The Teacher Leadership Effectiveness (TLE) Evaluation Tool for Oklahoma Teachers: A Foucauldian Analysis

Naomi K. Poindexter, Oklahoma State University

Introduction

In 2011, Tulsa Public Schools, working with the Gates Foundation, developed the Teacher Leader Effectiveness program to improve teaching. After a year of piloting the system at select school sites, the evaluation tool became official district policy. During the 2012–2013 school year, 90 percent of all Oklahoma public school districts adopted the TLE as their teacher evaluation tool. Other districts have adopted other evaluation methods and the requirement is all public-school teachers must be evaluated to maintain employment. The lead administrator for each school, always the evaluator, observes the teacher in two, informal observations (i.e. the administrator may do an informal walk-through as an observation). Conferences are held after each observation. Subsequently, a formal evaluation—a third observation—is completed, discussed, and the administrator assigns a score. Each interaction is documented on the TLE and signed by the teacher. A non-tenured teacher is observed and evaluated twice as many times and has more attempts to earn a three (3) or "Effective" score. The score ranges from one to five; a five represents the rating "Highly Effective." Presently, district officials are including student survey scores and student test scores as part of the teacher evaluation process; however, in this paper I focus on the evaluation process completed by the evaluator on a teacher's teaching performance.

Now retired, Lynn Stockley was the former Tulsa Teacher Classroom Association President during the creation of the aforementioned teacher evaluation tool (personal communication, April 20, 2014). She states despite the combined efforts of administrators, teachers, and association representatives, difficulties with the evaluation tool exist. Namely, the inconsistency that may exist between a teacher's informal observation and the final evaluation score, which decides employment may be problematic. Secondly, some teachers believe they perform at a higher level than their scores reflect; this potentially results in termination when a final evaluation score falls below a three. When a

teacher falls below a three on any of the 20 indicators, he or she is placed on a Professional Development Plan (PDP), increasing the possibility of not being rehired (L. Stockley, personal communication, April 20, 2014).

In this paper I analyze the summary, "Overview to the Teacher Leadership Effectiveness Evaluation," presented as a preface to the Oklahoma TLE program, through the Foucauldian lens of correct training. My purpose is to illuminate power and pitfalls of teacher evaluation and to suggest educators and administrators begin to shift perceptions of evaluation in order to use discussion, understanding, and dialogue between the evaluator and the teacher as a method to determine evaluation scores rather than to privilege only the evaluator's reality. Now I turn to defining Foucault's (1977) notion of Correct Training.

Hierarchical Observation

Foucault (1977) posits that discipline is exacted through hierarchical observation using an apparatus, techniques to observe the effects of power, and visible coercion (p. 170). The purpose of hierarchized observation is beyond surveillance; its purpose is to control, transform, implement conformity, and create accountability (p. 177). The perfect apparatus is a tool in which a single look will identify and judge every offense and activity. This observation supervises the worker, the work, and all places of work. Hierarchical observation is enacted from the top through to the bottom of the organization and is laterally interdependent, "functioning like a piece of machinery" (p. 177); it determines what does not measure up, departs from the norm, or is pathology in need of remedy.

Normative Judgment

Normative judgment lies within the power structure (Foucault, 1977), and one can be punished when he or she departs from the norm. When one conforms, gaps are reduced and corrective measures are designed to reduce gaps between subjects as "intensive forms of training," normalizing the individual (p. 179). The purpose of correction is to regulate individual behavior and correct any nonconformity, which is seen as a defect. Any defect is then eliminated by the exercise of discipline, enforced by use of "the fear of punishment" (p. 180). Foucault explains normative judgment allows individuals to be known, ranked, and graded (p. 181), allowing power to become an integrated, machine-like system: "organized, automatic and anonymous" that makes control more efficient (p. 176).

The Examination

The examination (Foucault, 1977), comprised of hierarchical observation and normalizing judgment, is highly ritualized, formatted,

and administrated method through which individuals are judged, categorized, and potentially pathologized. The examination is a form of the experiment; it establishes truth about the individual. It embodies a whole corpus of knowledge and powers that Foucault calls *operational schema*, which creates power relations that make up the body of knowledge.

Written documents of each individual's rank (e.g. numbered evaluation score) position the subject within a stationary framework. In addition, the examination documents the results of normative judgment, the effectiveness of treatments, and the change of disease. The examination homogenizes individual features through norms that control subjects. The exam makes the individual a "case" to be examined; the individual becomes a subject whose identity is determined by the exam. In this way, the examination, through hierarchical observation and normative judgment, "fabricate(s) the cellular, organic, genetic, and combinatory individuality" (p. 192), labeling the individual.

Normative Judgment as the End of Evaluation Minimizes the Possibility of Teacher Growth

Hierarchical observation is manifest within the informal observation of every teacher utilized by any school district. In the case of the TLE, the observation occurs and is performed when the administrator inspects each teacher against the 20 dimensions of the rubric, previously set by administrators, teachers, and others. The observation is documented and allows the individual teacher to be seen and identified and is considered a necessary step to identify the work of each teacher. Foucault (1977) refers to this as the functionalism of power. Therefore, it becomes impossible for a teacher to go unnoticed or remain unobserved, as TLE expectations require administrators to observe and plot each teacher's performance on TLE rubrics. The TLE process is an integrated system of structures that includes planned and unplanned (walkthrough) observations, as well as electronically tracking teacher scores on 20 dimensions of the TLE rubric that will result in a final evaluation score defining each teacher. Observation is necessary and vital to begin conversations about what was observed, however, the process of assessment should not end there. The growth mindset should be a main focus of observation so as not to result in permanent asymmetrical relational power, which is top heavy and oppressive (Foucault paraphrased from Bess, 1980).

After an observation, the administrator measures a teacher against the 20 rubric dimensions often using normative judgment alone. The language of the rubric potentially creates commonality in the meaning and interpretation of the rubric words and vocabulary through which both evaluator and teacher communicate shared reality through the

post-observation discussion. The observations are the evaluator's "intentional study and analysis of the teacher's performance" (TLE, Handbook for Tulsa Model Evaluators, 2012–13). The Handbook states the observation and conference process is a "critical opportunity for teachers to receive meaningful feedback from evaluators on the improvement in their instructional practice" (p. 12). Glaringly absent is the mention of the leveraging conversation that *teachers* should have with their evaluator. This fact supports my argument for shared knowledge between the evaluator and teacher to reach a shared understanding and score that reflect this understanding. Next, I discuss Foucault's notion of multi-directional power as essential for integration necessary for the paradigm that fosters conversation and sharing.

Foucault (1978) posits the rules of variations indicate power-knowledge is not static, but are "matrices of transformations" and thus power does not have to be one way (p. 99). For example, the teacher may then be prescribed additional professional development if the observation scores are not at the norm, but rather fall below the norm of three or the designation "Effective." The teacher can dialogue with the evaluator after the third evaluation, which will lead to the final evaluation score. Stockley states some teachers are not prepared for this discussion and are unable to leverage their input, resulting in a lower TLE score (personal communication, April 20, 2014). The lowered score punishes the teacher and immobilizes him or her within the narrow frame of top-down teacher evaluation. In most cases, the teacher is known as a poor teacher if his or her score falls below a 3. The TLE is perfectly aligned to Foucault's correct training and focuses educators on the score as the desired end of evaluation. However, the inclusion of teacher input is the acknowledgement of reality constructed through dialogue and mutual understanding, potentially changing the present perception of evaluation as top-down surveillance held by most involved in public education. Might such a practice begin to shift methods and perceptions of evaluation by recognizing constructive feedback and improvement as the ultimate goal of instruction and not the evaluation score itself? I first analyze the role of hierarchical observation and normative judgment in teacher evaluation.

Foucault (in Dreyfus & Rabinow, 1982) states normalizing paradigms, similar to the strict use of rubrics, can have damaging effects when a narrow, objective view is used to view society as a whole. Hubert and Rabinow (1982) ask, "How is one to analyze the power relationship?" (p. 222). Foucault (Hubert & Rabinow, 1982) argues pre-ordaining, or diverging on a reality should be sublimated to the idea of nonnormalizing reality through an interpretive dimension.

The Subject and Power, Power and Communications Rationale for Educational Discussion Teacher Evaluation

Humans are made into subjects within our culture through power (e.g. Fascism and Stalinism were pathological forms of power and society largely compares all power relations to these forms) (Foucault, 1982). Foucault suggests a "new economy of power relations," that represents a relationship between rationalization and political power and is therefore historically situated and appears differently in various contexts. Historically, pastoral power through Christianity spread a code of group ethics throughout the ancient and new world. However, in the 1800s the concept of a "modern state" came to the fore. It brought with it a "matrix of individualization, or a new form of pastoral power" (p. 215). Within this new construct, individual "salvation" became important and meant an improvement to one's life here on earth. The number of officials increased and they implemented plans through a growing complex state apparatus (e.g. municipal police departments). To illustrate, 1800s police work not only encompassed a watch against crime, but also included social services. This more complex role of the police insured people received supplies, obtained standard health care, and had access to goods and services. This emphasis on equal access and a means equally to distribute goods and services helped usher in the ideals of the modern state and the new ideas of equity. In addition, this new knowledge or way of using power benefitted the general population and improved individual lives through a "series of powers as seen in the family, medicine, psychiatry, education and employers" (p. 215). This use of power promised positive consequences as well as introduced a capacity for great harm (e.g. goods and services may not be distributed fairly, marginalizing some people and causing debilitating inequity). Foucault (1980) demonstrates the new use of power citing the educational institution. Space and regulations govern the internal activities and people within the institution and each uses a capacity called communication power. Communication (i.e., lessons, questions, and answers, orders, coded signs of obedience, grades) ensures behavior within, learning or acquiring, and apprenticeship.

Dreyfus and Rabinow (1982), employing Foucauldian theory, explain state power is not restricted, but can be direct or multidirectional, operating top-down but also from the bottom up. Foucault (1982) states power exists only when it is exercised and exists from multiple directions rather than only in a hierarchical manner. Power relationships exists between free subjects and involves communication, therefore, the power relationship is not dualistic, as it was in the violent relationship of slave to master.

Communication is the key, but the exchange of information must benefit each involved entity. At present, the TLE functions as *Panopticism*. For instance, the Lessons Learned portion of TLE is used to leverage teacher and administrator input, however, the system relies on an administrator who will allow input. Untenured teachers are observed twice in a semester and evaluated twice per year, whereas tenured teachers (ones who have 3 years of successful employment) are observed and evaluated at half that amount. The final observation tool is a composite of scores from the first two observations for any semester, but a teacher is not required to give input for an evaluator to score. The idea assumes leveraging will occur, but does it?

Adding the Interpretive Dimension to Teacher Evaluation

Kuhn (as quoted in Dreyfus & Rabinow, 1982) states the interpretive dimension can neither be general moralizing nor personal preference. This dimension must be absent of judgment. Teacher and evaluator must use reasonable sense to determine the socially shared sense of language that explains good teaching as defined by the rubric. Evaluation is not based on personal emotion or fondness for a teacher. This is to say conversation utilizing the rubric language anchors the dialogue between teacher and evaluator, but does not limit it. Secondly, the evaluator must commit to open up the conversation around a teacher's performance. For example, the evaluator can open the conversation so the teacher must be willing and able to defend his or her performance with evidence. The evidence may be of the teacher's written original lessons, tape recorded or filmed lessons, student work products, and narrations of what was learned through teaching. This shared sense is determined through the rubric language itself and our socially accepted inherent meanings of words. Then there exists tension between what is socially accepted meaning and how the teacher's performance is being judged. In other words, what is being seen, and at what rubric level? Foucault (1990) states this interpretation is constructed. The investigator (i.e. teacher or evaluator) engages the other player with answers to the reasons that defend a position statement.

Conversation and dialogue are vital to shared knowledge and understanding paradigms, especially when a teacher receives an unsatisfactory score. A 2.8 score, in *any* indicator area, may result in a Professional Development Plan (PDP). Additional professional development and more administrative observations are issued to "assist" the teacher in correcting his or her behavior to meet the perceived meaning of the rubric. The teacher can enter the non-evaluative Quality Experiences Supporting Teaching (QUEST) program that offers teachers supportive mentors who offer suggestions and classroom assistance. This results in additional observation and more judgment for

the teacher who does not receive a score of 3 or Effective. A performance that differs from what is the most common type may be seen as undesirable. For instance, perhaps in a specific district the most common way of writing lesson plans is on a desired template sent by the principal to all teachers. A teacher coach observes the teacher for pathology that digresses from the norm since the overall weighted score must be at a 2.8 overall score on the evaluation or the teacher can be removed from employment in the district. Foucault states that power is an "open, more or less coordinated cluster of relations…" (quoted in Dreyfus and Rabinow, 1982, p.184). Power is not subject to one theory or origin, but is an analytic of relations of power. Foucault (1980) states power analytics must be employed to begin to conceive of its complexity (Dreyfus & Rabinow, 1982). One possibility is through discursive practice to "isolate, identify, and analyze the web of unequal relationships" (Dreyfus & Rabinow, 1982, p. 185). Therefore, teacher evaluation is not simply to inform the evaluator as presently recommended in the TLE training. Based on the example of teacher lesson plans, the style of communicating lessons may not fit the required format, but be very creatively written and implemented. However, at present the evaluator's view can label a teacher's lessons as inadequate because it does not fit within the evaluator's personal preference. This type of narrow view often resulting in punishment is what I argue is in need of remedy, and remedy begun by our shifting to view evaluation as a dynamic work in progress.

Dreyfus & Rabinow (1982) expound on Foucault's (1980) notion of enunciative modalities and discursive practice by referring to archeology as ideas historically situated. Enunciative modalities are the speech uttered when speaking about the principles concerning the formal discourse between, for instance, evaluator and teacher. Historically this speech has been hierarchical with the administrator situated above the teacher. Foucault (as quoted in Dreyfus & Rabinow, 1982) states this type of discourse can be decompressed through rarefaction. Rarefaction means the opposite of compression however, Foucault (1980) uses this term to discuss the density of language (speech act) operated within and around discourse that limits or constrains who can speak and what concepts can be used in the speaking. Foucault states points of choice can "reanimate already existing themes, which opens the formation to make room for possible options" (quoted in Dreyfus and Rabinow, 1982, p. 72). For example, the teacher may introduce another point or classroom example and create an additional reality, one not readily seen by the evaluator. Power through discourse will exist as it is exercised through language and ideas through conversation while discussing TLE evaluations, therefore, the teacher must exercise to integrate paradigms

and to modify the use of the other person's power to change permanent structures (i.e. one person's reality) into possibilities of others (Foucault, 1982).

In my work as a Tulsa Public School District New Teacher Mentor, I attempt to empower new teachers to discuss their performance with their evaluator. I suggest evaluation can assist the teacher in becoming an increasingly better instructor, one capable of relating to students and parents and observations and evaluation that critique and give feedback to "grow" the teacher. Yet, the opposite seems more of an occurrence; that being teachers experience the anxiety and pain of judgment. Foucault (1980) states the new pastoral power constantly questions, "Who are we?" (p. 212). Individuals struggle against this government of individualization especially when the TLE scores fall below an Effective rating. Teachers find the exposure humiliating when they receive TLE scores below what they think the scores should be. Many oppose the secrecy, deformation, and mystifying (Foucault, 1980) associated with scores. Why mystifying? Many teachers receive a score from their administrator without input on the score. Even if the composite score concludes with an Effective rating, the lack of teacher input results in a one-sided, top-down reality. A perfect composite score of 5 without input from both parties will still result in a sense of mysticism.

Evaluation scores on the TLE identify a teacher as superior, good, poor, or mediocre, labeling one's competence and making known this competence to teacher, administrators, and other teachers. The teacher who does not receive an Effective rating is punished. According to the Foucauldian theory of discipline, a person must be judged after attempts to norm his or her behavior (Foucault, 1977, p.183). This is clearly seen in the multiple attempts to norm teacher behavior in this high-stakes process that are the PDP and QUEST remedial programs. Teacher behavior is normed through further observation and scoring, and if the evaluator sees the teacher's performance as ineffective, the teacher will be dismissed from employment.

One of the *TLE Impacts,* a section within the TLE overview, states *its purpose* is to "exit ineffective teachers and evaluators." Although evaluators must qualify yearly in order to score, teachers do not know their evaluator's score, the level of mastery achieved by that evaluator, or evaluator inter-rater reliability and accuracy. The administrator exercises power to begin the exiting of teachers who might erroneously be labeled ineffective based on a single evaluation score or the teacher's noncompliance with recommended professional development. The teacher is known to the administration, but the administrator remains hidden to the teacher: the panoptic condition.

Foucault theorizes these dialogues (e.g. the conversation between teacher and evaluator) are not just "antiauthority struggles" (Foucault, 1982, p. 211), rather they are "transversal" and not limited to a country, or political or economic forms of government, and have wide application to societal institutions like education (Foucault, 1990). He sees these "struggles" as dialogical attempts between the teacher and administrator to affect the use of power that have personal consequence, especially since the teacher's job is the one at risk. These struggles question the status of both individual and governmental individuation (i.e. labeling of the individual by the institution). In other words, this type of questioning critically challenges the imposed reality of labeling and evaluation to create another reality in which power is shared through mutual exercises. However, our present historical circumstance of audit culture in education is a serious obstacle to the antiauthority struggle.

Recently, an experienced, well respected teacher I know, told me, through tears, her concern over what she considered low evaluation numbers. She had been recruited by her present principal and was concerned that he would think poorly of her if her scores were "average" (3). I asked her if she had provided documentation of her accomplishments. For example, did she include in her conversation with the principal (evaluator) the interesting and ingenious project she was presently engaged in with her classes? Perplexed, she said she had never thought of doing so. Evaluation is a high-stakes test for teachers, but I suggest it is possible for teachers and evaluators to think and therefore act differently through the use of dialogue about evidence presented during the evaluation meeting.

A Method to Mutually Exercised Power

Asymmetrical power is visible when the evaluator assigns a numerical score to a teacher's performance, thereby labeling the teacher Foucault (1983). As discussed previously, I suggest rubric language be used as a calibration of ideas for both teacher and evaluator. Both parties must learn the rubric language and what the descriptive words mean and look like in a school. Then, with common understandings both parties can enter into discussion over how the teacher's performance is judged and the teacher can provide evidence of a performance that matches a level on the rubric. Teachers must recognize their lack of exercised-power contributes to their own punishment and feeds the helplessness of asymmetrical power. In the age of audit culture, and over-emphasis on accountability through testing and blaming (Taubman, 2011) we must make the effort to influence our own culture through listening to one another with regard to teacher evaluation rather than multiply its already detrimental effect.

Presently teacher evaluation operates as panopticism (Foucault, 1977) and mysticism (Foucault, 1983), much like the distant control and nebulous fear of audit culture. Healthy "push-back" empowers teachers and helps them to master curricula. Broadly defined, curriculum is the path of learning and being; it is one's planned and unplanned experience. Curriculum is cultivation of the mind, self-reflexivity and an interdisciplinary "currere" (Pinar's term for the infinite form of "curriculum") (Pinar, 2012, p. 34) or life path. I suggest teacher evaluation must exist within this prodigious notion of currere, contributing to open and divergent educational processes in which teachers take risks, employing curricular methods to improve student learning outcomes.

School Reform and Teacher Professionalism and Identity

The TLE is considered an improvement over the older form of evaluation in Tulsa's school district because it involves a rubric process and not the checklist method of the older evaluation system, however unless we adopt the understanding of the growth made possible by feedback and discussion through the evaluation process, little real improvement or growth may be actualized, rather what becomes inevitable is a reinforcement of the older evaluation system of top-down judgment. Similar to Oklahoma's state agenda to improve student outcomes by having all students meet a cut score, teacher evaluation requires all teachers meet a certain numerical measure on the TLE rubric. Day (2003) states teaching has entered a post-professional historical period in which government challenges teachers' agency and curriculum content. Present reforms have created a sea change, radically restructuring what it means to be a teacher. For example, changes to teacher identity have shifted to corporate goals, managing a range of students, and documentation for accountability. Day (2003) suggests several ideas positively to promote teacher identity, most notably increasing teachers' opportunity to exercise judgment over their own teaching and of their care of students and a collaborative culture built to deal with problems. Such an attitude shift can promote improved practice. Day (2003) also argues teachers should be encouraged to undertake self-directed study and be recognized for high-level task complexity in order to enrich their teacher identity and practice.

Nias (paraphrased in Day, 2003) states teacher identity results from personal experience and the cultural and institutional environment in which they function. The architecture of teachers' professional lives is not stable, but subject to tension and difficulties. Day (2003) reminds us excellence can only be motivated, not coerced, and that, once a teacher's identity exhibits commitment and job satisfaction, he or she becomes

better able to build relationships with students and make effective adjustments to his or her practice. As a result, both pedagogy and teacher efficaciousness are improved, also the goal of the TLE. Evaluation triggers teachers' fear of judgment and a static identity over which they have no control. How would teacher dialogue as a method to an evaluation score affect teacher identity?

Shifting Teacher Paradigms

Initially, I supported the TLE because I maintain it is superior to the checklist that previously defined my teaching. Under the old system, with very little knowledge or effort, an administrator could check off that my teaching had been viewed and evaluated. Very little of my improvement was noted or ever documented and the previous method of evaluation did nothing to build my capacity, or attempts at improvement. Because the new TLE is a complex rubric with measureable levels of performance (1–5 rating), and boasts dialogue as a method to a score, I assumed all teachers would receive constructive feedback, accurate scores, and an opportunity to defend their desired scores. Unfortunately, my hopes run contrary to the testimony of many classroom teachers within the district (L. Stockley, personal communication, April 20, 2014). According to the *TLE Board of Education Mandate* (2013) there were 202 teachers on PDPs, an increase of 66 to the 136 teachers listed from the 2010–11 school year. Thirty-six teachers are in the QUEST program, compared to seven from the 2010–11 school year. Clearly, hierarchical observation has identified more teachers who are perceived as operating below the norm. Administrators likely characterize this as success of the apparatus because the TLE rubric makes this observation possible. However, normative judgment and the examination increases teacher stress and causes some prematurely to exit the teaching field (Perryman, 2007; L. Stockley, personal communication, April 20, 2014).

The overview introductory page to the TLE states this system allows districts to identify levels of professional performance, support less-effective teachers, and exit the ineffective. According to Foucault, this discipline of correct training is negative especially if the evaluator is only looking for the norm and the evaluator punishes behaviors that fail to conform to what is listed as normed teacher behavior (Foucault, 1977). The punishment increases if the evaluator is ill-trained or does not welcome the teacher to "leverage "teacher input as stated in the "Overview to the Teacher Leadership Effectiveness Evaluation."

Conclusion

We should ask ourselves, "What truth about a teacher is established?" through the TLE. Why do we as teachers fail critically to

examine the power in programs that examine our teaching? Has the teacher been given knowledge to exercise appropriate power in creating the truth that identifies his or her teaching? What or who ensures that a bullying administrator will not create his or her own untethered, unevidenced truth about the interpretation of the rubric, making the attainment of the norm impossible? Given my analysis of this teacher evaluation tool and its processes, I conclude the relationship between evaluator and teacher must change, affording a culture of collegial adult development. It is the paradigm of shared power that builds inclusiveness and respect for others. Within this culture are creativity, collaborative listening, and sharing that results in a reality created by both evaluator and teacher (Blase & Blase, 1999). Therefore, teachers should be taught and encouraged to take risks by discussing their own inquiry questions especially at the observation phase. Teachers can demonstrate their educational process to discover or create sound solutions that buttress their classroom performance. They should document their work with students and learn to defend that performance with evidence that aligns to the levels of use on the TLE rubric. Principals, usually the TLE evaluator for each school, must shift their paradigm to one of inclusion and collegiality concerning not only child, but also adult learning (Drago-Severson, 2012).

Seasoned evaluators and other administrators should realize the implicit, negative effects of evaluation, and promote openness to teacher dialogue and the consideration of teaching evidence within the TLE system. The importance of the score, essentially a high-stakes exam, threatens the teacher's continued employment and livelihood. Judgment is felt, regardless of the teacher's aptitude or of past scores. Therefore, teacher documentation should be taught and expected as a means to show mastery. Initially, the evaluator should be supportive by getting to know the teacher he or she evaluates and by practicing dialogical methods to counter the negative possibilities of evaluation. In this age of testing accountability and blaming of teachers and school personal, we should practice evaluation that does not mirror this type of ridicule. We can, however, practice shared power in educational practice and embrace mutual understanding, and consideration through dialogue in teacher evaluation.

References

Bess, M. (1980, 3 November). Power, moral values, and the intellectual: An interview with Michel Foucault. *History of the Present, 4*(1–2), 11–13.

Blase, J., & Blase, J. (1999). Effective instructional leadership. *Journal of Educational Administration*, *38*(2), 130–141.

Day, C. (2002). School reform and transitions in teacher professionalism and identity. *International Journal of Educational Research, 37*(8), 677–692.

Drago-Severson, E. (2012). New opportunities for principal leadership: Shaping school climates for enhanced teacher development. *Teachers College Record*, *114*(3), 1–44.

Dreyfus, H., & Rabinow, P. (1982). *Michel Foucault: Beyond structuralism and hermeneutics* (2nd ed.). Chicago: The University of Chicago Press.

Foucault, M. (1982). The subject and power. *Critical Inquiry*, *8*(4), 777–795.

Foucault, M. (1984). *The history of sexuality; Volume 1: An introduction*. New York: Vintage.

Foucault, M. (1995). *Discipline & punish: The birthplace of the prison*. New York: Vintage.

Perryman, J. (2007). Inspection and emotion. *Cambridge Journal of Education*, *37*(2), 173–190.

Taubman, P. (2009). *Teaching by numbers: Deconstructing the discourse of standards and accountability in education*. New York: Routledge.

Tulsa Public Schools (2014). *TLE observation & evaluation handbook for evaluators using the Tulsa model*. Retrieved from https://www.tulsaschools.org/4_About_District/_documents/TLE/Handbook_TLE_Observation_and_Evaluation_System.pdf