Winter 2015

Principal Dispositions Regarding the Ohio Teacher Evaluation System

Theodore J. Kowalski
University of Dayton, tkowalski1@udayton.edu

David Alan Dolph
ddolph1@udayton.edu

Research Article

Principal Dispositions Regarding the Ohio Teacher Evaluation System

Theodore J. Kowalski, PhD
Professor and Kuntz Family Endowed Chair
University of Dayton
Dayton, OH

David A. Dolph, PhD
Assistant Professor
Chair, Department of Educational Leadership
University of Dayton
Dayton, OH

Abstract

The Ohio Teacher Evaluation System (OTES) was first implemented during the 2013-14 school term. This study examined principals' dispositions at the end of this school term. Findings revealed several major concerns. The most prominent were (a) not having sufficient time to implement the program properly, (b) basing a teacher's performance heavily on student value-added data, and (c) being required to assist teachers in developing their annual improvement plans. Three independent variables, *teaching experience*, *administrative experience*, and *level of school assignment*, were found to have only a low level of association with principal dispositions. With respect to teacher evaluation generally, findings here were consistent with earlier studies reporting mixed principal dispositions; with respect to OTES specifically, findings here were consistent with studies in other states reporting that principal dispositions were more negative than positive.

Key Words

superintendent, leadership, school administration

Principal Dispositions Regarding the Ohio Teacher Evaluation System

Recent federal programs, such as the *No Child Left Behind Act* and the *Race to the Top Initiative*, reflect a commonly held belief: improving the accuracy and effectiveness of teacher evaluation and making school officials more accountable for the process are essential reforms (Strong, Gargani, & Hacifazlioglu, 2011). In 2009, the Ohio legislature responded to federal incentives by directing the state's Educators Standards Board to recommend a rigorous statewide approach for assessing teacher performance.

Subsequently, the Ohio State Board of Education approved a new model, naming it the Ohio Teacher Evaluation System (OTES). Initially implemented in the 2013-14 school term, the system included two requirements previously uncommon in Ohio; 50% of a teacher's annual performance had to be determined by student value-added scores and teachers, assisted by principals, had to develop annual individual growth plans.

The overall purpose of this study was to determine principal dispositions toward performance evaluation generally and toward OTES specifically. According to the National Council for Accreditation of Teacher Education (2009), educator dispositions are relevant to reforms because they represent values and commitments that define the performance of those who implement change. In the case of employee performance evaluation, principal opinions are especially germane because they affect both personal behavior and teachers' attitudes and beliefs (Youngs, 2007).

Data were collected after the principals had implemented OTES for the first time. Findings indicate that the respondents' general views about performance evaluation were rather typical when compared to previous studies. Their temperament toward OTES, however, was primarily negative. Three variables often linked to principal dispositions, *teaching experience*, *administrative experience*, and *level of school assignment* (elementary or secondary) were examined. All three were found to have a low level association with the dispositions.

Prior Literature

Historically, teacher performance evaluation has evolved from end of the year checklists to far more sophisticated models that emphasized both summative and formative judgments (Danielson, 2002).

Recognizing the growing complexity of the process, Medley and Coker (1987) examined its effectiveness nearly 3 decades ago. They found the validity of teacher evaluations conducted by administrators to be unacceptably low. Since then, countless other studies have been conducted in an effort to better understand and improve the procedure. Two aspects of previous research are especially relevant here: educator dispositions toward teacher performance evaluation and research on state-mandated performance evaluation systems.

Performance Evaluation Dispositions

Dispositions are relevant because they have a behavioral component. That is, attitudes and feelings toward a responsibility influence behavior, particularly in relation to pursuing that duty. Thus, if administrators believe differentiating between good and bad instruction is impossible or if they believe that candid discussions with teachers do more harm than good, they act accordingly. Equally important, their personal behavior then

influences what teachers believe about the efficacy of performance evaluation and how they feel about being subjected to the process (Tuytens & Devos, 2010).

Most studies examining teacher and principal dispositions have yielded rather consistent findings. With regard to the former, teacher temperaments have been mixed but skewed toward being more negative than positive. For instance, in a national study, Duffett, Farkas, Rotherham, and Silva (2008) reported that only 26% of the teachers thought their evaluations were effective and useful.

Another study (Louis et al., 2010) reported that only 38% of teachers considered classroom observations helpful in relation to improving instruction. With respect to the latter, principal dispositions also have been mixed but skewed slightly toward being more positive than negative. For instance, studying Iowa principals, Armendt, (2004) found that 68% said the process had improved and 52% said they did not require additional training to conduct the process effectively.

Comparing the two groups, Armstrong (1988) found a statistically significant difference between them with principals expressing the more positive opinions; however, in-group variance among principals was considerably higher than it was among teachers.

Much of the literature on educator opinions has centered on problems and constraints. The following are notable examples of these findings:

- School culture has been identified as a primary barrier. Donaldson (2013), for instance, reported that the effectiveness of teacher evaluation in many schools has been diminished by shared negative values, beliefs and norms. Likewise, Louis and associates (2010) found that educator dispositions on performance evaluation often contravened professional norms and public policy.

- Dandoy (2012) and Kersten and Israel (2005) found that collective bargaining agreements unduly restricted what could be assessed, how assessments occurred, and when and where they occurred.

- Marshall (2005) and Youngs (2013) concluded that classroom observations often were conducted using invalid or unreliable instruments. In addition, Marzano (2012) found that the effectiveness of classroom observations often has been diminished because the evaluator did not understand the process; specifically, sampling errors resulted in principal ratings not being based on actual behavior.

- The presence of evaluator bias and subjectivity in areas such as age, experience, gender, and race has been reported in multiple studies such as those conducted by Donaldson (2013) and Tucker and Stronge (2005).

- Another pervasive problem identified in previous research is inadequate human and material resources (e.g., Coulter, 2013).

- The most pervasive problem in the eyes of principals has been time restrictions (e.g., Donaldson, 2013; Hill, 2013; Kersten & Israel, 2005).

- Painter (2001) reported that principals believe that defining and measuring effective teaching is inherently difficult.

Opinions about principal self-efficacy also are relevant. In both an Ohio study (Himmelein, 2009) and Massachusetts study (Ford, 2014), the researchers found that a majority of respondents believed they had the requisite knowledge and skills to evaluate teacher performance. In a study of nearly 300 Arizona principals, however, Painter (2001) found that a majority were dissatisfied with the level of training they had received in this area.

Although findings regarding principal self-efficacy have been mixed, teachers' opinions about principal expertise have been largely negative and consistent. Specifically, teachers have expressed doubt about principals being able to assess teachers across multiple subject areas or grade levels (e.g., Duffet et al., 2008; Oppenhiem, 1994), to conduct assessments relevant to instructional improvement (e.g., Louis et al, 2010; Peterson, 2000), and to apply assessment procedures correctly and consistently (e.g., Zimmerman & Deckert-Pelton, 2003).

Research examining evaluation outcomes also has revealed problems. For example, in a study spanning 12 districts in four states, the vast majority of teachers received the highest rating possible but conversely, dismissals in this defined population were extremely rare (Weisberg, Sexton, Mulhern, & Keeling, 2009). After analyzing numerous studies, Menuey (2005) noted that various researchers have estimated the level of incompetent teachers to be between 2% and 20%, with 5% being the modal approximation. Yet, research reveals that less than 1% of teachers have been dismissed annually. Menuey described the discrepancy between the level of incompetent teachers and teacher dismissals as "gross" and "staggering" (p. 310).

Studies also have revealed the presence of subjectivity and bias. Typically, these conditions have resulted in discrimination, especially in the areas of gender and race (e.g., Rinehart & Young, 1996). In addition to subjectivity, leniency in performance evaluations has been found to exist across all types of organizations, especially when performance ratings determined or influenced high-stake decisions, such as job retention, promotion, or tenure (Jacob & Lefgren, 2007).

Studies examining possible associations between principal attitudes and personal characteristics have been limited and their findings mixed. Studying dispositions toward Iowa's mandated evaluation program, Amendt (2004) found a significant difference between relatively inexperienced principals (less than 4 years) and their peers regarding program effectiveness with the former group having more positive beliefs. Conversely, Fisicaro (2010), studying New Jersey principals, found highly experienced principals (over 15 years) to have more positive views about teacher evaluation than their peers.

Several other studies have looked at possible associations between leadership style, a factor arguably relevant to conducting evaluations, and levels of professional experience. Results of these inquiries also have been mixed with most having found no statistically significant association between the two variables (e.g., Bentley, 2011; Cooper, 2011).

A few studies have examined a possible association between principal opinions and the

level of school assignment (elementary or secondary). Often, educators assume performance evaluation is more difficult for secondary principals, primarily because the quantity is greater and the nature (across multiple subject areas) is more complex. Nevertheless, most studies examining level of school assignment as an independent variable and principal opinions as a dependent variable (e.g., Cardine, 1998) have found no statistically significant association.

State-Mandated Evaluation Systems

The number of state-mandated paradigms proliferated over the past 2 decades, largely because of fiscal incentives embedded in the federal program, A Race to the Top. Commonly, state systems include two mandates: student performance, assessed by value-added achievement scores, must be a component of a teacher's evaluation and each teacher must develop an annual professional growth plan. Both obligations have been and remain controversial.

With respect to the former requirement, many teachers and principals believe that placing considerable weight on value-added learning data, a condition that currently exists in 40 states (Collins & Amrein-Beardsley, 2014), is unfair. Although some researchers (e.g., Kimball & Milanowski, 2009; Taylor & Tyler, 2012) have urged state policymakers to rely on these metrics, others either have challenged the validity of these measures (e.g., Kerstling, Mei-kuang, & Stigler, 2013) or have concluded that they are invalid (e.g., Berliner, 2013; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Konstantopoulous, 2014). Expectedly, teacher opinions about using value-added data to determine their performance have been predominantly negative. In a California study, for instance, Lee (2012) found that most teachers believed that the mandate was not only unfair, it likely would force them to change curriculum and instructional methods.

To a lesser extent, concerns also have been expressed about requiring teachers to develop individual growth plans under the guidance of a principal. Teacher opposition to this mandate appears to be nested in skepticism; that is, many teachers have been unconvinced that principals can provide them with meaningful guidance (Stark & Lowther, 1984; Zimmerman & Deckart-Pelton, 2003). Although principals' opinions about assisting teachers to develop growth plans are largely unknown, persistent concerns about the amount of time spent evaluating teachers (e.g., Hill, 2013; Maharaj, 2014) suggest that their attitudes are likely to be negative.

Recently, researchers have examined opinions of specific state-mandated programs. This body of research has disclosed myriad concerns. Educator apprehensions were not unexpected given the fact that state programs often contained as many or more constraints than the models they replaced (Hinchey, 2010). In a Colorado study, for example, Ramirez, Clouse, and Davies (2014) described that state's policy as over-reaching, unduly time consuming, and poorly designed.

Other state studies reveal the depth of educator concerns. As examples, in Georgia (Eady & Zepeda, 2007), Washington (Coulter, 2013), and Missouri (Killian, 2010), researchers reported mostly negative dispositions. Equally notable, disapproval of using value-added data was pervasive and concerns about specific state programs were nearly identical to those recorded in studies addressing performance evaluation in general

(e.g., lack of resources, excessive time requirements, inclusion of value-added metrics, and unrealistic expectations).

A notable exception among the state studies is research conducted by Lasswell, Pace, and Reed (2008) in Iowa. They found that principal opinions toward that state's system were primarily positive; however, their study population included only principals from small rural districts. Limited research (e.g., Ferguson, 1981) suggests that principals in small-enrollment districts have received substantially less performance evaluation training than have principals from large-enrollment districts. Thus, the nature of the Iowa study population may largely explain the atypical finding.

Ohio Study of Principals' Dispositions

Description

This study of Ohio principals was conducted immediately after teachers were evaluated under OTES for the first time. The research was guided by three questions:

1. What are the principals' opinions regarding teacher performance evaluation?
2. What are the principals' opinions regarding OTES?
3. What level of association exists between the dependent variable (opinions of OTES) and each of three independent variables (respondent *teaching experience*, *administrative experience*, and *level of school assignment*)?

The first question focused on opinions regarding educator dispositions toward performance evaluation generally; the second question focused on opinions regarding OTES. Both questions were answered using descriptive statistics (frequencies, percentages, and rank order). The rank order of the responses was determined by calculating the percentage of respondent agreement for each statement.

Because opinion data were continuous and demographic data were dichotomous, the third research question was answered by calculating point biserial correlation coefficients. The coefficients were then applied as descriptive statistics using a typology recommended by Cohen and Cohen (1983):

- Small association: (+ or -) correlations from .01 to .29
- Moderate association: (+ or -) correlations from .30 to .49
- Large association: (+ or -) correlations of .50 and higher

The defined study population consisted of 89 principals employed in public elementary and secondary schools located in three Southwestern Ohio counties. Data were collected in May and June of 2014 using a paper survey developed by the researchers. Content validity was established by a panel of experts, all of whom were former principals and current professors.

Limitations

The study had three notable limitations. First, the defined population only included public elementary or secondary school principals in three Ohio counties. Second, findings relied on the accuracy of self-reported beliefs. As such, validity depends on principals having sufficient self-awareness and responding honestly. Third, no inferences could be made about the study population's non-responders.

Findings

Completed surveys that could be analyzed were returned by 50 principals, a return rate of 56%. The respondents were almost equally divided in terms of the level of assigned schools, with 54% being secondary school principals and the remainder being elementary school principals.

Response percentages for nine statements about teacher performance evaluation in general are in Table 1. The statements appear in rank order based on the percentage of respondent agreement (highest to lowest). A majority of respondents disagreed that they and teachers had a positive disposition toward the evaluation process in general.

Table 1

Opinions about Teacher Performance Evaluation in General

*Rank**	*Statements*	*Percentages*			
		SD	*D*	*A*	*SA*
1	Principals have the skills necessary to complete teacher evaluations effectively.	2	10	68	20
2	Evaluation data are used by principals to improve the quality of instruction.	4	18	56	22
3	Principals consider teacher evaluation to be one of their most important duties.	10	14	52	24
4	Principals have the knowledge necessary to complete teacher evaluations effectively.	0	28	54	18
5	Evaluation data are used by principals to determine if a teacher is competent.	2	30	60	8
5	Teachers have confidence in the evaluation data generated by principals.	4	28	62	6
7	Evaluation data are used by principals to determine if a teacher should be reemployed.	4	33	47	16
8	Principal dispositions regarding performance evaluation are positive.	16	38	40	6
9	Teacher dispositions (attitudes/beliefs) regarding performance evaluation are positive.	14	48	36	2

*Statements are ranked from highest to lowest respondent agreement

Legend: SA = Strongly Agree; A = Agree; D = Disagree; SD = Strongly Disagree

Response percentages for the 16 OTES-related statements are in Table 2. Again, the statements appear in rank order based on the percentage of respondent agreement. Overall, the principals' responses reveal that opinions regarding OTES were substantially more negative than opinions about performance evaluation generally.

Table 2

Opinions about Ohio Teacher Evaluation System

Rank	*Statements*	*Percentages*			
		SD	*D*	*A*	*SA*
1	The amount of time I spend on the OTES is excessive.	4	0	10	86
2	The pre-conference requirement is an effective OTES element	8	16	62	14
3	The amount of time teachers I supervise spend on the OTES is excessive.	4	24	24	48
4	I know how to apply the OTES correctly.	2	30	56	12
5	The OTES is increasing the quantity of time I spend with supervising teachers.	10	24	28	38
6	The scope of the OTES is understood by the teachers I supervise.	6	38	50	6
7	Instructions for applying the OTES are clear to me.	6	42	46	6
8	The professional growth plan requirement is an effective OTES element.	12	40	42	6
8	The teachers I supervise know how to apply the OTES correctly.	6	46	42	6
10	Instructions for applying the OTES are clear to the teachers I supervise.	12	42	40	6
11	The OTES is increasing the accuracy of teacher evaluations.	18	45	31	6
12	I have a positive disposition regarding the OTES.	20	44	28	8
13	The OTES is having a positive effect on teaching and learning.	27	45	20	8
14	The OTES has improved my relationships with the teachers I supervise.	18	54	24	4
15	The weight placed on student growth measures (50%) in the OTES is fair.	36	44	18	2
16	The teachers I supervise have a positive disposition regarding the OTES.	24	58	14	4

*Statements are ranked from highest to lowest respondent agreement

Legend: SA = Strongly Agree; A = Agree; D = Disagree; SD = Strongly Disagree

Responses to three demographic questions were dichotomous and the percentages are shown in Table 3. Associations between respondent beliefs about OTES and each of the demographic variables were determined by calculating point bi-serial correlations. The coefficients were then categorized as being large, medium, or small as described earlier. The coefficients and categorization outcomes are in Table 4. As these data reveal, all three association were small, with the highest level of association being negative.

Table 3

Demographic Characteristics of the Respondents

Teaching experience		*Administrative experience*		*Level of assignment*	
< 11 years	11 > years	< 11 years	11 > years	Elementary	Secondary
38%	62%	62%	38%	46%	54%

Table 4

Levels of Association between Opinions about Ohio Teacher Evaluation System and Demographic Variables

Variable	Correlation coefficient	Level of association
Teaching experience	+.19	Small positive
Administrative experience	+.03	Small positive
Assignment level	-.27	Small negative

Discussion

The main purpose of this study was to determine dispositions of a defined population of Ohio principals toward performance evaluation generally and OTES specifically. At the time the study was conducted, the following three pieces of evidence suggested that their temperaments would be more negative than positive.

1. Relatively recent studies (e.g., Duffett et al., 2008; Louis et al., 2010) have revealed that educator skepticism regarding the validity and usefulness of teacher evaluations remains considerable.
2. Using value-added student data to determine teacher performance has been criticized not only by education associations but also by several prominent scholars, such as Berliner (2013) and Darling-Hammond and associates (2012). Expectedly, studies examining reactions to this mandate (e.g., Lee, 2012) have reported substantial teacher opposition to it.
3. Mandating educators to implement changes (i.e., using a power-coercive strategy), especially those they do not support, almost always have failed to be institutionalized (Kowalski, 2011; St. John, Griffith, & Allen-Haynes, 1997).

Nevertheless, in light of the fact that OTES constituted a radical change in teacher evaluations, there was a need to determine if this assumption was accurate.

With respect to opining about performance evaluation nationally, outcomes reported in this study are congruent with previous research findings. As examples, most respondents in the Ohio study thought that principals understood the importance of performance evaluation and possessed the requisite knowledge and skills to apply it properly. These findings reinforce evidence reported earlier by Armstrong (1988), Himmelein (2009) and Kersten and Israel (2005). Nevertheless, principal self-perceptions should be weighed in relation to teachers' perceptions of principal efficacy.

Teachers have tended to rate principals' expertise much lower as demonstrated in investigations conducted by Armstrong (1988), Duffet et al. (2008), and Oppenhiem (1994); unfortunately, explanations for the disparate views remain imprecise. Equally notable, the Ohio study found that most principals believed that educator dispositions toward performance evaluation were more negative than positive. This outcome reinforces data reported in a recent national study conducted by Louis and associates (2010).

With respect to OTES specifically, several findings are noteworthy. First, 96% of the principals agreed that the time they had devoted to implementing the new system was excessive. This finding is congruent with numerous studies reporting that principals consider insufficient time to be their most serious constraint (e.g., Hill, 2013; Kersten & Israel, 2005; Killian, 2010). Instead of attempting to mitigate this problem, OTES, especially the mandate for principal involvement in teacher professional growth plans, exacerbates time requirements.

Second, the level of opposition to using student value-added measures reported here was considerable. A similar finding was reported in a recent California study conducted by Lee (2012). As previously noted, resistance to judging teacher performance on the basis of value-added scores appears to be pervasive in

the education profession. From a political perspective, widespread opposition to basing 50% of a teacher's evaluation on this metric already has resulted in legislation that lowers the percentage in OTES for the next school year.

Third, the level of skepticism expressed about individual teacher growth plans merits attention; 52% of the principals did not believe this provision is effective. This finding is relevant in light of studies revealing that many teachers are skeptical about the ability of principals to conduct formative evaluations (e.g., Stark & Lowther, 1984; Zimmerman & Deckart-Pelton, 2003).

Many questions about the principal's responsibility to assist individual teacher growth remain unanswered. As examples what will occur if a teacher refuses to apply the advice provided by his or her principal? What will occur if a principal fails to meet his or her responsibility to provide advice?

Fourth, a majority of respondents did not believe that OTES would produce positive outcomes in several critical areas including (a) overall school-improvement, (b) principal-teacher relationships, and (c) the validity and reliability of performance evaluations. Similar levels of pessimism about state-mandated systems have been reported in studies in Georgia (Eady & Zepeda, 2007), Missouri (Killian, 2010), and Washington (Coulter, 2013). This growing body of evidence suggests that widespread cynicism will fuel resistance to state mandates.

Last, this study examined the extent to which principal dispositions about OTES were associated with three independent variables: *teaching experience*, *administrative experience*, and *level of school assignment* (elementary or secondary).

Both experience variables were found to have a low level of association with the dependent variable (dispositions). Likewise, level of school assignment had a small-negative association with the dependent variable, indicating that being an elementary or secondary school principal did not heavily influence dispositions toward OTES.

Recognizing structural and application deficiencies in teacher evaluation, state policymakers have been applauded by many stakeholders for taking actions intended to improve the situation. Unfortunately, many state systems appear to include more problems than the systems they have replaced (Hinchey, 2010).

As such, the growing body of research on mandated teacher evaluation programs indicates two primary concerns. One is that the requirements may exacerbate rather than resolve persistent reliability and validity problems. The other is that political resistance will incrementally result in reversion; that is, considerable human and material resources will be expended on state programs that will have a short lifespan.

Although limited in scope, this study provides additional insights regarding three highly relevant issues: (a) educator skepticism about the validity and reliability of teacher evaluation; (b) judging teacher performance on the basis student value-added metrics; (c) the possible effects of educator disposition on the institutionalization of state-mandated evaluation systems.

Given the proliferation of state systems, the need for additional investigations is axiomatic. Specifically, future research is encouraged in the following areas: studies of mandated systems in other states, examining variables underlying the disparity between principal and teacher views of principal efficacy, examining variables associated with educator dispositions toward teacher evaluation, and the development of alternative models that include reliable and valid components aligned with the existing knowledge base on performance evaluation.

Author Biographies

Theodore Kowalski is professor and the Kuntz Family Endowed Chair in educational administration at the University of Dayton. E-mail: tkowalski1@udayton.edu

David Dolph is an assistant professor and chair at the University of Dayton's Department of Educational Leadership. He is a former school principal and district superintendent. E-mail: ddolph@udayton.edu

References

Amendt, B. C. (2004). *A study of administrator perceptions of state-mandated teacher evaluation: The student achievement and Iowa teacher quality law*. Doctoral dissertation, Drake University, Des Moines, Iowa.

Armstrong, S. R. (1988). *Perceptions of principals and teachers toward mandated teacher evaluation*. Doctoral dissertation, Oklahoma State University, Stillwater, Oklahoma.

Bentley, K. (2011). *An investigation of the self-perceived principal leadership styles in an era of accountabilit*y. Doctoral dissertation. University of South Florida, Tampa, Florida.

Berliner, D. C. (2013). Problems with value-added evaluations of teachers? Let me count the ways! *Teacher Educator, 40*(4), 235-243.

Cardine, L. B. (1998). *Principals' perception of teacher evaluation as a support for decision-making*. Doctoral dissertation, Columbia University, Teachers College, New York, New York.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, C. & Amrein-Beardsley, A. (2014). Putting growth and value-added models on the map. *Teachers College Record, 116*(1). Retrieved from http://www.tcrecord.org/Content.asp?ContentId=17291

Cooper, G. J. (2011). *The influences of school variables on the principals' instructional leadership style in elementary schools in an urban setting*. Doctoral dissertation, University of North Carolina, Chapel Hill, North Carolina.

Coulter, M. P. (2013). *A qualitative study of teacher and principal perceptions of Washington state teacher evaluation instruments*. Doctoral dissertation, Washington State University, Pullman.

Dandoy, J. R. (2012). *Principals' perceptions of barriers to dismissal of poor-performing teachers*. Doctoral dissertation, University of Kansas, Lawrence, Kansas.

Danielson, C. (2002). *Enhancing student achievement: A framework for school improvement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 96*(6), 5-6.

Donaldson, M. L. (2013). Principals' approaches to cultivating teacher effectiveness: Constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Educational Administration Quarterly, 49*(5), 838-882.

Duffett, A., Farkas, S., Rotherham, A., & Silva, E. (2008). *Waiting to be won over: Teachers speak on the profession, unions, and reform*. Washington, DC: Education Sector.

Eady, C. K., & Zepeda, S. J. (2007). Evaluation, supervision, and staff development under mandated reform: The perceptions and practices of rural middle school principals. *Rural Educator, 28*(2), 1-7.

Ferguson, V. S. (1981). *Principals' perceptions concerning teacher evaluation in the state of Washington*. Doctoral dissertation, Seattle University, Seattle, Washington.

Fisicaro, R. J. (2010). *Teacher evaluation: Assessing principals' perceptions in the state of New Jersey*. Doctoral dissertation, Liberty University, Lynchburg, Virginia.

Ford, M. C. (2014). *Instructional self-efficacy of principals in the context of a statewide educator evaluation system*. Doctoral dissertation, University of Massachusetts, Lowell, Massachusetts.

Hinchey, P. H. (2010). *Getting teacher assessment right: What policymakers can learn from research.* Boulder, CO: National Education Policy Center.

Hill, K. A. (2013). *Walking the tightrope: Secondary school principals' perspectives on teacher evaluation*. Doctoral dissertation, George Washington University, Washington, DC.

Himmelein, M. E. (2009). *An investigation of principals' attitudes toward teacher evaluation processes*. Doctoral dissertation, University of Toledo, Toledo, Ohio.

Jacob, B. A., & Lefgren, L. (2007). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics, 26*(1), 101–136.

Kersten, T. A., & Israel, M. S. (2005). Teacher evaluation: Principals' insights and suggestions for improvement. *Planning & Changing, 36*(1/2), 47-67.

Kersting, N. B., Mei-kuang, C. & Stigler, J. W. (2013). Value-added teacher estimates as part of teacher evaluations: Exploring the effects of data and model specifications on the stability of teacher value-added scores. *Education Policy Analysis Archives, 21*(6/7), 1-39.

Killian, B. R. (2010). *Administrators' and teachers' perceptions of the efficacy of the Missouri performance-based teacher evaluation model*. Doctoral dissertation, Saint Louis University, St. Louis, Missouri.

Kimball, S., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly, 45*(1), 34-70.

Konstantopoulos, S. (2014). Teacher effects, value-added models, and accountability. *Teachers College Record, 116*(1). Retrieved from https://www.tcrecord.org/library/Abstract.asp?ContentId=17290

Kowalski, T. J. (2011). *Public relations in schools* (5th ed.). Boston, MA: Pearson, Allyn and Bacon.

Lasswell, T. A., Pace, N. J., & Reed, G. A. (2008). Weighing in: Rural Iowa principals' perceptions of state-mandated teaching evaluation standards. *Rural Educator, 29*(3), 40-44.

Lee, L. (2012). *Voices less heard: Teachers' perspectives on the evaluation of teacher effectiveness through the use of value-added modeling*. Doctoral dissertation, California State University, Los Angeles, California.

Louis, K. S., Leithwood, K., Walhlstrom, K. L., Anderson, S. E., Michlin, M., Mascall, B., Gordon, M., Strauss, T., Thomas, E., & Moore, S. (2010). *Investigating the links to improved learning: Final report of research findings*. Minneapolis, MN: Center for Applied Research and Educational Improvement.

Marshall, M. (2005). It's time to rethink teacher supervision and evaluation. *Phi Delta Kappan, 86*, 727-730.

Maharaj, S. (2014). Administrators' views on teacher evaluation: Examining Ontario's teacher performance appraisal. *Canadian Journal of Educational Administration & Policy*, (152), 1-58.

Marzano, R. J. (2012). Reducing error in teacher observation scores. *Educational Leadership,70*(3), 82-83.

Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *Journal of Educational Research, 80*, 242-247.

Menuey, B. P. (2005) Teachers' perceptions of professional incompetence and barriers to the dismissal process. *Journal of Personnel Evaluation in Education, 18*(4), 309-325.

National Council for Accreditation of Teacher Education (2009). *NCATE 2009 standards*. Washington, DC: author.

Oppenheim, C. (1994). *Encouraging evaluation utilization by preserving teacher self-esteem*. Washington, DC: Center for Research in Educational Accountability and Teacher Evaluation.

Painter, S. R. (2001). Barriers to evaluation: Beliefs of elementary and middle school principals. *Planning & Changing, 32*(1/2), 58-70.

Ramirez, A., Clouse, W., & Davis, K. W. (2014). Teacher evaluation in Colorado: How policy frustrates practice. *Management in Education, 28*(2), 44-51.

Rinehart, J., & Young, P. (1996). Effects of teacher gender and principal gender on ratings of teacher performance. *Journal of Personnel Evaluation in Education, 10(*4), 313-323.

St. John, E. P., Griffith, A. I., & Allen-Haynes, L. (1997). *Families in schools: A chorus of voices in restructuring*. Portsmouth, NH: Heinemann.

Stark, J. S. & Lowther, M. A. (1984). Predictions of teachers' preferences concerning their evaluation. *Educational Administration Quarterly, 20*(4), 76-106.

Strong, M., Gargani, J., & Hacifazlioglu, O. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education, 62*(4), 367-382.

Taylor, E. S., & Tyler, J. H. (2012). Can teacher evaluation improve teaching? *Education Next, 12*(4), 78-84.

Tucker, P. & Strong, J. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: Association for Supervision and Curriculum Development

Tuytens, M., & Devos, G. (2010). The influence of school leadership on teachers' perceptions of teacher evaluation policy. *Educational Studies, 36*(5), 521-536.

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

Youngs, P. (2007). How elementary principals' beliefs and actions influence new teachers' experiences. *Educational Administration Quarterly, 43*(1), 101-137.

Youngs, P. (2013). *Using teacher evaluation reform and professional development to support common core assessment*. Retrieved from www.americanprogress.org/wp-content/uploads/201304/Using-Teacher-Evaluation-and-PD-to-Support

Zimmerman, S., & Deckert-Pelton, M. (2003). Evaluating the evaluators: Teachers' perceptions of the principal's role in professional evaluation. *NASSP Bulletin 87*(636), 28-37.