it is not used in practice or school policy. Another example is “script taping,” a form of notetaking as an alternative to video or audiotaping, that was once characterized as an essential supervisory tool (Hunter, 1983). However, it was short-lived in practice, did not appear in law or policy, and thus, was omitted.

Concepts were then grouped by sub-discipline of origin, defined, and situated by their use in evaluation, i.e., purpose, instrument, evaluator, or procedure. While a concept like “feedback” can belong to multiple disciplines, the most relevant was selected for its disciplinary origin and defined in plain language using sources such as *The Oxford English Dictionary*, *Wikipedia*, *Oxford Scholarship Online*, *Oxford Research Encyclopedias*, or *Online Etymology Dictionary*. The sub-disciplines within education were identified and defined by drawing upon sources such as Furlong and Lawn’s (2011) *Disciplines of Education*, Furlong’s (2013) *Education---An Anatomy of the Discipline*, and Donaldson’s (2021) *Multidisciplinary Perspective on Teacher Evaluation*.

This is a reflective essay from a scholar who has spent a career pondering teacher evaluation, supervision, and their relationship. In this essay, I argue that many concepts restrict the practice of teacher evaluation. Its concepts represent a “bedrock,” making it impervious to change, and helping to explain how its summative purpose overshadows its formative. I hope to help scholars navigate the underpinnings of its practice and situate his/her contribution in the discourses. In this article, I situate its origins at the intersection of supervision and administration, explicate the many discourses of this field of study, illustrate these discourses with examples from in and outside of supervision, and conclude with insights.

## **Teacher Evaluation at the Intersection of Supervision and Administration**

Teacher evaluation is the annual rating of teachers to determine personnel decisions. Supervision scholars today tend to see supervision as the helping (formative) function and evaluation as the annual rating (summative), where one overshadows the other (McGhee, 2020; Mette et al., 2020). I tend to see the two as fraternal twins – similar, yet not identical – both with the purpose of improving instruction, requiring evidence from the classroom, and involving judgment; yet different in their practice, dilemmas, and scholarship.

In the early history of supervision, these two functions were vested in different people. Male principals, who were generalists, conducted personnel evaluations, while female supervisors, who were content specialists, helped teachers through lesson planning and demonstrations in classrooms. In the 1930s as schools expanded and the depression affected the funding of schools, female supervisors were let go and the rating and helping functions were combined into the role of principal (Glanz & Hazi, 2019).

As principals took over both functions, and its scholarship evolved, teacher evaluation became a personnel practice in educational administration. As the role of supervisor continued in some larger school systems, teacher evaluation as improvement of instruction was a studied practice of supervision in the early days of the Council of Professors of Instructional Supervision (COPIS) where educational administration scholars dominated. Scholars such as Harris (1978), McQuarrie and Wood (1991), and McGreal (1983) wrote about teacher evaluation. However, important events occurred in the 1990s that chronicle its tumultuous relationship with supervision.

Those in COPIS argued that the field should abolish supervision (Gordon, 1992; Sergiovanni, 1992; Starratt, 1992) as it became a vehicle of accountability (Garman, 2020). Many scholars in supervision shunned teacher evaluation viewing it as a “contaminant” of the field (Garman, 2010). Some called to abolish supervision because it was tainted by the practice of evaluation (Starratt, 1997) and moved on to more “palatable” topics such as staff development, mentoring, teacher leadership and collaborative teaming (Holland & Garman, 2001).

Debates among scholars were also reflected in practice. My case study of grievances filed against a school supervisor best illustrates the dilemma. This supervisor was responsible for both evaluating and helping teachers to improve. In her mind and practice, the supervisor knew when she did each, and tried to explain their differences in orientation meetings and classroom visits. Instead, the teachers’ association saw her actions as harassment and intimidation, claiming “that any time a supervisor is in the classroom it is an evaluation” (Hazi, 1994, p. 208) because evaluation and supervision were (and still are) entangled in law and practice (Ponticell & Zepeda, 2004). This sentiment was echoed in actions by the leadership of Association for Supervision and Curriculum Development (ASCD) who wanted to better attract teachers as members and to eliminate the term supervision which had become toxic to practitioners (Glanz & Hazi, 2019).

Also, around this time, Ron Brandt, then editor of *Educational Leadership*, with Charlotte Danielson approached COPIS to endorse her Framework for Teaching. The Framework was “not well received” and COPIS “urge[d] ASCD to distance itself from any association with this

specific teacher assessment framework” because it was “dated and behind in current research,” that “its research base was not evident,” that it “show[ed] little promise in guiding teacher improvement,” and that it focused on “individual teacher accountability at a time [of] ... [teacher] empowerment” (Hazi, 2016; Killian, 1994, p.1). After a decade of debate, ASCD ironically eliminated supervision from its title and mission (Glanz & Hazi, 2019), yet promoted Charlotte Danielson’s teacher evaluation framework in publications and workshops.

### The Many Discourse Communities of Teacher Evaluation

Teacher evaluation is a field of study within the discipline of education.<sup>3</sup> A *field of study*, according to Pinar et al. (1995) is “a tradition of language or discourse” (p. 7) that develops over time. Fields of study are not fixed edifices, but fluid systems that are informed by theories from many disciplines (Lawn & Furlong, 2011). Boundaries are crossed and knowledge is poached, especially where problems are considered complex and difficult to solve (Becher & Trowler, 2001). Scholars cross borders for concepts and language to study, understand, and explain beyond the superficial.

A field of study may have one or more discourse communities. A *discourse community* is a “grouping of people who share common language, norms, characteristics, patterns, or practices as a consequence of their ongoing communications and identification with each other” (Bazerman, 2009, para 1). Borg (2003) explains that in a discourse community scholars use fora (conferences, newsletters, articles, emails) to pursue their common goals. Scholars may participate within or across discourses. While size is less important, stability is “with experts who perform gatekeeping roles” and novices enter to renew the community over time (Borg, 2003, p. 3).

Teacher evaluation has many discourse communities in education.<sup>4</sup> The discourse communities have contributed concepts that influence its practice and make it difficult to change.<sup>5</sup> Table 1 presents the many concepts, defined, and their use in evaluation, organized by discourse. This is one way I see them and think about their influence on practice.

---

<sup>3</sup> A *discipline* is “a branch of learning or knowledge; a field of study or expertise; a subject” (Discipline, 2021). Education has been called a branch of social science, a discipline, and a university-based field of study. Some say education is the second largest discipline (Furlong, 2013). There are no formal criteria for defining a discipline nor how one should be classified (e.g., whether psychology is a subfield of education or whether education is a sub-field of psychology (“Outline of academic disciplines,” 2021).

<sup>4</sup> The discourse communities are limited to education, and do not include those such as 360° Feedback (Church et al., 2019) that addresses performance assessment in other organizational settings, situated in Industrial-Organizational Psychology. Ironically, there has been debate about whether 360° feedback should be developmental only (formative) or used for organizational decision making (summative) (Bracken et al, 2016).

<sup>5</sup> While this article draws from scholarship, it does not synthesize it or research. For these, readers are directed to seminal works such as Darling-Hammond et al. (1983), Wise et al. (1984), Millman & Darling-Hammond (1990), Kennedy (2010), Murphy et al. (2013), and most recently Donaldson (2021).

Table 1: Teacher Evaluation Concepts with Definition, Use, and Discourse

<b>Concept</b>	<b>Definition</b>	<b>Use in Evaluation</b>	<b>Discourse of Origin</b>
<i>Formative &amp; summative</i>	Purposes of evaluation	Purposes	Psych: Evaluation
<i>Generic &amp; particular</i>	Common & unique teacher duties	Instrument (items)	Psych: Evaluation
<i>Multiple methods &amp; sources of data</i>	More methods to increase data & improve judgment	Procedure	Psych: Evaluation
<i>Accuracy vs. utility</i>	Evaluate should be useful to teachers	Purposes	Psych: Evaluation
<i>Valid</i>	Content of instrument	Instrument	Psychometrics
<i>Reliable</i>	Principal accuracy	Evaluator	Psychometrics
<i>Research-based</i>	One basis of instrument items	Instrument (items)	Psychometrics
<i>Rubric</i>	Detailed guide for scoring	Instrument (scoring)	Psychometrics
<i>Inter-rater reliability</i>	Training to improve observer accuracy	Training	Psychometrics
<i>Bias</i>	Prejudice to be avoided	Evaluator	Psychometrics
<i>Inflated rating</i>	Too high rating by principal	Instrument scoring	Psychometrics
<i>Observer drift</i>	Moving away from coding accuracy	Evaluator	Psychometrics
<i>(Re)calibration</i>	(Re)training in the instrument to increase accuracy and avoid drift	(Re)training	Psychometrics
<i>Walkthrough</i>	Time saver for observation	Procedure	Psych: Organizational
<i>Goal-setting</i>	Conference goal for teacher change	Procedure	Psych: Organizational
<i>Feedback</i>	Principal delivers to change teacher	Evaluator delivers	Psych: Developmental
<i>Reflection</i>	Conference goal for teacher change	Procedure	Psych: Developmental
<i>Arbitrary &amp; capricious</i>	Label for not following procedure	Procedure	Labor Law
<i>Due process</i>	Notice & hearing before dismissal	Procedure	Labor Law
<i>Evidence</i>	Documentation	Procedure	Labor law
<i>Remediation</i>	Opportunity for improvement	Procedure	Labor Law
<i>Uniform Procedures</i>	Procedure to avoid discrimination	Procedure	Labor Law
<i>Value-added</i>	Student gains/loss in testing	Procedure (scoring)	Econometrics
<i>Preconference</i>	Meeting before observation	Procedure	Supervision
<i>Portfolio</i>	A file or folder of work	Procedure	Teacher Education

The recent decade's emphasis on teacher quality in the U.S. has only strengthened their hold on practice.<sup>6</sup> These concepts, from eight identifiable discourses, make it difficult to change its practice since many appear in federal and state law, negotiated in collective bargaining agreements, or established in local policy through the vote of elected officials. They appear in italics in the text that follows, are defined in simple terms, and organized by discourse starting with the one with the most number of concepts.

## **Educational Psychology**

Psychology contributes the most number of concepts influencing teacher evaluation and dominates its practice. Within psychology are the discourses of evaluation, psychometrics, organizational psychology, and developmental psychology. The discourse of evaluation provides the overarching principles of practice; psychometrics guide instrument design, measurement, and training; organizational psychology situates evaluation in its context; and developmental psychology focuses on behavior change.

### ***Evaluation***

This discourse provides foundational principles of evaluation with scholars such as Scriven, Stake and Stufflebeam with Scriven being the most influential to personnel evaluation. The *formative* purpose of teacher evaluation is usually associated with helping and improving teachers, while the *summative* is the end of year rating for personnel decisions (Scriven, 1967). Many believe that principals can both help and evaluate teachers (e.g. McGhee, 2020). However, Scriven, who introduced the terms, never meant for the two to be done by the same individual. In fact, Scriven (1996) believed that the evaluator should have distance from the teacher to maintain objectivity.

Scriven believed that teachers had duties in common, but also had specific duties due to their subject, grade, or specialty. Therefore, Scriven recommended that both *generic and particularized duties* be evaluated (Stufflebeam, 2013). These duties should be evaluated with *multiple methods and sources of data*, rather than one method, so that data provide a more comprehensive and rigorous view of the teacher (Scriven, 1994).

*Utility* is the final evaluation principle and is important to understanding one of its failures. Scriven observed that accuracy of personnel rating is emphasized in instrument use at the expense of usefulness. Most teachers do not find feedback useful (e.g., Lochmiller, 2019). If the evaluation has no use to the recipient, then an organization is producing information, and not

---

<sup>6</sup> I believe that teacher evaluation remains seriously flawed and dysfunctional for many reasons: it is not reliable or valid, or cost-effective to make them so (Berliner, 2018). Before the national focus on teacher evaluation as a vehicle to improve teacher quality, there was limited evidence that it led to instructional improvement (Donaldson, 2009) or improve employee performance or company effectiveness (Mueller-Hanson, 2021). This has not changed (Donaldson, 2021). Second, many of our teachers have experienced teaching in an accountability climate of standards, incentives, and sanctions, first as public school students themselves, then teacher education students, and now teacher employees. Thus, teachers have become normalized to accountability and many of its high-stakes consequences (Holloway & Brass, 2018). This may explain why “today’s teachers may view teacher evaluation more favorably than their counterparts in prior decades” (Donaldson, 2021, p. 160).

feedback (Bracken, et al., 2016). While evaluation delivers to some degree on its summative purpose, it rarely succeeds on its formative.<sup>7</sup>

### ***Psychometrics***

Psychometrics is theory and techniques of measurement. While instruments are the most visible influence on practice, they come and go, reflecting what is most valued about teaching at the time (Kennedy, 2010). However, the psychometric concepts about instrument design, measurement, and training endure to influence principal practice and training. Concepts for instrument design include: valid, reliable, research-based and rubric. Concepts for principal training include: bias, observer drift, inflated rating, inter-rater reliability, and recalibration.

When an instrument's content reflects what teachers do, it is said to be *valid*. When an instrument's content comes from research, it is called *research-based*. A *rubric* is a guide for scoring or grading and was most used in the field of evaluation to grade student compositions (Scriven, 1991b). Because of its details and levels, some believe that rubrics help clarify expected performance (Brookhart, 2013). When the principal accurately and consistently uses the instrument to evaluate teaching, it is considered *reliable*.

Principals can be *biased* or have prejudice, resulting in errors that can adversely affect teachers (Scriven, 1991b). Thus, principals are trained and re-trained so they do not *drift* away from the meaning of instrument items or *inflate their ratings* by being too generous or lenient. Retraining principals on the instrument is *recalibration* so that they score according to the standard established (Scriven, 1991b).

### ***Organizational Psychology***

Organizational psychology is concerned with workplace performance and attitudes such as motivation and job satisfaction. The walkthrough and goal setting are two of its concepts. The classroom *walkthrough* was used to encourage principals to get into classrooms briefly and often. It originated with Sam Walton of Walmart fame and made popular in education by Downey et al. (2004). This was how principals were to emphasize classroom teaching and devote more time with teachers on curriculum and instructional problems (Fraser et al., 1999).

*Goal setting* is considered a practice for teacher improvement and recommended as a culminating activity of the post-observation conference (McGreal, 1983). It is based on the assumptions that the teacher would find principal feedback useful, agree to the prescriptions to fix practice, and change their behavior by the next observation (Hazi, 2019a).

---

<sup>7</sup> School practitioners have followed some of Scriven's principles more than others. For example, while Scriven recommended the formative be separated from the summative purposes, they are usually combined in the principal. Teachers have been evaluated generically, based on what they have in common, not the particularized, because it benefits the credibility of the principal who lacks content knowledge. Observation by the principal has dominated, until the principle of multiple methods was rediscovered and used in the Gates' Measures of Effective Teaching project, and then established in many state statutes and regulations due to Race to the Top (Hazi, 2019a).

### ***Developmental Psychology***

Developmental psychology is concerned with learning and behavior change. Feedback is a term originating in the 20<sup>th</sup> century to describe the output signal of a circuit (“Feedback,” 2021).

*Feedback*, about those items on the evaluation form, is believed to drive improvement in teacher evaluation. Feedback, according to advice, should be frequent, timely, specific, and private to be effective. However, our knowledge about giving feedback to teachers is based on research about giving feedback to students. Also, more has been written on how to deliver feedback than how its recipients process, use or reject feedback (Hazi, 2019b).

*Reflection* is “a way of learning from what we do, experience, and understand about any knowledge or teaching method we are using. It is a way of rigorously examining our actions and decisions and improving the overall quality of our work” (Arredondo Rucinski, 2005, p. 79). Reflective thinking, according to Dewey (1933), is a large part of learning. Schon (1983) helped reintroduce reflection into thinking about teaching practice at the height of the influence of process-product research on teacher evaluation. Some look to the post-conference to provide opportunities for teacher reflection (Ponticell et al., 2019).

### **Labor Law**

Laws are “the minimum forms of protection for persons, property and promises” (Hart, 1961, p. 195). Starting in the 1960s, when teachers were wrongfully dismissed and aggressively fought in courts to be re-instated, concepts from law began to influence teacher evaluation. Teachers and their unions helped to establish employee rights first in case law, then in collective bargaining contracts and eventually in federal and state statutes. Concepts from labor law focus on transparency and fairness for the benefit of teachers and include: arbitrary and capricious, due process, evidence, remediation, and uniform procedures.

*Arbitrary and capricious* is used to characterize administrators who fail to follow procedures for evaluation and teacher dismissal. *Due process* is the safeguards of notice, hearing, and opportunity to improve prior to dismissal so that teachers’ 14<sup>th</sup> Amendment rights have not been violated. Administrators must show *substantial evidence* (usually observation over time with documentation) that tenured teachers should be dismissed on statutory grounds. Administrators are asked whether a teacher’s behavior is *remediable*, and thus, given an opportunity to improve before dismissal in most cases.

### **Econometrics**

Economics is concerned with “how best to allocate scarce resources” (Furlong & Lawn, 2011, p. 85). Econometrics addresses the measurement of quality, cost efficiency, effectiveness, incentives, and sanctions; and the relationship between education and pupil outcomes such as achievement, earnings, job quality, and crime rates. Economists looked at inputs such as teacher characteristics (i.e., graduate point average, certification, graduate course work, and exam scores) as indicators of teacher quality but found little relationship between them and student achievement. During the last two decades, economists looked at the relationship between student achievement and *value-added measurers* (VAMs) of teachers. VAMs were first used by

Williams Sanders in the field of agricultural for genetic and reproductive trends in livestock (Amrein-Beardsley, 2014). In education, a VAM is a statistical tool used to measure the relationship between a teacher and student growth or loss on standardized achievement tests, which some argue is flawed despite its appearance of objectivity (Paige & Amrein-Beardsley, 2020).

## Supervision

Supervision is concerned with the learning and improvement of preservice and teachers in service. Some take a big-tent view where practice includes curriculum work, professional development, and classroom visitation. Some take a narrow view where supervision is synonymous with teacher evaluation. Supervisory concepts do have a place, albeit small, in teacher evaluation. The preconference was made popular in the 1980s through Morris Cogan's (1973) clinical supervision. *Preconference*, a meeting between the supervisor and teacher to plan observation, helped teachers to protect themselves from the principal's unannounced visits. The preconference found its way into collective bargaining agreements as well as state statutes (Hazi, 1998).

## Teacher Education

Teacher education is concerned with the preparation and learning of teachers. More common in teacher education, the portfolio is another method of evaluation "rediscovered" in the recent national focus on teacher quality in teacher evaluation (Lavigne & Good, 2014). Originating in 18<sup>th</sup> century Italy, a *portfolio* is a movable case for carrying loose papers ("Portfolio," 2021). It is a term, like feedback, that has been used and defined in multiple occupations and fields to include art, architecture, finance, law, and writing. "The contents of these containers tend to be the work's end product, direct evidence of its quality" (Bird, 1990, p. 242). The process for compiling the portfolio can be an exercise in amassing paper, or an opportunity for authentic assessment, professional development, and reflection on practice (Zepeda, 2002).

Thus, in Table 1 there are many concepts that influence teacher evaluation. Many have found their way into local policy, collective bargaining contracts, case law and statute. These concepts result in rules that must be followed, or actions avoided to promote accuracy and fairness, and to balance the needs of the organization with the rights of teachers. *Their embeddedness in law and policy make teacher evaluation impervious to change*<sup>8</sup> because it takes a majority vote of board members or legislators or a judge's ruling to change what exists in law and policy. An example is school board members and teachers modifying teacher evaluation and other work conditions during the COVID pandemic by negotiating a memorandum of understanding (Hazi, 2022).

Educational psychology has had the most influence on practice. Psychometric concepts such as validity and reliability are the basis of instrument design and training. In fact, because teacher evaluation is a "test" in federal law, employers must treat teachers in a *uniform* way so that they

---

<sup>8</sup> I think of these concepts as the "bedrock" of practice. In our history of reforming teacher evaluation, we have been filling and patching potholes with substance that never seems to hold. To revise practice, one would have to rehabilitate its road bed by pulverizing and mixing cement into the existing structural section ("Reconstruction of road beds," 2021).

do not discriminate due to race, color, religion, sex, or national origin (Hazi, 1998; *Uniform Guidelines*, 1978, 2014). Subsequently, teachers can challenge the validity or reliability of teacher evaluation in court, claiming that an evaluator discriminates or violates law.

*Concepts address the summative purpose more than its formative.* They influence its *procedure* (12), the *evaluator* (4), *instrument design* (5), its *purposes* (2) and *training* (2). Few address help or improvement. *Remediation*, *goal-setting*, and *preconference* appear to help teachers and address “the opportunity to improve.” Furthermore, the psychometric concepts are used in the name of principal accuracy in the instrument--not its *utility* or usefulness for teachers.

*The concepts have become the language of teacher evaluation.* These concepts address instrument design and delivery and evaluator training and procedure. They reflect a patchwork rather than a coherent, conceptual framework for practice. Nonetheless, these concepts have become the *common property* (Scriven, 1991a) of teacher evaluation. When an idea from academic discourse is used without acknowledgement, it is a useful invention, and endures within the everyday familiar, that concept is considered common property.

### Supervision Scholarship in the Discourses

Table 2 presents examples of supervision scholarship within the discourses, with an eye towards including members of COPIS and those more recently published in this journal. The academic department, journal/article title, or literature cited was used to situate it in a discourse. Citations to authors outside the field are also provided to help locate other scholars in the discourse.<sup>9</sup>

The writings are illustrative, not exhaustive. A work can be situated in more than one discourse, especially with a multi-vocal literature (e.g., Scalzo Wilson, 2018; Tang & Chow, 2007). The placement of a scholar or work in a category can “appear contentious, by design or accident” (Bazerman, 2009, para 2), but is meant to provide a point for dialogue within and across discourses, not to fix a work in any one discourse for all time. After all, “most people move within and between different discourse communities” (“Discourse community,” 2021, para 2).

Two additional discourse communities have been added to Table 2 because they have produced scholarship, but not restrictive concepts of practice. Administration/Leadership is concerned with administrators and their challenges in evaluation, while the Policy discourse usually is focused on its implementation and that which facilitates or hinders.<sup>10</sup>

From the titles and content of articles presented in Table 2, many supervision scholars have written about teacher evaluation in the past two decades. This aligns with the national interest that began in 2001 with No Child Left Behind’s focus on teacher quality. Supervision seems to be the only discourse to have its scholars go through a period of abandonment begun in the 1990s, as noted earlier, then reconciliation decades later, when many wrote about the challenges of high-stakes evaluation and principals as instructional leaders. Supervision scholars have

<sup>9</sup> Citations of chapters and books were kept to a minimum, since they are more difficult to search.

<sup>10</sup> I considered: entrepreneurial (e.g., Danielson & McGreal, 2000), journalistic (Goldstein & Shapiro, 2020), and international (e.g., Tuytens & Devos, 2017), but excluded them because they reflect fora.

contributed to most of the discourses within the past two decades, except the Econometric. They also address the evaluation of teachers in service more than preservice.

Table 2: The Discourses of Teacher Evaluation with Examples

<b>The discourse</b>	<b>Supervision scholars in the discourse</b>	<b>Scholars beyond supervision in the discourse</b>
Evaluation	Hazi (2022)	Scriven (1967, 1991a, 1991b) Scalzo Wilson (2018)
Psychometric	Zepeda & Jimenez (2019)	Good (2014), Good & Lavigne (2015)
Psych: Organizational	Allen & Topolka-Jorisenn (2013) Starrett (2015)	Guskey (2020) Maslow & Kelley (2012)
Psych: Developmental	Ponticell et al. (2019)	Guskey (2002) Flushman et al. (2019) Le & Vasquez (2011)
Law	Hazi (1989, 1994, 2017)	Hemphill & Marianno (2021)
Administration/Leadership	Glanz et al. (2006), Holland (2005), Tuytens & Devos (2017),	Murphy et al. (2013), Donaldson (2021)
Policy	Hazi & Arredondo Rucinski (2009), Mette et al. (2017), Paufler et al. (2020a), Flores & Derrington (2018)	Adams et al. (2018), Berliner (2018), Reinhorn et al. (2017)
Econometrics		Chetty et al. (2012), Kane et al. (2011), Kane et al. (2013), Malinowski (2011)
Supervision	McGreal (1983), Oliveras-Ortiz (2015, 2017), McGhee (2020), Sullivan (2016)	Hoy & Forsyth (1987)
Teacher Education	Burns & Badiali (2015), Paufler et al. (2020b), Currin et al. (2019)	Lustick & Sykes (2006) Scheeler et al. (2004) Scalzo Willson (2018) Tanguay (2020) Willis & Davies (2002)

Nolan and Hoover (2011) characterize the field's complex response towards evaluation: compatibility "plagues" the scholars, while practitioners experience supervision as teacher evaluation, despite the attempt to "soften the threat" with the language of development and improvement. Supervision is considered by some to be the formative (helping) aspect. Others reconcile the formative and summative functions with an inclusive view of improvement of the individual and of the organization done by teachers, administrators, and clinical supervisors with a variety of supports such as peer coaching, mentoring and action research (e.g., Sullivan & Glanz, 2002). Some see the two purposes vested in the principal, but one done less frequently (e.g., McGhee, 2020) or done sequentially with the summative picking up where the formative ends (Zepeda, 2016).

While many have written about teacher evaluation, *supervision scholars have contributed few concepts to its practice*. Instead, scholars have written about many promising techniques that can lead to improvement such as action research (Glanz, 2014), coaching (Costa & Garmston, 2015), mentoring (Blasé, 2009), professional learning communities (Arredondo Rucinski, 2016), reflection and inquiry (Yendol-Hoppey et al., 2019).<sup>11</sup> They are considered by some as alternatives to supervision (Sullivan & Glanz, 2000), a way to differentiate supervision, and teacher support structures within the professional development arm of supervision (Zepeda, 2006).

However, they are “add-ons” (Zepeda et al., 2020) that depend upon the will of the principal or superintendent who sees value in their use. They are *not* in the bedrock of practice, even though “considerable effort has been devoted over the years to ‘soften’ the bureaucratic language of teacher evaluation with these supports” (Murphy et al., 2013, p. 349). These “softer” concepts have been difficult to graft onto a practice that has control over teacher labor.

As a field of study grows, multiple discourses appear as a natural part of its evolution. A discourse can have its own definition of evaluation and hold different foci for research and assumptions about its practice. A few examples are presented.

In the administration/leadership discourse, Hallinger et al. (2014, p. 3) define it as “the formal assessment of a teacher by an administrator, conducted with the intention of drawing conclusions about his/her instructional performance for the purpose of making employment decisions.” In addition, those in this discourse may look to the principal to be the instructional leader and seek to correlate teacher quality with student achievement.

In the policy discourse, closely allied to administration and psychometrics, Darling-Hammond et al. (2011), taking a broad view of teacher quality, want successful systems that improve and support career decisions for both pre-service and in-service teachers. They advocate for standards-based evaluation with evidence of student learning (but not with VAMs which tend to be unstable), and well-designed performance assessments such as the Educational Teacher Professional Assessment (edTPA) and National Board Certification that use an array of evidence about teaching practice.

*Teacher evaluation is a practice difficult to research.* Few studies capture the complexity of conducting evaluations or its many challenges within an accountability context. Some scholars research different topics over time. For example, Derrington, studied teacher evaluation through the eyes of superintendents (2014), principals (Derrington & Campbell, 2018) and their preparation (Derrington, 2016), teachers (Derrington & Martinez, 2019), and needed professional development (Derrington & Kirk, 2017). Others organize a team to look at the influence of political culture on legislation and policy across many states, and principals at the intersection of supervision and evaluation (Mette et al., 2017, 2020).

---

<sup>11</sup> I list these techniques because some supervision scholars advocate their use (e.g., Arredondo Rucinski, 2016; Ponticell et al., 2019; Sullivan & Glanz, 2000) but their concepts have not become well-established in practice as those in Table 1.

Still others develop rich case studies drawing from multiple discourses. Reinhorn et al. (2017) is one such example, who show both depth and breadth by researching how principals and teachers made sense of and implemented teacher evaluation policy (for both accountability and teacher development) in schools. It is not enough to find that teachers believe their evaluations to be positive, but also to understand why. They discovered many contextual details such as policy spillovers that supported teachers, incentives, teacher-principal relationships, and once criticized initiatives that became accepted over time.

Another example comes from Holloway and Brass (2018) who looked at differences over time. Early career English teachers implementing standards were interviewed, then again a decade later in the implementation of value-added teacher assessments. In the first study teachers expressed concern about ethical dilemmas, students, and their mental health, as they struggled to follow standards, exercise their professional judgment, and resist what they criticized. However, a decade later, they complied with all mandates, willingly attended professional development, competed against colleagues for merit pay, and agreed with how the system evaluated their performance. Thus, over time, teachers abandoned their professional identity and perspectives about reform efforts, that once might have outraged them because these initiatives had become normalized.

Two topics are worthy of closer examination: performance assessment of preservice teachers and the complications of feedback. Topics include: the tensions between the formative and summative (e.g., Scalzo Willson, 2018), the high stakes consequences of the Educational Teacher Professional Assessment (edTPA) (e.g., Tanguay, 2020), and how highly effective cooperating teachers contribute to preservice teacher effectiveness (Ronfeldt et al., 2018).

Feedback for both preservice and in service teachers is in need of critique. Its logic is troubling because it assumes that with sufficient feedback, teachers will improve their practice. However, the simplistic advice of giving frequent, timely, specific, and private feedback is rooted in our understanding of giving students feedback and a behavioral view that takes teaching out of context of the lesson, subject, and students. It ignores that many teachers do not find principal feedback useful. Feedback will continue to be important, but perhaps less from evaluators. Yet, we know limited information about what happens with feedback when it comes from peers in coaching, mentoring, and cooperative teaching relationships. We know little about how teachers process feedback. The seminal work of Ilgen et al. (1979) helps us understand just how complex feedback is and that teachers tend *not* to be receptive to criticism or negative feedback. However, Tang and Chow (2007) account for variables absent from most studies of teacher feedback--the use of evidence and feedback in contexts of both the lesson itself and in the post-conference. They help us imagine how feedback can be discussed and evaluated, allowing the teacher to be engaged in evaluating their own performance.

In summary, this has been an excavation of sorts – “fieldwork” to examine the scope of teacher evaluation through its many discourse communities. Many discourse communities are natural in a field’s evolution with each helping us better understand its many niches with lens to examine many aspects of a complex practice. Evaluation provides its foundational concepts. Psychometrics is the measurement of various teacher qualities or standards. Organizational psychology addresses attitudes, behavior, and conditions in the workplace. Developmental

psychology is concerned with learning and behavioral change. Labor law is concerned with fairness, transparency, and protection from arbitrary and capricious action of administrators during evaluation. Econometrics deals with the relationship between teacher quality and effectiveness and various student outcomes. Supervision is about preservice and in-service teacher improvement and learning. Teacher education is concerned with the preparation and learning of teachers. Administration/leadership examines the role and challenges of the principal and others in evaluating teachers, while policy is concerned with conditions for the implementation of teacher evaluation.

Many concepts restrict teacher evaluation and its improvement, making it difficult to change, and promoting its summative purpose over its formative. Educational psychology has influenced teacher evaluation the most, with its foci of instrument design, delivery, and training and with an emphasis on accuracy to the exclusion of teacher usefulness. Supervision's influence on practice has been limited, but most importantly, can offer promising techniques for the formative purpose. Because their use is "add-on," depending on adoption of willing principals and superintendents, scholars can hope to influence practice through their advocacy of teaching, writing and research. Both fields may need to consider improvement as a complementary system to teacher evaluation (Hill & Grossman, 2013).

## **Conclusion**

In conclusion, two trends appear to encourage more research and scholarship. First, there is an international interest in the formative:

Scholars in some countries have recently begun to call for teacher evaluation to become a meaningful form of professional development with potential benefits for both teachers and schools.... As teachers are afforded a greater array of professional development alternatives as evaluation options teachers' preferences need to be further investigated, and qualitative approaches could be helpful in this regard. (Conley et al., 2016, p. 168)

Since scholars in other countries tend to cite U.S. scholars, continued efforts with research on feedback and promising techniques such as action research, coaching, mentoring, professional learning communities, and reflection can be influential.

Second, instruments and technologies will continue to shape teacher evaluation and are worth studying to better understand their seductions and challenges. Topics here include: distant supervision (Schroeder & Currin, 2019), rubrics (Tenam-Zemach & Flynn, 2015), on-line platforms (Hazi, 2014), bug-in-ear technology (Scheeler et al., 2018), telepresence robots (Burbank et al., 2021), and virtual evaluation (M.L. Derrington, personal communications, January 16, 2022). So, while some may be lured away to other topics, staying with one's niche may prove consequential when schools get back to the business of improvement.

This fieldwork has been a journey of thinking differently about teacher evaluation: its foundations, disciplinary influences, and scholars. Recognizing the many discourses helped me discover concepts I'd not before considered (e.g., Scriven's principles), revisited concepts in unique ways (e.g., feedback), see other audiences and fora for my work (e.g., teacher education,

the international community), and wisely poach the important and useful, given the topic, time, and audience. While it may seem presumptuous to place a scholar's work in a discourse, it should be viewed as a momentary opportunity "to consider the fit." A scholar must be the one to responsibly place him/herself in the discourse(s), assess its audiences, and consider how one's thinking advances its concepts and influence.

A discourse community is a group of people who share common goals, a common language, and continue to produce scholarship for practitioners and scholars. Some may be *transients* who join, then leave depending upon interest. After all, discourse communities "both influence and are influenced by the larger communities within which they are situated" (Swales, 2016, p. 9). Sometimes individual interest depends upon personal history, location, career stage, national interest, and funding. It may also depend on whether supervision concepts have relevance and currency within the evolving focus on improvement in practice. After all, a discourse community is formed one scholar at a time.

## References

- Adams, P., Mombourquette, C., Brandon, J., Hunter, D., Friesen, S., Koh, K., Parsons, D., & Stelmach, B. (2018). A study of teacher growth, supervision, and evaluation in Alberta: Policy and perception. *Journal of Educational Supervision*, 1(2), 1-16. <https://doi.org/10.31045/jes.1.2.1>
- Allen, A. S., & Topolka-Jorisenn, K. (2013). Using teacher learning walks to build capacity in a rural elementary school: Repurposing a supervisory tool. *Professional Development in Education*. <https://doi.org/10.1080/19415257.2013.851104>
- Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. Routledge.
- Arredondo Rucinski, D. (2005). Standards for reflective practice. In S. Gordon (Ed.), *Standards for instructional supervision: Enhancing teaching and learning*, (pp. 77-90.) Eye on Education.
- Arredondo Rucinski, D. (2016). *Real world professional learning communities: Their use and effects*. Rowman and Littlefield.
- Bazerman, C. (2009, November 22). Issue brief: Discourse community theory. <https://web.archive.org/web/20151117012340/http://www.ncte.org/college/briefs/dc>
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of discipline* (2<sup>nd</sup> ed.). Open University Press.
- Berliner, D. (2018). Between Scylla and Charybdis: Reflections on and problems associated with the evaluation of teachers in an era of metrification. *Education Policy Analysis Archives*, 26(54), 1-24. <https://epaa.asu.edu/ojs/article/view/3820>
- Bird, T. (1990). The schoolteacher's portfolio: An essay on possibilities. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 241-256). Sage.
- Blasé, J. (2009). The role of mentors of preservice and inservice teachers. In L. Saha & A.G. Dworkin (Eds.), *International handbook of research on teachers and teaching* 21, (pp.171-181). Springer.
- Blumberg, A. (1990) Comment: Toward a scholarship of practice. *Journal of Curriculum and Supervision*, 5(3), 236-243.
- Bolin, F. (1988). Does a community of scholars in supervision exist? *Journal of Curriculum and Supervision*, 3(4), 296-307.
- Borg, E. (2003). Discourse community. *ELT Journal*, 57(4), 398-400. <https://openlab.citytech.cuny.edu/eng1121-o436-spring2021/files/2021/01/Borg-Discourse-Community.pdf>
- Bracken, D., Rose, D., & Church, A. (2016). The evolution and devolution of 360<sup>0</sup> feedback. *Industrial and Organization Psychology*, 9(4), 761-794. DOI: 10.1017/iop.2016.93.
- Brookhart, S. (2013). *How to create and use rubrics for formative assessment and grading*. Association for Supervision and Curriculum Development.
- Burbank, M., Goldsmith, M., Bates, A., Spikner, J., & Park, K. (2021). Teacher observations using telepresence robots: Benefits and challenges for strengthening evaluations. *Journal of Educational Supervision*, 4(1), 68-93. <https://doi.org/10.31045/jes.4.1.6>
- Burns, R., & Badiali, B. (2015). When supervision is conflated with evaluation: Teacher candidates' perceptions of their novice supervisor. *Action in Teacher Education*, 37(4), 418-437. <https://doi.org/10.1080/01626620.2015.1078757>

- Chetty, R., Friedman, J., & Rockoff, J. (2012, January). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. NBER Working Paper No. 17699. National Bureau of Economic Research. <https://www.nber.org/papers/w17699>
- Church, A., Bracken, D., Fleenor, J., & Rose, D. (2019). *The handbook of strategic feedback*. Oxford Scholarship Online. <https://doi.org/10.1093/oso/9780190879860.001.0001>
- Cogan, M. (1973). *Clinical supervision*. Houghton Mifflin Company.
- Conley, S., Smith, J., Collinson, V., & Palazuelos, A. (2016). A small step into the complexity of teacher evaluation as professional development. *Professional Development in Education*, 42(1), 168-170. <https://doi.org/10.1080/19415257.2014.923926>
- Costa, A. & Garmston, R. (2015). *Cognitive coaching: Developing self-directed leaders and learners (3<sup>rd</sup> ed.)*. Rowman & Littlefield.
- Currin, E., Schroeder, S., Bondy, E., & Castanheira, B. (2019). Tinker, tailor, supervisor, spy: Lessons learned from distant supervision. *Journal of Educational Supervision*, 2 (1). <https://doi.org/10.31045/jes.2.1.5>
- Danielson, C., & McGreal, T. (2000). *Teacher evaluation to enhance professional practice*. Association for Supervision and Curriculum Development.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., Rothstein, J. (2011, September 14). Getting teacher evaluation right: A background paper for policy makers. National Academy of Education. <https://files.eric.ed.gov/fulltext/ED533702.pdf>
- Darling-Hammond, L., Wise, A., Pease, S. (1983). Teacher evaluation in the organization context: A review of the literature. *Review of Educational Research*, 53(3), 259-328.
- Derrington, M.L. (2014). Teacher evaluation initial policy implementation: Superintendent and principal perception *Planning and Changing*, 45(1/2), 120-137.
- Derrington, M.L. (2016). Implementing teacher evaluation: Lattice of leadership. *Journal of Research on Leadership Education*, 11(2), 181-199. <https://doi.org/10.1177/1942775116658689>
- Derrington, M. L., & Campbell, J. (2015). Implementing new teacher evaluation systems: Principals' concerns and supervisor support. *Journal of Educational Change*, 16, 305-326. <https://doi.org/10.1007/s10833-015-9244-6>
- Derrington, M. L., & Campbell, J. (2018). High-stakes teacher evaluation policy: US principals' perspectives and variations in practice, *Teachers and Teaching: Theory and Practice* 24(3), 246-262. <https://doi.org/10.1080/13540602.2017.1421164>
- Derrington, M. L., & Kirk, J. (2017). Linking job-embedded professional development and mandated teacher evaluation: teacher as learner. *Professional Development in Education*, 43(4), 630-644. <https://doi.org/10.1080/19415257.2016.1231707>
- Derrington & Martinez, J. (2019). Exploring teachers' evaluation perceptions: A snapshot. *NASSP Bulletin*, 102(1), 32-50. <https://doi.org/10.1177/0192636519830770>
- Dewey, J. (1933). *How we think*. D.C. Heath and Company.
- Discipline. (2021, February 14). *Oxford English dictionary* (3<sup>rd</sup> ed.).
- Discourse community, (2021, February 14). *Wikipedia*. [https://en.wikipedia.org/wiki/Discourse\\_community](https://en.wikipedia.org/wiki/Discourse_community)
- Donaldson, M. L. (2009). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality*. Center for American Progress. <https://www.americanprogress.org/article/so-long-lake-wobegon/>
- Donaldson, M. (2021). *Multidisciplinary perspectives on teacher evaluation: Understanding the research and theory*. Routledge.

- Downey, C., Steffy, B. English, F., Frase, L., & Poston, W. (2004). *The three-minute classroom walk-through: Changing school supervisory practice one teacher at a time*. Corwin Press.
- Feedback. (2021, August 3). *Oxford English dictionary*.
- Flores, M., & Derrington, M. L. (2018). Improving teacher evaluation: Key issues for appraisers in a globalized era. *Teachers and Teaching; Theory and Practice*, 24(3), 203-208. <https://doi.org/10.1080/13540602.2018.1427017>
- Frase, L., Downey, C., & Canciamilla, L. (1999). Putting principals in their place: The classroom. *Thrust for Educational Leadership*, 28 (5), 36-40.
- Flushman, T., Guise, M., & Hegg, S. (2019). Improving supervisor written feedback: Exploring the what and why of feedback provided to pre-service teachers. *Issues in Teacher Education*, 28(2), 46-66. <https://files.eric.ed.gov/fulltext/EJ1239633.pdf>
- Furlong, J. (2013). *Education—anatomy of the discipline: Rescuing the university project?* Routledge.
- Furlong, J. & Lawn, M. (2001). *Disciplines of education: Their role in the future of education research*. Routledge.
- Garman, N. B. (2010). The dream of clinical supervision and the nightmare of political accountability nightmare. An address presented at the AERA 2010 Distinguished Research Award presented by the SIG: Supervision and Instructional Leadership, and published in its Fall 2010 Newsletter.
- Garman, N. B. (2020). The dream of clinical supervision and the nightmare of political accountability nightmare. *Journal of Educational Supervision*, 3(3), 7-17. <https://digitalcommons.library.umaine.edu/cgi/viewcontent.cgi?article=1092&context=jes>
- Glanz, J. (2014). *Action research: An educational leader's guide to school improvement (2nd ed.)*. Rowman & Littlefield.
- 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 on the phenomenon of supervision traveling incognito: A field's struggles for visibility. *Journal of Educational Supervision*, 2(1), 1-21. <https://digitalcommons.library.umaine.edu/jes/vol2/iss1/1/>
- Glanz, J., Sullivan, S., & Shulman, V. (2006, April). *Usurpation or abdication of instructional supervision in the New York City public schools?* Paper presented at the annual conference of the American Educational Research Association (AERA), San Francisco.
- Goldstein D., & Shapiro, E. (2020, June 30). Full-time return to classrooms when summer ends? 'That's wishful thinking.' *The New York Times*. <https://www.nytimes.com/2020/06/26/us/coronavirus-schools-reopen-fall.html?searchResultPosition=2>
- Good, T.L. (2014). What do we know about how teachers influence student performance on standardized tests: And why do we know so little about other student outcomes? *Teachers College Record*, 116(1), 1-41. <http://www.tcrecord.org/content.asp?contentid=17289>
- Good, T. L., & Lavigne, A. (2015). Rating Teachers cheaper, faster, and better: Not so fast. *Journal of Teacher Education*, 66(3), 288-293. <https://doi.org/10.1177%2F0022487115574292>

- Gordon, S. (1992). Paradigms, transitions and the new supervision. *Journal of Curriculum and Supervision*, 8(1), 62-76.
- Gordon, S. (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>
- Guskey, T. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, 8(3/4), 381-391. <https://doi.org/10.1080/135406002100000512>
- Guskey, T. (2020). Flip the script on change: Experience shapes teachers' attitudes and beliefs. *The Learning Professional*, 41(2), 18-22. <https://learningforward.org/journal/beyond-the-basics/flip-the-script-on-change/>
- Hallinger, P., Heck, R., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment and Evaluation Accountability*, 26, 5-28, <https://doi.org/10.1007/s11092-013-9179-5>
- Harris, B. (1978). Supervisor competence and strategies for improving instruction. *Educational Leadership*, 33(5), 332-335.
- Hart, H. (1961). *The concept of law*. Clarendon Press.
- Hazi, H. M. (1989). Measurement versus supervisory judgment: The case of Sweeny v. Turlington. *Journal of Curriculum and Supervision*, 4(3), 211-229.
- Hazi, H. M. (1994). The teacher evaluation-supervision dilemma: A case of entanglement and irreconcilable differences. *Journal of Curriculum and Supervision*, 9(2), 195-216.
- Hazi, H. M. (1998). Policy and legal considerations in supervision. In G. Firth & E. Pajak (Eds.) *Handbook of research on school supervision* (pp. 968-986). Macmillan.
- Hazi, H. M. (2012). *Expert judgment: A concept for teacher evaluation in a postmodern world*. A paper presented at the annual meeting of the American Educational Research Association, Vancouver, BC.
- Hazi, H. M. (2014). The marketing of teacher evaluation: The seductive claims of instruments. *The WERA Educational Journal*, 6(1), 2-9.
- Hazi, H. M. (2016, October 15). A letter to the field: An elder proposes reconciliation with teacher evaluation. A paper presented at the annual meeting of the Council of Professors of Instruction Supervision, Tampa, FL.
- Hazi, H. M. (2019a). Coming to understand the wicked problem of teacher evaluation. In S. J. Zepeda, & J. Ponticell (Eds.), *Handbook of educational supervision* (pp. 183-208). Wiley Blackwell Publishing.
- Hazi, H. M. (2019b). The language of instructional improvement in the U.S. In M. L. Derrington & J. Brandon (Eds.), *Differentiated teacher evaluation and professional learning: Policies and practices for promoting career growth* (pp. 149-171). Palgrave Macmillan.
- Hazi, H. M. (2022). Basic principles of teacher evaluation examined for school leaders in the 21<sup>st</sup> century. In J. Glanz (Ed.), *Managerial school leadership: New skills for the 21<sup>st</sup> century*. Rowman & Littlefield.
- 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), 1-22. <http://epaa.asu.edu/epaa/v17n5/>
- Hemphill, A., & Marianno, B. (2021). Teachers unions, collective bargaining, and the response to COVID-19. *Education Finance and Policy*, 16(1), 170-182. [https://doi.org/10.1162/edfp\\_a\\_00326](https://doi.org/10.1162/edfp_a_00326)

- Hill, H., & Grossman, P. (2013). Learning from teacher observation: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, 83(2), 371-384. <https://doi.org/10.1016/j.stueduc.2019.01.006>
- Holland, P. (2005). The case for expanding standards for teacher evaluation to include an instructional supervision perspective. *Journal of Personnel Evaluation in Education*, 18(1), 67-77. DOI: 10.1007/s11092-006-9009-0
- Holland, P. & Garman, N.B. (2001). Toward a resolution of the crisis of legitimacy in the field of supervision. *Journal of Curriculum and Supervision*, 16(2), 95-111.
- Holloway & Brass, (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>.
- Hoy, W., & Forsyth, P. (1987). Beyond clinical supervision: A classroom performance model. *Planning & Change*, 18, 210-223.
- Hunter, M. (1983). Script taping: An essential supervisory tool. *Educational Leadership*, 41(3), 43.
- Ilgen, D., Fisher, C., & Taylor, M.S. (1979). Consequences of individual feedback on behaviors in organizations. *Journal of Applied Psychology*, 64(4), 349-371. <https://doi.org/10.1037/0021-9010.64.4.349>
- Kane, T.J., McCaffrey, D.M., Miller, T., & Staiger, D.O. (2013, January). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Bill and Melinda Gates Foundation. <https://files.eric.ed.gov/fulltext/ED540959.pdf>
- Kane, T., Taylor, E., Tyler, J., & Wooten, A., (2011). Evaluating teacher effectiveness: Can classroom observations identify practices that raise achievement? *Education Next*, 11(3), 54-60. <http://educationnext.org/evaluating-teacher-effectiveness/>
- Kennedy, M. (Ed.) (2010). *Teacher assessment and the quest for teacher quality: A handbook*. Jossey-Bass.
- Killian, J. (1994, November 9). Letter to Ron Brandt, Director of Publications of ASCD. Copy in possession of Helen M. Hazi, COPIS member, West Virginia University.
- Lavigne, A. & Good, T. (2014). *Teacher and student evaluation: Moving beyond the failure of school reform*. Routledge.
- Lawn, M., & Furlong, J. (2011). The disciplines of education: between the ghost and the shadow. In J. Furlong & M. Lawn (Eds.) *Disciplines of education: Their role in the future of education research*. (pp.1-10). Routledge.
- Le, P. T.A. & Vasquez, C. (2011). Feedback in teacher education: Mentor discourse and intern perceptions. *Teacher Development*, 15(4), 453-470. <https://doi.org/10.1080/13664530.2011.635264>
- Lochmiller, C. R. (2019). Credibility in instructional supervision: Catalyst for differentiation. In M. L. Derrington & J. Brandon (Eds.), *Differentiated teacher evaluation and professional learning: Policies and practices for promoting career growth* (pp. 83-105). Palgrave Macmillan.
- Lustick, D., & Sykes, G. (2006). National board certification as professional development: What are teachers learning? *Education Policy Analysis Archives*, 14(5), 1-43. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1013.3474&rep=rep1&type=pdf>
- Malinowski, A. (2011). Strategic measures of teacher performance. *Phi Delta Kappan*, 92(7), 19-25. <https://doi.org/10.1177%2F003172171109200705>

- Maslow, V., & Kelley, C. (2012). Does evaluation advance teaching practice? The effects of performance evaluation on teaching quality and system change in large diverse high schools. *Journal of School Leadership*, 22(3), 600-632.  
<https://doi.org/10.1177/105268461202200307>
- McGreal, T. (1983). *Successful teacher evaluation*. Association for Supervision and Curriculum Development.
- McQuarrie, F. & Wood, F. (1991). Supervision, staff development, and evaluation connections. *Theory Into Practice*, 30(2), 91-97.
- McGhee, M. (2020). The supervision-evaluation debate meets the theory-to-practice conundrum: Contemplations of a practitioner turned professor. *Journal of Education Supervision*, 3(2), 1-5. <https://doi.org/10.31045/jes.3.2.1>
- 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/https://doi.org/10.31045/jes.2.2.1>
- Mette, I., Aguilar, I., & Wiczorek, D. (2020). A thirty state analysis of teacher supervision and evaluation systems in the ESSA era. *Journal of Educational Supervision*, 3(2).  
<https://doi.org/10.31045/jes.3.2.7>
- Mette, I., Fairman, J., Lech, P., & Frankland, M. (2019). Principal and teacher perceptions of performance evaluation and professional growth (PE/PG) system implementation. Maine Education Policy Research Institute.  
[https://usm.maine.edu/sites/default/files/cepare/Principal\\_and\\_Teacher\\_Perceptions\\_of\\_Performance\\_Evaluation\\_and\\_Professional\\_Growth\\_PE\\_PG\\_System\\_Implementation.pdf](https://usm.maine.edu/sites/default/files/cepare/Principal_and_Teacher_Perceptions_of_Performance_Evaluation_and_Professional_Growth_PE_PG_System_Implementation.pdf)
- Mette, I., 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.  
<https://files.eric.ed.gov/fulltext/EJ1134210.pdf>
- Millman, J. & Darling-Hammond (1990). *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*. Sage.
- Mueller-Hanson, R. (2021, June 28). Work performance management and assessment. *The oxford research encyclopedia of psychology*. Oxford University Press.  
<https://doi.org/10.1093/acrefore/9780190236557.013.25>
- Murphy, J., Hallinger, P., & Heck, R. (2013). Leading via teacher evaluation: The case of the missing clothes: *Educational Researcher*, 42(6), 349-354.  
<https://doi.org/10.3102%2F0013189X13499625>
- Nolan, J. & Hoover, N. (2011). *Teacher supervision and evaluation (3rd ed.)*. John Wiley & Sons, Inc.
- Oliveras-Ortiz, Y. (2015). Instructional leadership amidst the accountability frenzy: Using quality data to drive instruction. Advancing Improvement in Education Conference. San Antonio, TX.
- Oliveras-Ortiz, Y. (2017). School administrators as instructional coaches: Teachers' trust and perceptions of administrators' capacity. *School Leadership Review*, 12(1), 39-46.  
<https://scholarworks.sfasu.edu/slr/vol12/iss1/6>
- Outline of academic disciplines. (2021, February 14). In *Wikipedia*.  
[https://en.wikipedia.org/wiki/Outline\\_of\\_academic\\_disciplines?oldid=776838216](https://en.wikipedia.org/wiki/Outline_of_academic_disciplines?oldid=776838216)

- Paige, M. & Amrein-Beardsley, A. (2020). "Houston, we have a lawsuit": A cautionary tale for the implementation of value-added models for high-stakes employment decisions. *Educational Researcher*, 49(5), 350-359. <https://doi.org/10.3102/0013189X20923046>
- Paufler, N., King, K., & Zhu, P. (2020a). Promoting professional growth in new teacher evaluation systems: Practitioners' lived experiences in changing policy contexts. *Studies in Educational Evaluation*, 65, 100873. <https://doi.org/10.1016/j.stueduc.2020.100873>
- Paufler, N. A., King, K. M., & Zhu, P. (2020b). Delivering on the promise of support for growth? Evaluator perceptions of a new state teacher evaluation system. *Journal of Educational Supervision*, 3(2), 32. <https://doi.org/10.31045/jes.3.2.3>
- Peshkin, A. (1988). In search of subjectivity—One's own. *Educational Researcher*, 17(7), 17-21.
- Pinar, W. F., Reynolds, W.M., Slattery, P., & Taubman, P.M. (1995). *Understanding curriculum*. Peter Lang.
- Ponticell, J., & Zepeda, S. J. (2004). Confronting well-learned lessons in supervision and evaluation. *NASSP Bulletin*, 88(639), 43-59. DOI:[10.1177/019263650408863905](https://doi.org/10.1177/019263650408863905)
- Ponticell, J., Zepeda, S.J., Lanoue, P., Haines, J.G., Jimenez, A.M., & Ata, A. (2019). Observation, feedback and reflection. In S.J. Zepeda & J. Ponticell (Eds.), *The Wiley handbook of educational supervision* (pp. 251-280). Wiley Blackwell.
- Portfolio. (2021, July 27). *Online etymology dictionary*. <https://www.etymonline.com/word/portfolio>
- Reconstruction of road beds. (2021, July 4). [https://www.google.com/search?q=reconstruction+of+road+beds&client=safari&rls=en&ei=stfhYKyrKsTbtAbaz7zYAg&oq=reconstruction+of+road+beds&gs\\_lcp=Cgdnd3Mtd2l6EAxKBAhBGABQAFgAYI\\_5xChoAHACeACAAVqIAVqSAQExmAEAqgEHZ3dzLXdpesABAQ&sclient=gws-wiz&ved=0ahUKEwizKms4cnxAhXELc0KHdonDysQ4dUDCA4](https://www.google.com/search?q=reconstruction+of+road+beds&client=safari&rls=en&ei=stfhYKyrKsTbtAbaz7zYAg&oq=reconstruction+of+road+beds&gs_lcp=Cgdnd3Mtd2l6EAxKBAhBGABQAFgAYI_5xChoAHACeACAAVqIAVqSAQExmAEAqgEHZ3dzLXdpesABAQ&sclient=gws-wiz&ved=0ahUKEwizKms4cnxAhXELc0KHdonDysQ4dUDCA4)
- Reinhorn, S., Moore Johnson, S., & Simon, N. (2017). Investing in development: Six high-performing, high-poverty schools implement the Massachusetts teacher evaluation policy. *Educational Evaluation and Policy Analysis*, 39(3). 383-406. <https://doi.org/10.3102/0162373717690605>
- Ronfeldt, M., Brockman, S., & Campbell, S. (2018). Does cooperating teachers' instructional effectiveness improve preservice teachers' future performance? *Educational Researcher*, 47(7), 405-418. <https://doi.org/10.3102%2F0013189X18782906>
- Scalzo Willson, A. (2018). Coach and evaluator: Exploring how to negotiate both functions in the role of supervisor. *Journal of Practitioner Research*, 3(1). <https://doi.org/10.5038/2379-9951.3.1.1065>
- Scheeler, M., Morano, S., & Lee, D. (2018). Effects of immediate feedback using bug-in-ear with paraeducators working with students with autism. *Teacher Education and Special Education*, 41(41), 24-28. <https://doi.org/10.1177/0888406416666645>
- Scheeler, M., Ruhl, K., & McAfee, J. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education*, 27(4), 396-407. <https://files.eric.ed.gov/fulltext/EJ694123.pdf>
- Schon, D. (1983). *The reflective practitioner: How professionals think in action*. Basic Books, Inc.
- Schroeder, S., & Currin, E. (2019). Syncing our cycles: An inquiry-based coaching model for distant supervision. *Journal of Practitioner Research*, 4(1), Article 6. <https://doi.org/10.5038/2379-9951.4.1.1098>

- Scriven, M. (1967). The methodology of evaluation. In R. Stake (Ed.), *Perspectives of curriculum evaluation: An AERA monograph series on curriculum evaluation. No. 1*. Rand McNally.
- Scriven, M. (1991a). Beyond formative and summative evaluation (pp. 19-64). In M. W. McLaughlin & D.C. Phillips (Eds.), *Evaluation and education: At quarter century, ninetieth yearbook of the National Society for the Study of Education*. The University of Chicago Press.
- Scriven, M. (1991b). *Evaluation thesaurus* (4<sup>th</sup> ed.). Sage Publications.
- Scriven, M. (1994). Duties of a teacher. *Journal of Personnel Evaluation in Education*, 8(2), 151-184.
- Scriven, M. (1996). Truth and objectivity in evaluation. *Evaluation for the 21st century: A handbook* (pp. 477-500). Sage.
- Sergiovanni, T. (1992). Why we should seek substitutes for leadership. *Educational Leadership*, 49(5), 41-45.
- Starratt, R. (1992). After supervision. *Journal of Curriculum and Supervision*, 8(1), 77-86.
- Starratt, R. (1997). Should supervision be abolished? Yes. In J. Glanz & R. Neville (Eds.), *Educational supervision: Perspectives, issues, and controversies* (pp. 4-12). Christopher-Gordon Publishers.
- Starrett, T. M. (2015). Principal perceptions of walkthrough effectiveness. *School Leadership Review*, 10(1), 44-51. <https://scholarworks.sfasu.edu/slr/vol10/iss1/6>
- Stufflebeam, D. (2013). My tribute to a trail blazer: Evaluation iconoclast—Professor Michael Scriven. *The future of evaluation in society: A tribute to Michael Scriven* (pp. 73-91). Information Age Publishing, Inc.
- Sullivan, S. (2016). High-stakes accountability and the demise of formative supervision of instruction: The case of New York City. In J. Glanz & S. J. Zepeda (Eds.), *Supervision: New perspectives in theory and practice* (pp. 145-162). Rowman & Littlefield.
- Sullivan, S. & Glanz, J. (2000). Alternative approaches to supervision: Cases from the field. *Journal of Curriculum & Supervision* 15(3), 212-235.
- Swales, J. (2016). Reflections on the concept of discourse community. *Concepts and Frameworks in English for Specific Purposes*, 69, 7-19. <https://doi.org/10.4000/asp.4774>
- Tang, S., & Chow, A. (2007). Communicating feedback in teaching practice supervision in a learning-oriented field experience assessment framework. *Teaching and Teacher Education*, 23, 1066-1085. <https://doi.org/10.1016/j.tate.2006.07.013>
- Tanguay, C. (2020). High-stakes assessment in elementary education teacher preparation: Educators' perceptions and actions resulting in curriculum change. *Education Policy Analysis Archives*, 28(53), 1-36. <https://doi.org/10.14507/epaa.28.4840>
- Tenam-Zemach, M., & Flynn, J. (2015). *Rubric nation: Critical inquiries on the impact of rubrics in education*. Information Age Publishing, Inc.
- Tuytens, M., & Devos, G. (2017). The role of feedback from the school leader during teacher evaluation for teacher and school improvement. *Teaches and Teaching: Theory and Practice*, 23(1), 6-24. <https://doi.org/10.1080/13540602.2016.1203770>
- Uniform Guidelines on Employee Selection Procedures. (1978, 2014). <http://www.uniformguidelines.com/uniformguidelines.html>
- Willis, E., & Davies, M. (2002). Promise and practice of professional portfolios. *Action in Teacher Education*, 23(4), 18-27. <https://doi.org/10.1080/01626620.2002.10463084>

- Wise, A., Darling-Hammond, L., McLaughlin, M., & Bernstein, H. (1984). *Teacher evaluation: A study of effective practices*. Rand.
- Yendol-Hoppey, D., Jacobs, J., & Burns, R. (2018). Improving teacher practice-based knowledge: What teachers need to know and how they come to know it. In S.J. Zepeda & J. Ponticell (Eds.), *Handbook of educational supervision* (pp. 511-532). Wiley Blackwell Publishing.
- Zepeda, S. J. (2002). Linking portfolio development to clinical supervision: A case study. *Journal of Curriculum and Supervision*, 18(1), 83-102.
- Zepeda, S. J. (2006). High stakes supervision: We must do more. *International Journal of Leadership in Education*, 9(1), 61-73. <https://doi.org/10.1080/13603120500448154>
- Zepeda, S. J. (2016). *The leader's guide to working with underperforming teachers: Overcoming marginal teaching and getting results*. Eye on Education.
- Zepeda, S. J., Alkaabi, A., & Ravernier, M. (2020, January 30). Leadership and supervision. *Oxford research encyclopedia: Education*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190264093.013.617>
- Zepeda, S. J., & Jimenez, A. (2019). Teacher evaluation and reliability: Additional insights gathered from inter-rater reliability analyses. *Journal of Educational Supervision*, 2(2), 11-26. <https://doi.org/10.31045/jes.2.2.2>

## Author Biography

**Helen M. Hazi** is a professor emerita of Educational Leadership at West Virginia University and was a student of the east coast version of clinical supervision represented by Morris Cogan and Noreen B. Garman. She has been a teacher, a supervisor of curriculum and instruction, and an expert witness. Helen is a member of the Council of Professors of Instructional Supervision and a founder of the AERA SIG: Supervision and Instructional Leadership. She writes about legal issues that have consequence for supervision and the discourse communities of teacher evaluation in books and journals such as *Journal of Curriculum & Supervision*, *Journal of Educational Supervision*, *Journal of Staff Development*, and *Educational Policy Analysis Archives*. Legal issues in recent writings include: instructional improvement and its commodification; judgment; teacher evaluation instruments, statutes, litigation, and marketing; and professional development statutes. Her web page is <https://helenhazi.faculty.wvu.edu/home>. Contact is welcomed at [hmhazi@verizon.net](mailto:hmhazi@verizon.net).